

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 700 274**

51 Int. Cl.:

C12N 1/21 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **14.03.2014 PCT/US2014/029804**

87 Fecha y número de publicación internacional: **25.09.2014 WO14153253**

96 Fecha de presentación y número de la solicitud europea: **14.03.2014 E 14769797 (3)**

97 Fecha y número de publicación de la concesión europea: **05.09.2018 EP 2970872**

54 Título: **Microorganismos y métodos para producir oligosacáridos sialilados y que contienen N-acetilglucosamina**

30 Prioridad:

14.03.2013 US 201361782999 P

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

14.02.2019

73 Titular/es:

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

72 Inventor/es:

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

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 700 274 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Microorganismos y métodos para producir oligosacáridos sialilados y que contienen N-acetilglucosamina

Campo de la invención

5 La invención proporciona composiciones y métodos para producir oligosacáridos purificados; en particular, ciertos oligosacáridos que contienen N-acetilglucosamina y/o sialilados que se encuentran típicamente en la leche humana.

Antecedentes de la invención

10 La leche humana contiene un conjunto diverso y abundante de oligosacáridos neutros y ácidos (oligosacáridos de la leche humana, hMOS). Muchas de estas moléculas no son utilizadas directamente por los bebés en su nutrición, pero, sin embargo, tienen un papel fundamental en el establecimiento de un microbioma intestinal sano, en la prevención de enfermedades y en la función inmunológica. Antes de la invención, la capacidad de producir hMOS de forma económica a gran escala era problemática. Por ejemplo, la producción de hMOS a través de la síntesis química se ha visto limitada por problemas estereo-específicos, disponibilidad de precursores, impurezas del producto y alto costo general. Como tal, existe una necesidad apremiante de nuevas estrategias para fabricar grandes cantidades de hMOS a bajo costo para una variedad de aplicaciones comerciales.

Sumario de la invención

15 La invención presenta métodos eficientes y económicos para producir oligosacáridos que contienen N-acetilglucosamina y/o sialilados.

20 La invención proporciona un método para producir un oligosacárido que contiene N-acetilglucosamina en una bacteria que comprende los siguientes pasos: proporcionar una bacteria que comprenda una UDP-GlcNAc:Gala/β-R β 3-N-acetilglucosaminiltransferasa exógena, una permeasa de lactosa funcional, y una mayor capacidad de producción de UDP-GlcNAc que comprende la sobreexpresión de nagC, de modo que la bacteria produzca al menos un 10% más de UDP-GlcNAc que una bacteria nativa; y cultivo de la bacteria en presencia de lactosa. El oligosacárido que contiene N-acetilglucosamina se recupera entonces de la bacteria o de un sobrenadante de cultivo de la bacteria.

25 La invención proporciona además un método para producir un oligosacárido sialilado en una bacteria que comprende los siguientes pasos: proporcionar una bacteria que comprenda un gen de sialil-transferasa exógeno, una vía catabólica deficiente en ácido siálico, una capacidad sintética de ácido siálico y un gen funcional de permeasa lactosa, y una mayor capacidad de producción de UDP-GlcNAc que comprende la sobreexpresión de nagC, de modo que la bacteria produzca al menos un 10% más de UDP-GlcNAc que una bacteria nativa; y el cultivo de la bacteria en presencia de lactosa. El oligosacárido sialilado se recupera de la bacteria o de un sobrenadante de cultivo de la bacteria. Específicamente, una capacidad sintética de ácido siálico comprende la expresión de CMP-Neu5Ac sintetasa exógena, una sintasa de ácido siálico exógena y una UDP-GlcNAc-2-epimerasa exógena, o una variante funcional o su fragmento.

35 En ambos métodos para producir oligosacáridos que contienen N-acetilglucosamina y/o sialilados, la bacteria comprende además la capacidad de aumentar la producción de UDP-GlcNAc. Por "mayor capacidad de producción" se entiende que la bacteria huésped produce más del 10%, 20%, 50%, 100%, 2 veces, 5 veces, 10 veces o más de un producto que la bacteria nativa, endógena. Por lo tanto, la bacteria sobreexpresa el regulador endógeno positivo de la síntesis de UDP-GlcNAc nagC. Por ejemplo, la bacteria sobreexpresa el gen nagC de Escherichia coli. Además, la bacteria sobreexpresa el gen glmS de Escherichia coli (L-glutamina:D-fructosa-6-fosfato aminotransferasa), o sobreexpresa el gen glmY de Escherichia coli (un regulador positivo de la traducción de glmS), o alternativamente sobreexpresa el gen glmZ de Escherichia coli (otro regulador de traducción positivo de glmS, glmY y glmZ se describen en Reichenbach et al, Nucleic Acids Res 36, 2570-80 (2008)). Alternativamente, la bacteria sobreexpresa cualquier combinación de tales enfoques además de nagC. Por ejemplo, la bacteria sobreexpresa nagC y glmS. Alternativamente, la bacteria sobreexpresa nagC y glmY. Alternativamente, la bacteria sobreexpresa nagC y glmZ. Los métodos también incluyen la sobreexpresión de cualquier variante o fragmento funcional de nagC, glmS, glmY y glmZ y cualquier combinación de los mismos. Por "sobreexpresión" se entiende que la transcripción del gen o el producto genético codificado es 10%, 20%, 50%, 2 veces, 5 veces, 10 veces o más que el nivel expresado o producido por el nativo correspondiente, o el gen endógeno natural.

50 La descripción detalla la manipulación de genes y rutas dentro de bacterias tales como la enterobacteria Escherichia coli K12 (E. coli) que conduce a un alto nivel de síntesis de hMOS. Otras cepas de E. coli adecuadas para su uso en la presente invención incluyen E. coli MG1655, E. coli W3110, E. coli DHSaE, E. coli B, E. coli C y E. coli W. Varias especies bacterianas son adecuadas para su uso en los métodos de biosíntesis de oligosacáridos, por ejemplo, Erwinia herbicola (Pantoea agglomerans), Citrobacter freundii, Pantoea citrea, Pectobacterium carotovorum o Xanthomonas campestris. Las bacterias del género Bacillus son adecuadas para su uso, incluyendo Bacillus subtilis, Bacillus licheniformis, Bacillus coagulans, Bacillus thermophilus, Bacillus laterosporus, Bacillus megaterium, Bacillus mycoides, Bacillus pumilus, Bacillus lentus, Bacillus cereus y Bacillus circulans. De manera similar, las bacterias de los géneros, Lactobacillus y Lactococcus se modifican utilizando los métodos descritos, incluidos, entre otros,

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. También se incluyen cepas, modificadas como se describe, de los géneros Enterococcus (p. ej., Enterococcus faecium y Enterococcus thermophilus), Bacteroides (p. ej., Bacteroides caccae, Bacteroides cellulosilyticus, Bacteroides dorei, Bacteroides eggerthii, Bacteroides finegoldii, Bacteroides fragilis, Bacteroides nordii, Bacteroides ovatus, Bacteroides salyersiae, Bacteroides thetaiotaomicron, Bacteroides uniformis, Bacteroides vulgatus y Bacteroides xylanisolvens), Bifidobacterias (p. ej., Bifidobacterium longum, Bifidobacterium infantis, y Bifidobacterium bifidum), Parabacteroides (p. ej., Parabacteroides distasonis, Parabacteroides goldsteinii, Parabacteroides johnsonii y Parabacteroides merdae), Prevotella (p. ej., Prevotella copri), Sporolactobacillus spp., Micromonospora spp., Micrococcus spp., Rhodococcus spp., y Pseudomonas (p. ej., Pseudomonas fluorescens y Pseudomonas aeruginosa). Las bacterias que comprenden las características descritas en el presente documento se cultivan en presencia de lactosa, y se recupera un oligosacárido sialilado o que contiene N-acetilglucosamina, ya sea de la propia bacteria o de un sobrenadante de cultivo de la bacteria. El oligosacárido sialilado o que contiene N-acetilglucosamina se purifica para su uso en productos terapéuticos o nutricionales, o las bacterias se usan directamente en dichos productos.

La bacteria de la presente invención comprende una capacidad de producción de UDP-GlcNAc que comprende la sobreexpresión de nagC, tal que la bacteria produce al menos un 10% más de UDP-GlcNAc que una bacteria natural. Además, la bacteria comprende un gen de β -galactosidasa endógeno eliminado o inactivado (es decir, no funcional). Por ejemplo, el gen de la β -galactosidasa comprende un gen lacZ de E. coli (por ejemplo, el Número de Acceso de GenBank V00296.1 (GI: 41901)). El gen lacZ endógeno de la E. coli se elimina o se inactiva funcionalmente, pero de tal manera que la expresión del gen de la lactosa permeasa (lacY) cadena abajo permanece intacta, es decir, un gen funcional de la lactosa permeasa también está presente en la bacteria. Por eliminado se entiende que una parte o la secuencia de codificación completa está ausente, de manera que no se produce ningún producto genético. Un gen "inactivado" no produce un producto genético que funcione como el gen nativo, natural o endógeno. Por ejemplo, la actividad funcional de un producto génico de β -galactosidasa inactivado se reduce al 10%, 20%, 50% o 100%, 1 vez, 2 veces, 5 veces o 10 veces menos que la actividad funcional del producto génico endógeno, nativo o de origen natural.

El gen de la lactosa permeasa es un gen endógeno de la lactosa permeasa o un gen exógeno de la lactosa permeasa. Por ejemplo, el gen de la lactosa permeasa comprende un gen lacY de E. coli (por ejemplo, Número de Acceso de GenBank V00295.1 (GI:41897)). Muchas bacterias poseen la capacidad inherente de transportar la lactosa desde el medio de crecimiento a la célula, utilizando una proteína de transporte que es un homólogo de la permeasa de la lactosa de E. coli (por ejemplo, como se encuentra en Bacillus licheniformis), o un transportador que es un miembro de la ubicua familia de transporte de azúcar PTS (por ejemplo, como se encuentra en Lactobacillus casei y Lactobacillus rhamnosus). Para bacterias que carecen de una capacidad inherente para transportar lactosa extracelular al citoplasma celular, esta capacidad se confiere por un gen transportador de lactosa exógeno (por ejemplo, lacY de E. coli) proporcionado en construcciones de ADN recombinante, y se suministra en un vector de expresión de plásmido o como genes exógenos integrados en el cromosoma huésped.

Para la producción de oligosacáridos que contienen N-acetilglucosamina, la bacteria comprende, además de la capacidad de producción de UDP-GlcNAc, un gen de UDP-GlcNAc exógeno:Gal α / β -R β 3-N-acetilglucosaminiltransferasa o una variante funcional o su fragmento. Este gen UDP-GlcNAc:Gal α / β -R β 3-N-acetilglucosaminiltransferasa exógeno se obtiene de una cualquiera de diversas fuentes, por ejemplo, el gen LgtA descrito de N. meningitidis (SEQ ID NO: 16 Acceso de la proteína del Genbank AAF42258.1) o N. gonorrhoeae (Acceso de la proteína del Genbank ACF31229.1). Además, se compara un gen de lactosa permeasa funcional. Opcionalmente, un gen de glicosiltransferasa exógeno adicional se coexpresa en la bacteria que comprende una UDP-GlcNAc:Gal α / β -R β 3-N-acetilglucosaminiltransferasa exógena. Por ejemplo, un gen de la β -1,4-galactosiltransferasa se coexpresa con el gen UDP-GlcNAc:Gal α / β -R β 3-N-acetilglucosaminiltransferasa. Este gen de la β -1,4-galactosiltransferasa exógeno se obtiene de una cualquiera de diversas fuentes, por ejemplo, la descrita de N. meningitidis, el gen LgtB (Acceso de la proteína del Genbank AAF42257.1), o de H. pylori, el gen Lex2B (SEQ ID NO: 17 Acceso de la proteína del Genbank NP_207619.1). Opcionalmente, el gen de glicosiltransferasa exógeno adicional coexpresado en la bacteria que comprende un gen de UDP-GlcNAc:Gal α / β -R β 3-N-acetilglucosaminiltransferasa exógena es un gen de β -1,3-galactosiltransferasa, por ejemplo, que se describe en O55:H7 de E. coli, el gen WbgO (SEQ ID NO: 18 Acceso de la proteína del Genbank YP_003500090.1), o de H. pylori, el gen jhp0563 (Acceso de la proteína del Genbank AEZ55696.1). Las variantes y fragmentos funcionales de cualquiera de las enzimas descritas anteriormente también están incluidas en la presente invención.

En una realización, los oligosacáridos que contienen N-acteilglucosamina producidos por los métodos descritos en este documento incluyen Lacto-N-triosa 2 (LNT2), Lacto-N-tetraosa (LNT), Lacto-N-neotetraosa (LNnT), Lacto-N-fucopentaosa I (LNF I), Lacto-N-fucopentaosa II (LNF II), Lacto-N-fucopentaosa III (LNF III), Lacto-N-fucopentaosa V (LNF V), Lacto-N-difucohexaosa I (LDFH I), Lacto-N-difucohexaosa II (LDFH II), y Lacto-N-neodifucohexaosa II (LFNnDFH II).

Para la producción de sialil-oligosacáridos, la bacteria comprende un gen exógeno de sialil-transferasa. Por ejemplo, el gen de sialil-transferasa exógeno codifica α (2,3) sialil-transferasa o el gen de sialil-transferasa exógeno codifica

$\alpha(2,6)$ sialil-transferasa o el gen de sialil-transferasa exógeno codifica $\alpha(2,8)$ sialiltransferasa. Los genes de sialil-transferasa exógenos se obtienen de una cualquiera de varias fuentes, por ejemplo, las descritas de *N. meningitidis*, *N. gonorrhoeae*, y de varios organismos del género *Photobacterium*. Ejemplos de $\alpha(2,8)$ sialiltransferasas, útiles para la producción de ácido polisialico, por ejemplo, se encuentran en *Campylobacter jejuni* (CstII: ADN52706) y *Neisseria meningitidis* (o siaD: AAA20478).

Las bacterias utilizadas en el presente documento para producir hMOS están modificadas por ingeniería genética para comprender un aumento de la acumulación de lactosa intracelular (en comparación con el tipo natural) y para comprender la actividad de UDP-GlcNAc:Gal α / β -R β 3-N-acetilglucosaminiltransferasa y/o sialil-transferasa. Opcionalmente, también comprenden la actividad de β -1,4-galactosiltransferasa o β -1,3-galactosiltransferasa, y/o la actividad de α -1,2-, α -1,3- y/o α -1,4-fucosiltransferasa. En algunos casos, la bacteria comprende además un gen funcional de *E. coli* lacZ⁺ de tipo natural insertado en un gen endógeno, por ejemplo, el gen lon en *E. coli* o el gen thyA en *E. coli*. De esta manera, la bacteria comprende además una mutación en un gen lon o una mutación en el gen thyA. En estos casos, el gen lacZ endógeno de la *E. coli* se elimina o se desactiva funcionalmente, pero de tal manera que la expresión del gen de la lactosa permeasa (lacY) cadena abajo permanece intacta. El organismo así manipulado mantiene la capacidad de transportar la lactosa desde el medio de crecimiento y de desarrollar una acumulación de lactosa intracelular para usar como un aceptor de azúcar en la síntesis de oligosacáridos, al tiempo que mantiene un bajo nivel de actividad de beta-galactosidasa intracelular útil para una variedad de fines. Por ejemplo, la invención también incluye: a) métodos para el marcado fenotípico de un locus genético en una célula huésped con β -galactosidasa negativa utilizando un inserto del gen de β -galactosidasa (p. ej., LacZ) diseñado para producir un nivel bajo, pero fácilmente detectable, de la actividad de la β -galactosidasa, b) métodos para detectar fácilmente la contaminación por bacteriófagos líticos en las series de fermentación mediante la liberación y detección de β -galactosidasa citoplásmica en el medio de cultivo celular, y c) métodos para agotar un cultivo bacteriano de lactosa residual al final de las series de producción. a), b) y c) se logran utilizando un inserto génico de β -galactosidasa funcional (por ejemplo, lacZ) cuidadosamente diseñado para dirigir la expresión de un nivel bajo, pero detectable, de la actividad de β -galactosidasa en una célula huésped de β -galactosidasa negativa. La bacteria opcionalmente comprende además una mutación en un gen lacA. Preferiblemente, la bacteria acumula un aumento de la acumulación de lactosa intracelular y produce un bajo nivel de beta-galactosidasa. Una acumulación intracelular aumentada es cuando la concentración de lactosa en la bacteria huésped es al menos 10%, 20%, 50%, 2 veces, 5 veces o 10 veces más alta que la de la bacteria nativa o de origen natural.

En un aspecto, el oligosacárido de la leche humana producido por bacterias modificadas genéticamente comprende una molécula de ácido nucleico exógena que codifica una UDP-GlcNAc:Gal α / β -R β 3-N-acetilglucosaminiltransferasa y un ácido nucleico exógeno que codifica β -1,4-galactosiltransferasa es lacto-N-neotetraosa (LNnT). En otro aspecto, el oligosacárido de la leche humana producido por bacterias modificadas genéticamente comprende una molécula de ácido nucleico exógena que codifica una UDP-GlcNAc:Gal α / β -R β 3-N-acetilglucosaminiltransferasa y un ácido nucleico exógeno que codifica β -1,3-galactosiltransferasa es lacto-N-tetraosa (LNT).

En este documento se describen composiciones que comprenden una célula bacteriana que produce el oligosacárido de la leche humana LNnT (lacto-N-neotetraosa), en donde la célula bacteriana comprende una UDP-GlcNAc exógena:Gal α / β -R β 3-N-acetilglucosaminiltransferasa y un ácido nucleico exógeno que codifica una β -1,4-galactosiltransferasa. La célula bacteriana puede ser *E. coli*. El gen UDP-GlcNAc:Gal α / β -R β 3-N-acetilglucosaminiltransferasa exógeno se obtiene de una cualquiera de varias fuentes, por ejemplo, el gen LgtA descrito de *N. meningitidis*. El gen de la β -1,4-galactosiltransferasa exógeno se obtiene de una cualquiera de varias fuentes, por ejemplo, la descrita de *N. meningitidis*, el gen LgtB, o de *H. pylori*, el gen jhp0765.

La bacteria comprende una producción incrementada de UDP-GlcNAc. El medio para lograr esto es mediante la sobreexpresión de un regulador endógeno positivo de la síntesis de UDP-GlcNAc; es decir, la sobreexpresión del gen nagC de *Escherichia coli*. En un aspecto, esta sobreexpresión de nagC se logra proporcionando copias adicionales del gen nagC en un vector plasmídico o integrando copias adicionales del gen nagC en el cromosoma de la célula huésped. Alternativamente, la sobreexpresión se logra modulando la fuerza de la secuencia de unión al ribosoma que dirige la traducción nagC o modulando la fuerza de la transcripción nagC que dirige el promotor. Además, la acumulación intracelular de UDP-GlcNAc puede mejorarse por otros medios, por ejemplo, por la sobreexpresión del gen glmS de *Escherichia coli* (L-glutamina:D-fructosa-6-fosfato aminotransferasa), o alternativamente por la sobreexpresión del gen glmY de *Escherichia coli* (un regulador de traducción positivo de glmS), o alternativamente mediante la sobreexpresión del gen glmZ de *Escherichia coli* (otro regulador de traducción positivo de glmS), o alternativamente utilizando simultáneamente una combinación de enfoques. En una realización preferida, por ejemplo, los genes nagC (SEQ ID NO: 19 Acceso de la proteína del Genbank BAA35319.1) y glmS (SEQ ID NO: 20 Acceso de la proteína del Genbank NP_418185.1) que codifican las secuencias proporcionadas en el presente documento se sobreexpresan simultáneamente en la misma célula huésped para aumentar la acumulación intracelular de UDP-GlcNAc. Otros componentes del metabolismo de UDP-GlcNAc incluyen: (GlcNAc-1-P) N-acetilglucosamina-1-fosfato; (GlcN-1-P) glucosamina-1-fosfato; (GlcN-6-P) glucosamina-6-fosfato; (GlcNAc-6-P) N-acetilglucosamina-6-fosfato; y (Fruc-6-p) fructosa-6-fosfato. Las bacterias que comprenden las características descritas en el presente documento se cultivan en presencia de lactosa, y se recupera la lacto-N-neotetraosa, ya sea de la propia bacteria (es decir, mediante lisis) o a partir de un sobrenadante de cultivo de la bacteria.

También dentro de la invención se encuentra una bacteria aislada de *E. coli* como se describe anteriormente y se caracteriza porque comprende un gen de β -galactosidasa endógeno eliminado o inactivado, un gen *lacA* inactivado o eliminado, y un gen funcional de lactosa permeasa (*lacY*).

5 En el presente documento también se describen composiciones que comprenden una célula bacteriana que produce el oligosacárido de la leche humana 6'-SL (6'-sialil-lactosa), en donde la célula bacteriana comprende un gen de sialil-transferasa exógeno que codifica $\alpha(2,6)$ sialil-transferasa. La célula bacteriana puede ser *E. coli*. El gen de sialil-transferasa exógeno utilizado para la producción de 6'-SL se obtiene de una cualquiera de varias fuentes, por ejemplo, las descritas de diversos organismos del género *Photobacterium*. En otro aspecto más, el oligosacárido de la leche humana producido por bacterias modificadas genéticamente que comprende una molécula de ácido nucleico exógena que codifica una $\alpha(2,3)$ sialiltransferasa es el 3'-SL (3'-sialil-lactosa). El gen de la sialiltransferasa exógeno utilizado para la producción de 3'-SL se obtiene de una cualquiera de varias fuentes, por ejemplo, las descritas a partir de *N. meningitidis* y *N. gonorrhoeae*.

15 Además, la bacteria contiene una vía catabólica deficiente en ácido siálico. Por "vía catabólica del ácido siálico" se entiende una secuencia de reacciones, generalmente controlada y catalizada por enzimas, que resulta en la degradación del ácido siálico. Una vía catabólica del ácido siálico ejemplar en *Escherichia coli* se describe en este documento. En la ruta catabólica del ácido siálico descrita en el presente documento, el ácido siálico (Neu5Ac; ácido N-acetilneuramínico) se degrada por las enzimas NanA (ácido N-acetilneuramínico liasa) y NanK (N-acetilmanosamina quinasa) y NanE (N-acetilmanosamina-6-fosfato epimerasa), todas codificadas en el operón *nanATEK-yhch*, y reprimidas por NanR (<http://ecocyc.org/ECOLI>). Una vía catabólica deficiente de ácido siálico se diseña en *Escherichia coli* por medio de una mutación en *nanA* endógena (N-acetilneuraminato liasa) (por ejemplo, Número de Acceso de GenBank D00067.1 (GI:216588)) y/o genes de *nanK* (N-acetilmanosamina quinasa) (por ejemplo, Número de acceso del GenBank (aminoácido) BAE77265.1 (GI:85676015)), y/o *nanE* (N-acetilmanosamina-6-fosfato epimerasa, GI: 947745). Opcionalmente, el gen *nanT* (transportador de N-acetilneuraminato) también está inactivado o mutado. Otros intermedios del metabolismo del ácido siálico incluyen: (ManNAc-6-P) N-acetilmanosamina-6-fosfato; (GlcNAc-6-P) N-acetilglucosamina-6-fosfato; (GlcN-6-P) glucosamina-6-fosfato; y (Fru-6-p) fructosa-6-fosfato. En algunas realizaciones preferidas, *nanA* está mutada. En otras realizaciones preferidas, *nanA* y *nanK* están mutados, mientras que *nanE* sigue siendo funcional. En otra realización preferida, *nanA* y *nanE* se mutan, mientras que *nanK* no se ha mutado, inactivado o eliminado. Una mutación es uno o más cambios en la secuencia de ácido nucleico que codifica el producto genético de *nanA*, *nanK*, *nanE* y/o *nanT*. Por ejemplo, la mutación puede ser 1, 2, 5, 10, 25, 50 o 100 cambios en la secuencia de ácido nucleico. Por ejemplo, la *nanA*, *nanK*, *nanE* y/o *nanT* está mutada por una mutación nula. Las mutaciones nulas, como se describen en este documento, abarcan sustituciones, adiciones, eliminaciones o inserciones de aminoácidos que causan una pérdida de la función de la enzima (es decir, una actividad reducida o nula) o una pérdida de la enzima (es decir, ningún producto genético). Por eliminado se entiende que la región de codificación se elimina en su totalidad o en parte de manera que no se produce ningún producto genético. Por inactivado se entiende que la secuencia de codificación se ha alterado de tal manera que el producto génico resultante es funcionalmente inactivo o codifica un producto génico con menos del 100%, 80%, 50% o 20% de la actividad del producto genético endógeno nativo o natural. Un gen o proteína "no mutada" no difiere de una secuencia codificante nativa, natural o endógena en 1, 2, 5, 10, 20, 50, 100, 200 ó 500 codones más, o en la correspondiente secuencia de aminoácidos codificada.

40 Además, la bacteria (por ejemplo, *E. coli*) también comprende una capacidad sintética de ácido siálico. Por ejemplo, la bacteria comprende una capacidad sintética de ácido siálico a través de la provisión de una 2-epimerasa UDP-GlcNAc exógena (por ejemplo, *neuC* de *Campylobacter jejuni* (SEQ ID NO: 13, GenBank AAK91727.1; GI:15193223) o equivalente (por ejemplo, *E. coli* S88 *neuC* GenBank YP_002392936.1; GI:218560023), una Neu5Ac sintasa (por ejemplo, *neuB* de *C. jejuni* (SEQ ID NO: 14 AAK91726.1 GenBank GI:15193222) o equivalente, (por ejemplo, *Flavobacterium limnosediminis* sintasa del ácido siálico, GenBank GI:559220424), y/o una CMP-Neu5Ac sintetasa (por ejemplo, *neuA* de *C. jejuni* (SEQ ID NO: 15 GenBank AAK91728.1; GI:15193224) o equivalente, (p. ej. *Vibrio brasiliensis* CMP-ácido siálico sintasa, GenBank GI:493937153). También se describen variantes funcionales y fragmentos.

50 La bacteria de la presente invención, que comprende una capacidad sintética de ácido siálico, tiene una producción mayor de UDP-GlcNAc por la sobreexpresión del regulador endógeno positivo de la síntesis de UDP-GlcNAc, *nagC*. Un medio ejemplar para lograr esto es, p. ej., mediante la sobreexpresión simultánea de los genes *nagC* y *glmS* de *Escherichia coli*. Esta sobreexpresión de *nagC* y *glmS* se logra proporcionando copias adicionales de los genes *nagC* y *glmS* en un vector plasmídico, o integrando copias adicionales de los genes *nagC* y *glmS* en el cromosoma de la célula huésped. Alternativamente, la sobreexpresión se logra mediante la modulación de la fuerza de la secuencia de unión al ribosoma que dirige *nagC* (descrito por Sleight et al, *Nucleic Acids Res.* Mayo 2010; 38 (8): 2624-2636) y/o la traducción de *nagC* y *glmS*, o mediante la modulación de la fuerza del promotor/es que dirige la transcripción de *nagC* y *glmS* (Sleight et al, *Nucleic Acids Res.* Mayo 2010; 38 (8): 2624-2636).

60 Las bacterias que comprenden las características descritas en el presente documento se cultivan en presencia de lactosa y, en el caso en que las células comprendan una $\alpha(2,6)$ sialiltransferasa (por ejemplo, *Photobacterium* spp JT-ISH-224 (SEQ ID NO: 21 Acceso de la proteína del Genbank BAF92026.1), se recupera 6'-sialil-lactosa, ya sea de la propia bacteria o de un sobrenadante de cultivo de la bacteria. En el caso en que las células comprendan una $\alpha(2,3)$ sialiltransferasa, (por ejemplo, *Neisseria meningitidis* 1st (Acceso de la proteína del Genbank NP273962.1) 3'-

sialilactosa se recupera de la propia bacteria (por ejemplo, por lisis de la bacteria) o a partir de un sobrenadante de cultivo de la bacteria.

También, dentro de la invención, se encuentra una bacteria aislado de *E. coli* que comprende (a) un gen de β -galactosidasa endógeno eliminado o inactivado, un gen de sialil-transferasa exógeno, una vía catabólica deficiente de ácido siálico, una capacidad sintética de ácido siálico, un gen de permeasa de lactosa funcional y una capacidad de producción de UDP-GlcNAc aumentada que comprende la sobreexpresión de nagC, de modo que la bacteria produce al menos un 10% más de UDP-GlcNAc que una bacteria nativa, por ejemplo, dicha bacteria comprende además un gen de β -galactosidasa recombinante que proporciona un nivel bajo, pero detectable, de actividad de β -galactosidasa; o (b) un gen UDP-GlcNAc:Gala/ β -R β 3-N-acetilglucosaminiltransferasa exógeno, un gen funcional de la lactosa permeasa y un aumento de la capacidad de producción de UDP-GlcNAc que comprende la sobreexpresión de nagC, de modo que la bacteria produce al menos el 10% más de UDP-GlcNAc que una bacteria nativa.

Se describe un oligosacárido purificado que contiene N-acetilglucosamina o sialilado producido por los métodos descritos anteriormente. Un oligosacárido purificado, por ejemplo, 6'-SL, es uno que es al menos 90%, 95%, 98%, 99% o 100% (p/p) del oligosacárido deseado en peso. La pureza se evalúa mediante cualquier método conocido, por ejemplo, cromatografía de capa fina u otras técnicas electroforéticas o cromatográficas conocidas en la técnica. Se describe un método para purificar un oligosacárido que contiene N-acetilglucosamina o sialilado producido por la bacteria de ingeniería genética descrita anteriormente, método que consiste en separar el oligosacárido sialilado o que contiene N-acetilglucosamina deseado (por ejemplo, 6'-SL) de contaminantes en un extracto o lisado de células bacterianas, o un sobrenadante de cultivo de células bacterianas. Los contaminantes incluyen el ADN bacteriano, las proteínas y los componentes de la pared celular, y los caramelos de azúcar amarillo/marrón algunas veces formados en reacciones químicas espontáneas en el medio de cultivo.

Los oligosacáridos se purifican y se usan en una serie de productos para el consumo humano y animal, como los animales de compañía (perros, gatos) y el ganado (bovino, equino, ovino, caprino o porcino, así como las aves de corral). Por ejemplo, una composición farmacéutica que comprende 6'-sialil-lactosa purificada (6'-SL) y un excipiente es adecuada para la administración oral. Grandes cantidades de 6'-SL se producen en huéspedes bacterianos, por ejemplo, una bacteria de *E. coli* que comprende una sialiltransferasa heteróloga, por ejemplo, una α (2,6) sialiltransferasa heteróloga. Una bacteria de *E. coli* que comprende una acumulación citoplásmica mejorada de cada uno de los siguientes: lactosa y CMP-Neu5Ac, es útil en tales sistemas de producción. En el caso de la lactosa, las vías metabólicas y los genes endógenos de *E. coli* se manipulan de manera que se genere un aumento de las concentraciones citoplásmicas de lactosa, en comparación con los niveles encontrados en *E. coli* de tipo natural. Por ejemplo, las bacterias contienen al menos un 10%, 20%, 50%, 2 x, 5 x, 10 x o más de los niveles en una bacteria de tipo natural correspondiente que carece de las modificaciones genéticas descritas anteriormente. En el caso de CMP-Neu5Ac, los genes de catabolismo Neu5Ac endógenos están inactivados y los genes de biosíntesis de CMP-Neu5Ac exógenos introducidos en *E. coli* dan como resultado la generación de una acumulación citoplásmica de CMP-Neu5Ac que no se encuentra en la bacteria de tipo natural.

Un método para producir una composición farmacéutica que comprende una hMOS purificada se lleva a cabo cultivando la bacteria descrita anteriormente, purificando la hMOS producida por la bacteria y combinando la hMOS con un excipiente o vehículo para producir un suplemento dietético para 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 usando métodos conocidos de terapia clínica.

La invención prevé aumentar, en *E. coli*, la concentración intracelular del nucleótido azúcar uridina difosfato N-acetilglucosamina (UDP-GlcNAc). Esto se logra mediante la sobreexpresión del regulador positivo endógeno bifuncional de la síntesis de UDP-GlcNAc y del represor del catabolismo de glucosamina y N-acetilglucosamina, nagC, opcionalmente de forma simultánea con el gen que codifica L-glutamina:D-fructosa-6-fosfato aminotransferasa, glmS.

La invención también proporciona un aumento de la concentración intracelular de lactosa en *E. coli*, para células cultivadas en presencia de lactosa, mediante el uso de manipulaciones de genes endógenos de *E. coli* implicados en la importación, exportación y catabolismo de la lactosa. En particular, en el presente documento se describen métodos para aumentar los niveles de lactosa intracelular en *E. coli* genéticamente diseñados para producir un oligosacárido de la leche humana mediante la incorporación de una mutación lacA en la *E. coli* modificada genéticamente. La mutación lacA evita 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, lo que facilita enormemente las manipulaciones intencionadas de la acumulación de lactosa intracelular de *E. coli*.

También se describen células huésped bacterianas con la capacidad de acumular un depósito de lactosa intracelular mientras que simultáneamente se poseen niveles bajos y funcionales de actividad citoplásmica de β -galactosidasa, por ejemplo, como lo proporciona la introducción de un gen recombinante funcional de *E. coli* lacZ, o por un gen de β -galactosidasa de cualquiera de una serie de otros organismos (por ejemplo, el gen lac4 de *Kluyveromyces lactis*

(por ejemplo, Número de Acceso de GenBank M84410.1 (GI:173304)). Los niveles bajos y funcionales de β -galactosidasa citoplásmica incluyen niveles de actividad de β -galactosidasa de entre 0,05 y 200 unidades, por ejemplo, entre 0,05 y 5 unidades, entre 0,05 y 4 unidades, entre 0,05 y 3 unidades, o entre 0,05 y 2 unidades (para una definición estándar véase: Miller JH, laboratorio CSH. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.; 1972). Sin embargo, este bajo nivel de actividad de β -galactosidasa citoplásmica, aunque no es lo suficientemente alto como para disminuir significativamente la acumulación de lactosa intracelular, es muy útil para tareas tales como el marcado fenotípico de loci genéticos deseables durante la construcción de fondos de células huésped, para la detección de la lisis celular debida a contaminaciones de bacteriófagos no deseados en los procesos de fermentación, para la eliminación fácil de la lactosa residual no deseada al final de las fermentaciones, o para fines de control de calidad de la fermentación durante el proceso (es decir, como un fenotipo no estándar, la provisión de un fenotipo débil *lacZ* ayuda en las evaluaciones de pureza del cultivo).

Los métodos para purificar un oligosacárido sialilado o que contiene N-acetilglucosamina producido por los métodos de la invención o descritos se llevan a cabo uniendo el oligosacárido de un lisado de células bacterianas o de un sobrenadante de cultivo de células bacterianas de la bacteria a una columna de carbono, y luego eluyendo de la columna. Los oligosacáridos purificados que contienen N-acetilglucosamina o sialilados se producen mediante los métodos de la invención o descritos en el presente documento.

Se describe un vector, por ejemplo, un vector que contiene un ácido nucleico. El vector puede incluir además uno o más elementos reguladores, por ejemplo, un promotor heterólogo. Los elementos reguladores pueden unirse operativamente a un gen de proteína, a un gen de proteína de fusión o a una serie de genes unidos a un operón para expresar la proteína de fusión. Para mantener estable el vector plasmídico dentro de la célula, se incluye un marcador seleccionable dentro de su secuencia, como un gen de resistencia a los antibióticos o un gen que complemente una auxotrofia nutricional de la bacteria huésped. Por ejemplo, en *E. coli*, una deficiencia de timidina causada por un defecto cromosómico en el gen de la timidilato sintasa (*thyA*) puede complementarse con una copia de tipo natural transmitida por el plásmido del gen *thyA* (M. Belfort, GF Maley, F. Maley, Proceedings of the National Academy of Sciences 80, 1858 (1983)). Alternativamente, una deficiencia de adenina causada por una deficiencia cromosómica en el gen de la adenilosuccinato sintetasa (*purA*) (S.A. Wolfe, J. M. Smith, J. Biol Chem 263, 19147-53 (1988)) puede complementarse con una copia de *purA* de tipo natural de plásmido. Se pueden utilizar dos vectores de plásmidos simultáneamente dentro de la misma célula bacteriana empleando marcadores seleccionables separados, por ejemplo, un plásmido que utilice la selección de *thyA* y uno que utilice la selección de *purA*, y utilizando dos replices de plásmidos compatibles, por ejemplo, en *E. coli*. Los dos dichos replices compatibles comprenden el replicón ColE1 (pUC) y el replicón p15A (pACYC) (RE Bird, J Bacteriol 145, 1305-9 (1981)). En otro aspecto más, la invención comprende una célula recombinante aislada, por ejemplo, una célula bacteriana que contiene la(s) molécula(s) de ácido nucleico mencionada(s) o vector(es). Las secuencias de ácido nucleico pueden integrarse opcionalmente en el genoma.

Se describe 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, purificado a partir de un cultivo de una cepa recombinante de la presente invención, en donde el hMOS se une a un patógeno y en el que el sujeto está infectado o en riesgo de infección con el patógeno. La infección puede estar causada por un virus tipo Norwalk o *Campylobacter jejuni*. El sujeto es preferiblemente un mamífero que necesita tal tratamiento. 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. Por ejemplo, las composiciones se formulan en piensos para animales (por ejemplo, peletes, croquetas, puré) o complementos alimenticios para animales de compañía, por ejemplo, perros o gatos, así como ganado o animales criados para el consumo de alimentos, por ejemplo, ganado vacuno, ovejas, cerdos, gallinas y cabras. Preferiblemente, el hMOS purificado puede formularse 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 zumo antes del consumo) o en forma de tabletas, cápsulas o pastas o se incorpora como componente en productos lácteos como la leche, cremas, quesos, yogures 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 destinadas a mejorar el crecimiento de microorganismos benéficos ya sea *in vitro* o *in vivo*. Por ejemplo, el azúcar purificado (por ejemplo, LNT o 6'-SL) se puede mezclar con un *Bifidobacterium* o *Lactobacillus* en una composición nutricional probiótica (es decir, las bifidobacterias son componentes beneficiosos de una flora intestinal normal y también se sabe que utilizan hMOS para el crecimiento).

Todos los genes descritos en el presente documento también incluyen una descripción de los productos génicos codificados correspondientes. Como tales, los usos de genes exógenos como se describen en este documento abarcan ácidos nucleicos que codifican las secuencias de productos génicos descritos en este documento. El experto en la materia podría generar fácilmente secuencias de ácido nucleico que codifiquen las secuencias de proteínas descritas en este documento e introducir dichas secuencias en vectores de expresión para llevar a cabo la presente invención.

La expresión "sustancialmente puro", en referencia a un polipéptido, polinucleótido u oligosacárido dado significa que el polipéptido, polinucleótido u oligosacárido está sustancialmente libre de otras macromoléculas biológicas. El polipéptido, polinucleótido u oligosacárido sustancialmente puro es al menos un 75% (por ejemplo, al menos 80, 85,

95 ó 99%) puro en peso seco. La pureza se puede medir mediante cualquier método estándar calibrado apropiado, por ejemplo, mediante cromatografía en columna, electroforesis en gel de poliacrilamida, cromatografía en capa fina (TLC) o análisis por HPLC.

5 Se describen polinucleótidos, polipéptidos y oligosacáridos que 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. Específicamente, como se usa en este 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 químicamente. Por ejemplo, las composiciones de hMOS purificadas
10 tienen al menos un 60% en peso (peso seco) del compuesto de interés. La preparación puede tener al menos el 75%, al menos el 90%, y al menos el 99% en peso del compuesto de interés. La pureza se mide mediante cualquier método estándar calibrado apropiado, por ejemplo, mediante cromatografía en columna, electroforesis en gel de poliacrilamida, cromatografía en capa fina (TLC) o análisis por 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. La proteína puede constituir al menos 10, 20, 50 70, 80, 90, 95, 99-100% en peso seco de la
15 preparación purificada.

Por "ácido nucleico aislado" se entiende un ácido nucleico que está libre de los genes que lo flanquean en el genoma natural del organismo del cual se deriva el ácido nucleico. El término cubre, por ejemplo: (a) un ADN que forma parte de una molécula de ADN genómico natural, pero no está flanqueado por las dos secuencias de ácido nucleico que flanquean esa parte de la molécula en el genoma del organismo en el que ocurre naturalmente (b) un ácido nucleico incorporado en un vector o en el ADN genómico de un procarionta 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 la 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 incluyen además moléculas producidas sintéticamente, así como cualquier ácido nucleico que se haya alterado químicamente y/o que tenga estructuras modificadas. Por ejemplo, el ácido nucleico aislado es un polinucleótido de ADNc o ARN purificado.
20
25

Un "promotor heterólogo", cuando está unido operativamente a una secuencia de ácido nucleico, se refiere a un promotor que no está asociado naturalmente con la secuencia de ácido nucleico.
30

El término "sobrexpresar", como se usa en este documento, se refiere a que la transcripción del gen o al producto genético codificado tiene 10%, 20%, 50%, 2 veces, 5 veces, 10 veces o más que el nivel expresado o producido por un gen nativo, natural o endógeno en una bacteria en la que ocurre naturalmente. Por ejemplo, la bacteria huésped descrita en el presente documento está diseñada para sobreexpresar una transcripción génica exógena o un producto génico codificado de UDP-GlcNAc:Gal α / β -R β 3-N-acetilglucosaminiltransferasa, nagC, glmS, glmY, glmZ, una sialil-transferasa, una β -galactosiltransferasa, una α -fucosiltransferasa, CMP-Neu5Ac sintetasa, una ácido sialico sintasa o una UDP-GlcNAc 2-epimerasa, es decir, un gen o producto génico con una secuencia correspondiente a la de una bacteria distinta de la bacteria huésped.
35

Los términos "tratar" y "tratamiento", como se usan en este documento, se refieren a la administración de un agente o formulación a un individuo clínicamente sintomático afectado por una afección, trastorno o enfermedad adversos, a fin de lograr una reducción en la severidad y/o frecuencia de los síntomas, eliminar los síntomas y/o su causa subyacente, y/o facilitar la mejora o la reparación del daño. Los términos "prevenir" y "prevención" se refieren a la administración de un agente o composición a un individuo clínicamente asintomático que es susceptible a una condición, trastorno o enfermedad adversa en particular y, por lo tanto, se relaciona con la prevención de la aparición de síntomas y/o o su causa subyacente.
40
45

Por las expresiones "cantidad eficaz" y "cantidad terapéuticamente eficaz" de una formulación o componente de formulación se entiende una cantidad no tóxica pero suficiente de la formulación o componente para proporcionar el efecto deseado.

La expresión de transición "que comprende", que es sinónimo de "que incluye", "que contiene" o "caracterizado por", es inclusiva o abierta y no excluye elementos o pasos de métodos no citados adicionales. Por el contrario, la frase de transición "que consiste en" excluye cualquier elemento, paso o ingrediente no especificado en la reivindicación. La frase de transición "que consiste esencialmente en" limita el alcance de una reivindicación a los materiales o pasos especificados "y aquellos que no afectan materialmente a las características básicas y novedosas" de la invención reivindicada.
50

Otras características y ventajas de la invención serán evidentes a partir de la siguiente descripción y de las reivindicaciones. A menos que se defina lo contrario, todos los términos técnicos y científicos utilizados en este documento tienen el mismo significado que entiende comúnmente cualquier experto en la técnica a la que pertenece esta invención. Aunque se puedan usar métodos y materiales similares o equivalentes a los descritos en el presente documento en la práctica o en el ensayo de la presente invención, a continuación, se describen métodos y
55

materiales adecuados. En caso de conflicto con todas las patentes y solicitudes de patentes extranjeras publicadas citadas en este documento, las presentaciones de Genbank y NCBI indicadas por el número de acceso citado en este documento y la literatura científica publicada en este documento, la presente memoria descriptiva, incluidas las definiciones, prevalecerá. Además, los materiales, métodos y ejemplos son solo ilustrativos y no pretenden ser limitativos.

Breve descripción de los dibujos

Fig. 1 es un esquema que muestra las rutas metabólicas y los cambios introducidos en ellas para diseñar la síntesis de 2'-fucosil-lactosa (2'-FL) en *Escherichia coli* (*E. coli*). Específicamente, se ilustran la ruta de la síntesis de lactosa y la ruta de la síntesis de GDP-fucosa. En la vía de la síntesis de GDP-fucosa: manA = fosfomanosa isomerasa (PMI), manB = fosfomanomutasa (PMM), manC = manosa-1-fosfato guanililtransferasa (GMP), gmd = GDP-manosa-4,6-deshidratasa, fcl = GDP-fucosa sintasa (GFS), y Δ wcaJ = UDP-glucosa transferida portadora lipídica de glucosa.

Fig. 2 es un esquema que muestra las rutas metabólicas involucradas en la síntesis de UDP-GlcNAc (uridina difosfato N-acetilglucosamina) y catabolismo de la glucosamina y N-acetilglucosamina en *E. coli*. En el esquema: (GlcNAc-1-P) N-acetilglucosamina-1-fosfato; (GlcN-1-P) glucosamina-1-fosfato; (GlcN-6-P) glucosamina-6-fosfato; (GlcNAc-6-P) N-acetilglucosamina-6-fosfato; y (fruc-6-P) fructosa-6-fosfato; glmS (L-glutamina: D-fructosa-6-fosfato aminotransferasa), glmM (fosfoglucoamina mutasa), glmU (N-acetil glucosamina-1-fosfato uridiltransferasa y glucosamina-1-fosfato acetiltransferasa fusionada), nagC (activador transactivo bifuncional/proteína represora), nagA (N-acetilglucosamina-6-fosfato desacetilasa) y nagB (glucosamina-6-fosfato desaminasa), nagE (transportador de N-acetilglucosamina) y manXYZ (transportador de glucosamina).

Fig. 3 es un esquema que muestra las rutas metabólicas y un ejemplo (que utiliza deleciones nanT, nanA y nanK) de los cambios introducidos en ellos para diseñar la síntesis de 6'-sialil-lactosa (6'-SL) en *E. coli*. Las abreviaturas incluyen: (Neu5Ac) ácido N-acetilneuramínico, ácido siálico; (Δ nanT) transportador del ácido N-acetilneuramínico mutado; (Δ nanA) liasa del ácido N-acetilneuramínico mutada; (ManNAc) N-acetilmanosamina; (Δ nanK) N-acetilmanosamina quinasa mutada; (nanE) epimerasa de N-acetilmanosamina-6-fosfato de tipo natural; (ManNAc-6-P) N-acetilmanosamina-6-fosfato; (GlcNAc-6-P) N-acetilglucosamina-6-fosfato; (GlcN-6-P) glucosamina-6-fosfato; (Fruc-6-P) Fructosa-6-fosfato; (neuA), sintetasa del ácido CMP-N-acetilneuramínico; (CMP-Neu5Ac) ácido CMP-N-acetilneuramínico; (neuB), sintasa del ácido N-acetilneuramínico; (neuC) UDP-GlcNAc-2-epimerasa; y (UDP-GlcNAc) uridina difosfato N-acetilglucosamina.

Fig. 4 es un esquema que ilustra la nueva configuración de genes diseñados en el locus thyA de *Escherichia coli* en cepas utilizadas para producir oligosacáridos que contienen N-acetilglucosamina.

Fig. 5 es un mapa de plásmido de pG292, que expresa el gen de la β (1,3)-N-acetilglucosaminiltransferasa de *N. meningitidis* lgtA.

Fig. 6 es un mapa plasmídico de pG221, que expresa, como un operón, el gen lgtA de β (1,3)-N-acetilglucosaminiltransferasa de *N. meningitidis* y el gen de *E. coli* O55:H7 wbgO β (1,3)-galactosiltransferasa.

Fig. 7 es un mapa plasmídico de pG222, que expresa, como un operón, el gen lgtA de *N. meningitidis* β (1,3)-N-acetilglucosaminiltransferasa y el gen 4GalT (jhp0765) β (1,4)-galactosiltransferasa de *H. pylori*.

Fig. 8 ilustra esquemáticamente las reacciones enzimáticas necesarias para producir a partir de la lactosa, a través del trisacárido intermedio lacto-N-triosa 2 (LNT2), los dos oligosacáridos de la leche humana: Lacto-N-tetraosa (LNT) y Lacto-N-neotetraosa (LNnT). Se presenta un cromatograma de capa fina (a la izquierda) de muestras de medio de cultivo tomadas de cultivos de *E. coli* a pequeña escala y que demuestran la síntesis de LNT2, LNT y LNnT. Se presenta un segundo cromatograma de capa fina (a la derecha) de muestras de medio de cultivo tomadas de un cultivo de biorreactor de *E. coli* de 15 L que demuestra la síntesis de LNnT.

Fig. 9 es un mapa plasmídico de pG317, un vector de baja copia que se expresa como un operón, bajo el control del promotor lac de *E. coli*, los genes neuB, neuC y neuA del *Campylobacter jejuni* ATCC43438, que codifican la N-acetilneuraminato sintasa, UDP-N-acetilglucosamina 2-epimerasa, y N-acetilneuraminato citidililtransferasa, respectivamente.

Fig. 10 es un mapa plasmídico de pG315, un vector de copias múltiples que expresa un gen que codifica una sialiltransferasa α (2,6) de *Photobacterium* spp JT-ISH-224, bajo el control del promotor lac de *E. coli*.

Fig. 11 es una fotografía de un cromatograma de capa fina que muestra 6'-SL en medio de cultivo producido por la cepa E547 de *E. coli* (AnanRATEK), que contiene plásmidos que expresan una α (2,3) sialiltransferasa bacteriana y neuA, neuB y neuC. La Fig. 11 también muestra un análisis de TLC de sobrenadantes de cultivos de dos fermentaciones que producen 6'-sialilactosa (6'-SL). Las muestras a la izquierda de la figura se toman de una fermentación de una cepa de *E. coli* que contiene pG315 (que lleva un RBS fuerte frente al gen de la α (2,6) sialiltransferasa en el vector). Las muestras a la derecha de la figura se toman de una fermentación de una cepa de *E. coli* que contiene una variante cercana de pG315 que porta un RBS más débil frente al gen de la α (2,6) sialiltransferasa.

Fig. 12 es un mapa plasmídico de pG345, un vector de copias múltiples que expresa un gen que codifica una $\alpha(2,6)$ sialiltransferasa de *Photobacterium* spp JT-ISH-224, bajo el control de un sitio de unión ribosómico más débil (SEQ ID NO: 8) y el promotor lac de *E. coli*.

Fig. 13 es un esquema que muestra las rutas metabólicas y un segundo ejemplo (que utiliza delecciones nanT, nanA y nanE) de los cambios introducidos en ellos para diseñar la síntesis de 6'-sialil-lactosa (6'-SL) en *E. coli*. Las abreviaturas incluyen: (Neu5Ac) ácido N-acetilneuramínico, ácido siálico; (Δ nanT) transportador del ácido N-acetilneuramínico mutado; (Δ nanA) liasa del ácido N-acetilneuramínico mutada; (ManNAc) N-acetilmanosamina; (nanK) N-acetilmanosamina quinasa de tipo natural; (Δ nanE) N-acetilmanosamina-6-fosfato epimerasa mutada; (ManNAc-6-P) N-acetilmanosamina-6-fosfato; (GlcNAc-6-P) N-acetilglucosamina-6-fosfato; (GlcN-6-P) glucosamina-6-fosfato; (Fruc-6-P) Fructosa-6-fosfato; (neuA), sintetasa del ácido CMP-N-acetilneuramínico; (CMP-Neu5Ac) ácido CMP-N-acetilneuramínico; (neuB), sintasa del ácido N-acetilneuramínico; (neuC) UDP-GlcNAc-2-epimerasa; y (UDP-GlcNAc) uridina difosfato N-acetilglucosamina.

Fig. 14 ilustra el análisis de TLC de sedimentos celulares y/o sobrenadantes de tres experimentos de fermentación a escala piloto que usan tres cepas de *E. coli* que llevan varias combinaciones de mutaciones nan.

Fig. 15 es un esquema que ilustra la ubicación de la eliminación del gen realizada dentro del operón nan de *E. coli* para generar el locus mutante [nanR⁺, nanA, nanT, nanE, nanK⁺] de las cepas E1017 y E1018.

Fig. 16 es una gráfica de la curva de crecimiento de la densidad celular de cuatro cultivos de E680 transformados con pG292, inducidos o no inducidos por la adición de triptófano, y en presencia o ausencia de lactosa en el medio de crecimiento. Se observa abundante lisis celular en los cultivos que contienen lactosa.

Fig. 17 es un mapa plasmídico de pG356, que expresa, como un operón, los genes glmS y nagC de *E. coli*. pG356 tiene un origen de replicación p15A y marcadores seleccionables tanto ampC como purA.

Fig. 18 es una traza de parámetros de fermentación y un análisis del sobrenadante del cultivo de TLC (para la producción de LNnT) de un cultivo de biorreactor de 1,5 l de E796 transformado con pG222.

Fig. 19 es una traza de parámetros de fermentación y un análisis del sobrenadante del cultivo de TLC (para la producción de LNnT) de un cultivo de biorreactor de 1,5 l de E866 transformado con pG222 y pG356.

Descripción detallada de la invención

En este documento se describen construcciones genéticas y métodos para la producción de oligosacáridos de leche humana (hMOS) que contienen N-acetilglucosamina y sialiloligosacáridos. Con el fin de producir tanto hMOS que contengan N-acetilglucosamina como que contengan sialilo, se necesita aprovechar la acumulación celular de UDP-GlcNAc. Hacerlo puede ser difícil, ya que UDP-GlcNAc es un metabolito esencial para las bacterias (usado para fabricar la pared celular). Las construcciones, composiciones y métodos de la invención o como se describen superan las dificultades del pasado al mejorar la acumulación de UDP-GlcNAc, una estrategia que representa una ventaja en la producción de ambas clases de hMOS. Otras distinciones sobre los enfoques anteriores representan mejoras y/o confieren ventajas sobre las estrategias anteriores.

hMOS

Los glicanos de la leche humana, que comprenden tanto oligosacáridos (hMOS) como sus glicoconjugados, desempeñan un papel importante en la protección y el desarrollo de los lactantes humanos y, en particular, en el tracto gastrointestinal (GI) infantil. Los oligosacáridos de leche que se encuentran en varios mamíferos difieren mucho, y su composición en 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 glucano 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). Anteriormente, una exploración completa de los roles de hMOS estaba limitada por la incapacidad de caracterizar y medir adecuadamente estos compuestos. En los últimos años se han desarrollado métodos cuantitativos sensibles y reproducibles para el análisis de hMOS neutros y ácidos (Erney, R., Hilty, M., Pickering, L., Ruiz-Palacios, G., y Prieto, P. (2001) *Adv Exp Med Biol* 501, 285-297. Bao, Y., y Newburg, D. S. (2008) *Electrophoresis* 29, 2508-2515). Se han identificado aproximadamente 200 oligosacáridos distintos en la leche humana, y las combinaciones de un pequeño número de epítomos simples son responsables de esta diversidad (Newburg DS, 1999 *Curr Med Chem*, 6: 117-127; Ninonuevo M. et al., 2006 *J Agric Food Chem*, 54: 7471-74801). Los hMOS se componen de 5 monosacáridos: D-glucosa (Glc), D-galactosa (Gal), N-acetilglucosamina (GlcNAc), L-fucosa (Fuc) y ácido siálico (ácido N-acetil neuramínico, Neu5Ac, NANA). Los hMOS 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 β 1-4Glc), y compuestos ácidos que incluyen los mismos azúcares y, a menudo, las mismas estructuras del núcleo, más NANA (Charlwood J. y otros, 1999 *Anal Biochem*, 273: 261-277; Martin-Sosa y otros, 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 el 70-80% de los oligosacáridos en la leche humana están fucosilados. Una proporción más pequeña de los oligosacáridos en la leche humana está sialilada, o está fucosilada y sialilada.

Curiosamente, los hMOS, como clase, sobreviven de manera muy eficiente a través del intestino de los bebés, en función de su transporte deficiente a través de la pared intestinal y de su resistencia a la digestión por las enzimas intestinales humanas (Chaturvedi, P., Warren, CD, Buescher, CR, Pickering, LK y Newburg, DS *Adv Exp Med Biol* 501, 315-323 (2001)). Una consecuencia de esta supervivencia en el intestino es que los hMOS pueden funcionar como prebióticos, es decir, están disponibles para servir como una fuente de carbono abundante para el crecimiento de microorganismos comensales del intestino residentes (Ward, RE, Niñonuevo, M., Mills, DA, Lebrilla, CB y German, JB (2007) *Mol Nutr Food Res* 51, 1398-1405). Recientemente, hay un creciente interés en el papel de la dieta y los agentes prebióticos en la determinación de la composición de la microflora intestinal, y en la comprensión del vínculo entre la microflora intestinal y la salud humana (Roberfroid, M., Gibson, GR, Hoyles, L., McCartney, AL, Rastall, R., Rowland, I., Wolvers, D., Watzl, B., Szajewska, H., Stahl, B., Guarner, F., Respondek, F., Whelan, K., Coxam, V., Davicco, MJ, Léotoing, L., Wittrant, Y., Delzenne, NM, Cani, PD, Neyrinck, AM y Meheust, A. (2010) *Br J Nutr* 104 Suppl 2, S1-63).

Varios glicanos de la leche humana poseen homología estructural con los receptores celulares para los enteropatógenos, y cumplen funciones en la defensa de los patógenos al actuar como "señuelos" de los receptores moleculares. Por ejemplo, las cepas patógenas de *Campylobacter* se unen específicamente a los glucanos en la leche humana que contiene el epítipo H-2, es decir, 2'-fucosil-N-acetilactosamina o 2'-fucosillactosa (2'-FL); la unión de *Campylobacter* y la infectividad son inhibidas por 2'-FL y otros glicanos que contienen este epítipo H-2 (Ruiz-Palacios, GM, Cervantes, LE, Ramos, P., Chavez-Munguia, B., y Newburg, DS (2003) *J Biol Chem* 278, 14112-14120). De manera similar, algunos patógenos diarreogénicos de *E. coli* son fuertemente inhibidos in vivo por hMOS que contiene restos de fucosa unidos por 2'. Varias cepas principales de calicivirus humanos, especialmente los norovirus, también se unen a los glucanos fucosilados unidos por 2', y esta unión es inhibida por los glucanos fucosilados unidos por enlaces 2' de la leche humana. El consumo de leche humana con altos niveles de estos fucosiloligosacáridos ligados en 2' se ha asociado con un menor riesgo de norovirus, *Campylobacter*, ST de diarrea asociada a *E. coli* y diarrea de moderada a grave de todas las causas en una cohorte mexicana de niños lactantes (Newburg DS et al., 2004 *Glycobiology*, 14: 253-263; Newburg DS et al., 1998 *Lancet*, 351: 1160-1164).

También se sabe que varios agentes patógenos utilizan glicanos sialilados como sus receptores anfitriones, como la influenza (Couceiro, JN, Paulson, JC & Baum, LG *Virus Res* 29, 155-165 (1993)), parainfluenza (Amonsén, M., Smith, DF, Cummings, RD y Air, GM *J Virol* 81, 8341-8345 (2007)) y rotovirus (Kuhlenschmidt, TB, Hanafin, WP, Gelberg, HB y Kuhlenschmidt, MS *Adv Exp Med Biol* 473, 309-317 (1999)). El epítipo sialil-Lewis X es usado por *Helicobacter pylori* (Mandavi, J., Sondén, B., Hurtig, M., Olfat, FO, y col. *Science* 297, 573-578 (2002)), *Pseudomonas aeruginosa* (Scharfman, A., Delmotte, P., Beau, J., Lamblin, G., y otros, *Glycoconj J* 17, 735-740 (2000)), y algunas cepas de norovirus (Rydell, GE, Nilsson, J., Rodríguez-Díaz, J., Ruvoën-Clouet, N., et al. *Glycobiology* 19, 309-320 (2009)).

El nucleótido azúcar uridina difosfato N-acetilglucosamina (UDP-GlcNAc) es un intermediario metabólico clave en las bacterias, donde participa en la síntesis y el mantenimiento de la envoltura celular. En todas las clases bacterianas conocidas, UDP-GlcNAc se usa para elaborar peptidoglicanos (mureína); un polímero que comprende la pared celular bacteriana cuya integridad estructural es absolutamente esencial para el crecimiento y la supervivencia. Además, las bacterias gramnegativas utilizan UDP-GlcNAc para la síntesis de lípido A, un componente importante de la membrana celular externa. Por lo tanto, para las bacterias, la capacidad de mantener una acumulación intracelular adecuada de UDP-GlcNAc es crítica.

La biosíntesis de ciertos oligosacáridos de la leche humana (hMOS) se ha logrado en cepas modificadas por ingeniería genética de la bacteria, *Escherichia coli* K12. Como se describe en este documento, hMOS fucosilados simples, p. ej. la 2'-fucosillactosa (2'-FL), la 3-fucosillactosa (3-FL) y la lactodifucotetraosa (LDFT) se producen de manera eficiente por la *E. coli* viva mediante la mejora artificial de las reservas intracelulares existentes de GDP-fucosa (el nucleótido donante de azúcar) y la lactosa (el azúcar aceptante), y para usar estas reservas mejoradas como sustratos para fucosiltransferasas recombinantes heterólogas (Figura 1). Dado que ni las reservas de lactosa ni de GDP-fucosa son esenciales para la supervivencia de *E. coli*, la biosíntesis de hMOS fucosilado simple se logra con buenos rendimientos sin consecuencias negativas sobre el crecimiento o la viabilidad de la bacteria huésped. Sin embargo, para sintetizar hMOS más complejos en *E. coli*, se requiere el uso de la reserva bacteriana crítica de UDP-GlcNAc, con los consiguientes impactos potenciales en la viabilidad celular.

La reserva de UDP-GlcNAc en *E. coli* se produce a través de la acción combinada de tres genes *glmS*, *glmS* (L-glutamina:D-fructosa-6-fosfato aminotransferasa), *glmM* (fosfoglucosamina mutasa) y el *glmU* bifuncional (N-acetilglucosamina-1-fosfato uridiltransferasa y glucosamina-1-fosfato acetil transferasa fusionadas) (Figura 2). Estos tres genes dirigen un flujo constante de carbono a UDP-GlcNAc, un flujo que se origina con fructosa-6-fosfato (una molécula abundante del metabolismo central de energía). La expresión de los genes *glm* está bajo control positivo por la proteína activadora de la transcripción, *NagC*.

Cuando *E. coli* encuentra glucosamina o N-acetil-glucosamina en su ambiente, estas moléculas se transportan a la célula a través de proteínas de transporte de membrana específicas y se usan para complementar el flujo de carbono a la reserva de UDP-GlcNAc o, alternativamente, son consumidos para generar energía, bajo la acción de los productos génicos del operón *nag* (es decir, *nagA* [N-acetilglucosamina-6-fosfato desacetilasa] y *nagB* [glucosamina-6-fosfato desaminasa]). En contraste con los genes *glm*, la expresión de *nagA* y *nagB* están bajo el

control transcripcional negativo, pero por la misma proteína reguladora que los genes *glm*, es decir, NagC. NagC es, por lo tanto, bifuncional, capaz de activar la síntesis de UDP-GlcNAc, mientras que al mismo tiempo reprime la degradación de la glucosamina-6-fosfato y la N-acetilglucosamina-6-fosfato.

5 La unión de NagC a secuencias de ADN reguladoras específicas (operadores), ya sea que dicha unión resulte en la activación o represión de genes, es sensible a las fluctuaciones en el nivel citoplásmico del inductor y metabolito de molécula pequeña, GlcNAc-6-fosfato. Las concentraciones intracelulares de GlcNAc-6-fosfato aumentan cuando la N-acetilglucosamina está disponible como una fuente de carbono en el medio ambiente y, por lo tanto, en estas condiciones, la expresión de los genes *glm* (esenciales para mantener la reserva vital de UDP-GlcNAc) disminuiría, a menos que se ponga en juego un mecanismo compensatorio. *E. coli* mantiene un nivel básico de síntesis de UDP-GlcNAc a través de la expresión continua de *nagC* dirigida por dos promotores constitutivos, ubicados dentro del gen *nagA* en sentido ascendente. Este nivel constitutivo de la expresión de *nagC* se complementa aproximadamente tres veces en condiciones en las que se induce el operón degradante de *nag* y, de este modo, *E. coli* garantiza un nivel adecuado de expresión del gen *glm* en todas las condiciones, incluso cuando se utiliza N-acetilglucosamina como fuente de carbono.

15 Muchos hMOS incorporan GlcNAc en sus estructuras directamente, y muchos también incorporan ácido siálico, un azúcar cuya síntesis implica el consumo de UDP-GlcNAc (FIG. 3, FIG. 13). Por lo tanto, la síntesis de muchos tipos de hMOS en *E. coli* diseñada por ingeniería genética conlleva el riesgo significativo de una reducción del rendimiento del producto y de la viabilidad celular comprometida que resulta del agotamiento de la reserva de UDP-GlcNAc de la bacteria. Una forma de abordar este problema durante la síntesis por ingeniería de hMOS que contiene GlcNAc o ácido siálico es aumentar la reserva de UDP-GlcNAc mediante la sobreexpresión simultánea de *nagC*, o preferiblemente mediante la sobreexpresión simultánea de *nagC* y *glmS*.

25 Si bien los estudios sugieren que los glicanos de la leche humana podrían usarse como prebióticos y como agentes antimicrobianos antiadherentes, la dificultad y el costo de producir cantidades adecuadas de estos agentes de una calidad adecuada para el consumo humano ha limitado sus pruebas a gran escala y su utilidad percibida. Lo que se ha necesitado es un método adecuado para producir los glicanos apropiados en cantidades suficientes a un costo razonable. Antes de la invención, hubo intentos de usar varios enfoques sintéticos distintos para la síntesis de glicanos. Los nuevos enfoques químicos pueden sintetizar oligosacáridos (Flowers, *HM Methods Enzymol* 50, 93-121 (1978); Seeberger, *PH Chem Commun (Camb)* 1115-1121 (2003)), pero los reactivos para estos métodos son caros y potencialmente tóxicos (Koeller, KM & Wong, *CH Chem Rev* 100, 4465-4494 (2000)). Enzimas expresadas a partir de organismos modificados (Albermann, C., Piepersberg, W. y Wehmeier, *UF Carbohydr Res* 334, 97-103 (2001); Bettler, E., Samain, E., Chazalet, V., Bosso, C., y otros, *Glycoconj J* 16, 205-212 (1999); Johnson, KF *Glycoconj J* 16, 141-146 (1999); Palcic, *MM Curr Opin Biotechnol* 10, 616-624 (1999); Wymer, N. & Toone, *EJ Curr Opin Chem Biol* 4, 110-119 (2000) proporcionan una síntesis precisa y eficiente (Palcic, *MM Curr Opin Biotechnol* 10, 616-624 (1999)); Crout, D. H. y Vic, G. *Curr Opin Chem Biol* 2, 98-111 (1998)), pero el alto costo de los reactivos, especialmente los nucleótidos de azúcar, limita su utilidad para la producción a gran escala y bajo costo. Los microbios han sido diseñados genéticamente para expresar las glicosiltransferasas necesarias para sintetizar los oligosacáridos de la reserva innata de bacterias de los azúcares de nucleótidos (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, los bajos rendimientos generales del producto y la alta complejidad del proceso han limitado la utilidad comercial de estos enfoques.

45 Antes de la invención, que permite la producción barata de grandes cantidades de hMOS neutros y ácidos, no había sido posible investigar completamente la capacidad de esta clase de molécula para inhibir la unión de patógenos, o incluso explorar su gama completa de potenciales funciones adicionales.

50 Antes de la invención, las síntesis químicas de hMOS eran posibles, pero estaban limitadas por problemas de estereo-especificidad, disponibilidad de precursores, impurezas del producto y alto costo general (Flowers, *HM Methods Enzymol* 50, 93-121 (1978); Seeberger, *PH Chem Commun (Camb)* 1115-1121 (2003); Koeller, KM & Wong, *CH Chem Rev* 100, 4465-4494 (2000)). Además, antes de la invención, también eran posibles las síntesis enzimáticas *in vitro*, pero estaban limitadas por un requisito de precursores costosos de nucleótido-azúcar. La invención supera los inconvenientes de estos intentos previos al proporcionar nuevas estrategias para fabricar grandes cantidades de oligosacáridos de la leche humana para usar como suplementos dietéticos. La invención hace uso de una bacteria diseñada por ingeniería genética *E. coli* (u otra bacteria) diseñada para producir oligosacáridos sialilados en niveles comercialmente viables, por ejemplo, los métodos de la invención, o como se describen, permiten la producción de 3'-SL a > 50 g/L en biorreactores.

Variantes y fragmentos funcionales.

60 La presente invención presenta la introducción de genes exógenos en bacterias para manipular las vías para aumentar las reservas de UDP-GlcNAc, para producir oligosacáridos sialilados y para producir oligosacáridos que contienen N-acetilglucosamina. En cualquiera de los métodos de la invención, o como se describen, los genes o productos génicos pueden ser variantes o fragmentos funcionales de los mismos.

Una variante de cualquiera de los genes o productos genéticos descritos en este documento puede tener 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98 %, o 99% de identidad de secuencia con las secuencias de ácido nucleico o aminoácidos descritas en el presente documento. La expresión "% de identidad", en el contexto de dos o más secuencias de ácido nucleico o polipéptido, se refiere a dos o más secuencias o subsecuencias que son iguales o tienen un porcentaje específico de residuos de aminoácidos o nucleótidos que son iguales, cuando se comparan y alinean para una máxima correspondencia, según lo medido utilizando 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 codificantes de las secuencias que se comparan, o la longitud de un fragmento particular o dominio funcional del mismo.

Las variantes, como se describen, también incluyen homólogos, ortólogos o parálogos de los genes o productos génicos, como se describen, que conservan la misma función biológica que los genes o productos génicos especificados en el presente documento. Estas variantes se pueden usar indistintamente con los genes enumerados en estos métodos. Tales variantes pueden demostrar un porcentaje de homología o identidad, por ejemplo, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, o 99% de identidad en dominios conservados importantes para la función biológica, tal como en un dominio funcional, por ejemplo, un dominio catalítico.

Para la comparación de secuencias, una secuencia actúa como una secuencia de referencia, con la que se comparan las secuencias de prueba. Cuando se utiliza un algoritmo de comparación de secuencias, las secuencias de prueba y referencia se ingresan en una computadora, las coordenadas de la subsecuencia se designan, si es necesario, y se designan los parámetros del programa de algoritmo de secuencia. El algoritmo de comparación de secuencia calcula luego el porcentaje de identidad de secuencia para la secuencia o secuencias de prueba en relación con la secuencia de referencia, en función de los parámetros de programa designados. El porcentaje de identidad se determina utilizando 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 ejemplares: (1) El umbral esperado era 10; (2) El costo de la brecha era Existencia: 11 y Extensión: 1; (3) La matriz empleada fue BLOSUM62; (4) El filtro para regiones de baja complejidad estaba "activado".

Los cambios pueden introducirse por mutación en la secuencia de ácido nucleico o secuencia de aminoácidos de cualquiera de los genes o productos genéticos descritos en el presente documento, lo que lleva a cambios en la secuencia de aminoácidos de la proteína o enzima codificada, sin alterar la capacidad funcional de la proteína o enzima. Por ejemplo, las sustituciones de nucleótidos que conducen a sustituciones de aminoácidos en residuos de aminoácidos "no esenciales" pueden realizarse en la secuencia de cualquiera de las secuencias expresamente descritas en este documento. Un residuo de aminoácido "no esencial" es un residuo en una posición en la secuencia que se puede alterar de la secuencia de tipo natural del polipéptido sin alterar la actividad biológica, mientras que un residuo de aminoácido "esencial" es un residuo en una posición que se requiere para la actividad biológica. Por ejemplo, no es probable que los residuos de aminoácidos que se conservan entre los miembros de una familia de proteínas sean susceptibles de mutación. Sin embargo, otros residuos de aminoácidos (por ejemplo, aquellos que están mal conservados entre los miembros de la familia de proteínas) pueden no ser tan esenciales para la actividad y, por lo tanto, es más probable que sean susceptibles de alteración. Por lo tanto, se describen moléculas de ácido nucleico que codifican las proteínas o enzimas descritas en el presente documento que contienen cambios en los residuos de aminoácidos con respecto a las secuencias de aminoácidos descritas en el presente documento que no son esenciales para la actividad.

Una molécula de ácido nucleico aislada que codifica una proteína homóloga a cualquiera de los genes descritos en el presente documento puede crearse introduciendo una o más sustituciones, adiciones o eliminaciones de nucleótidos en la secuencia de nucleótidos correspondiente, de manera que se introducen una o más sustituciones, adiciones o eliminaciones de aminoácidos. en la proteína codificada.

Se pueden introducir mutaciones en una secuencia de ácido nucleico de modo que la secuencia de aminoácidos codificada se altera mediante técnicas estándar, como la mutagénesis dirigida al sitio y la mutagénesis mediada por PCR. Las sustituciones de aminoácidos conservativas se realizan en uno o más restos de aminoácidos no esenciales predichos. Una "sustitución conservadora de aminoácidos" es aquella en la que el residuo de aminoácido se reemplaza por un residuo de aminoácido que tiene una cadena lateral similar. Se han definido en la técnica familias de residuos de aminoácidos que tienen cadenas laterales similares. Ciertos aminoácidos tienen cadenas laterales con más de una característica clasificable. Estas familias incluyen aminoácidos con cadenas laterales básicas (por ejemplo, lisina, arginina, histidina), cadenas laterales ácidas (por ejemplo, ácido aspártico, ácido glutámico), cadenas laterales polares sin carga (por ejemplo, glicina, asparagina, glutamina, serina, treonina, tirosina, triptófano, cisteína), cadenas laterales no polares (por ejemplo, alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, tirosina, triptófano), cadenas laterales beta-ramificadas (por ejemplo, treonina, valina, isoleucina) y cadenas laterales aromáticas (por ejemplo, tirosina, fenilalanina, triptófano, histidina). Por lo tanto, un residuo de aminoácido no esencial predicho en un polipéptido dado se reemplaza con otro residuo de aminoácido de la misma familia de cadenas laterales. Alternativamente, en otra realización, pueden introducirse mutaciones aleatoriamente a lo largo de todo o parte de una secuencia de codificación dada, tal como mediante mutagénesis de saturación, y los mutantes resultantes pueden seleccionarse para determinar la actividad biológica del polipéptido dado para identificar mutantes que retienen la actividad. A la inversa, la invención también proporciona variantes con mutaciones que mejoran o aumentan la actividad biológica endógena. Después de la mutagénesis de la secuencia

de ácido nucleico, la proteína codificada puede expresarse mediante cualquier tecnología recombinante conocida en la técnica y puede determinarse la actividad de la proteína. El experto en la materia puede medir fácilmente un aumento, disminución o eliminación de una actividad biológica dada de las variantes descritas como se describe, es decir, midiendo la capacidad para mediar en la modificación, síntesis o degradación de los oligosacáridos (a través de la detección de los productos).

También se describen fragmentos funcionales de los genes o productos génicos como se describe en el presente documento. Un fragmento, en el caso de estas secuencias y todas las demás que se proporcionan en este documento, se define como una parte del todo que es menor que el todo. Además, un fragmento varía en tamaño desde un solo nucleótido o aminoácido dentro de una secuencia de polinucleótidos o polipéptidos hasta unos pocos nucleótidos o aminoácidos que la secuencia completa de polinucleótidos o polipéptidos. Finalmente, un fragmento se define como cualquier porción de una secuencia completa de polinucleótidos o polipéptidos que es intermedia entre los extremos definidos anteriormente.

Por ejemplo, los fragmentos de cualquiera de las proteínas o enzimas descritas en este documento o codificadas por cualquiera de los genes descritos en este documento pueden ser de 10 a 20 aminoácidos, de 10 a 30 aminoácidos, de 10 a 40 aminoácidos, de 10 a 50 aminoácidos, de 10 a 60 aminoácidos, de 10 a 70 aminoácidos, 10 a 80 aminoácidos, 10 a 90 aminoácidos, 10 a 100 aminoácidos, 50 a 100 aminoácidos, 75 a 125 aminoácidos, 100 a 150 aminoácidos, 150 a 200 aminoácidos, 200 a 250 aminoácidos, 250 a 300 aminoácidos, 300 a 350 aminoácidos, 350 a 400 aminoácidos, 400 a 450 aminoácidos, o de 450 a 500 aminoácidos. Los fragmentos abarcados en la presente invención comprenden fragmentos que retienen fragmentos funcionales. Como tales, los fragmentos pueden retener los dominios catalíticos que se requieren o son importantes para la actividad funcional. Los fragmentos pueden determinarse o generarse utilizando la información de secuencia en el presente documento, y los fragmentos pueden probarse para determinar la actividad funcional utilizando métodos estándar conocidos en la técnica. Por ejemplo, la proteína codificada se puede expresar mediante cualquier tecnología recombinante conocida en la técnica y se puede determinar la actividad de la proteína. La función biológica de dicho fragmento puede medirse midiendo la capacidad para sintetizar o modificar un sustrato oligosacárido o, por el contrario, para catabolizar un sustrato oligosacárido.

Ejemplo 1. Ingeniería de *E. coli* para generar cepas huésped para la producción de oligosacáridos de leche humana que contienen N-acetilglucosamina

El protótipo de *E. coli* K12, W3110, se eligió como fondo progenitor para la biosíntesis de hMOS. Esta cepa había sido modificada previamente en el locus *ampC* por la introducción de un constructo represor $P_{trpB-cl+}$ inducible por triptófano (McCoy, J. y Lavallie, E. Current protocols in molecular biology/edited by Frederick M. Ausubel et al., (2001)), que permite la producción económica de proteínas recombinantes del promotor del fago λ PL (Sanger, F., Coulson, AR, Hong, GF, Hill, DF y Petersen, GB J Mol Biol 162, 729-773 (1982)) mediante la inducción con concentraciones milimolares de triptófano (Mieschendahl, M., Petri, T. & Hänggi, U. Nature Biotechnology 4, 802-808 (1986)). La cepa G1724, un derivado de *E. coli* W3110 que contiene la construcción del represor $P_{trpB-cl+}$ inducible por triptófano en *ampC*, se usó en la base para otras manipulaciones de la cepa de *E. coli*.

La biosíntesis de hMOS requiere la generación de un reserva celular mejor de lactosa. Esta mejora se logró en la cepa G1724 mediante varias manipulaciones del cromosoma usando la recombinación λ Red (Court, DL, Sawitzke, JA y Thomason, LC Annu Rev Genet 36, 361-388 (2002)) y la transducción de fagos generalizada P1 (Thomason, LC, Costantino, N. & Court, DL Mol Biol, Capítulo 1, Unidad 1.17 (2007)). La capacidad de la cepa huésped de *E. coli* para acumular lactosa intracelular se diseñó por primera vez mediante la eliminación simultánea del gen endógeno de β -galactosidasa (*lacZ*) y el gen represor del operón de lactosa (*lacI*). Durante la construcción de esta delección, el promotor *lacIq* se colocó inmediatamente cadena arriba del gen de la lactosa permeasa, *lacY*. La cepa modificada mantiene así su capacidad para transportar lactosa desde el medio de cultivo (a través de *LacY*), pero se elimina para la copia de tipo natural del gen *lacZ* (β -galactosidasa) responsable del catabolismo de la lactosa. Por lo tanto, se crea una reserva de lactosa intracelular cuando la cepa modificada se cultiva en presencia de lactosa exógena.

Una modificación adicional útil para aumentar la acumulación citoplásmico de lactosa libre (y, por lo tanto, el rendimiento final de hMOS) es la incorporación de una mutación *lacA*. *LacA* es una acetiltransferasa de lactosa que solo es activa cuando se acumulan altos niveles de lactosa en el citoplasma de *E. coli*. La alta osmolaridad intracelular (p. ej., causada por una alta acumulación de lactosa intracelular) puede inhibir el crecimiento bacteriano, y la *E. coli* ha desarrollado un mecanismo para protegerse de la alta osmolaridad intracelular causada por la lactosa al "marcar" 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* diseñada para producir oligosacáridos de la 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, ya que llevar una delección del gen *lacA* hace que la bacteria sea incapaz de sintetizar acetil-lactosa.

Se introdujo una mutación *thyA* (timidilato sintetasa) eliminando casi por completo el gen *thyA* y reemplazándolo por un gen de *E. coli* *lacZ*⁺ funcional insertado, de tipo natural, pero sin promotor, que portaba el sitio de unión del

ribosoma 2.8 (SEQ ID NO: 10) (Δ thyA::(2.8RBS lacZ+, kan')). Se usó la recombinación λ Red para realizar la construcción. La Fig. 4 ilustra la nueva configuración de genes así diseñados en el locus thyA. Se describe la secuencia completa de ADN de la región, con anotaciones en formato GenBank. La secuencia de ADN genómico que rodea la inserción lacZ+ en la región thyA se establece en la SEQ ID NO: 1.

5 El defecto thyA puede complementarse in trans suministrando un gen thyA de tipo natural en un plásmido multicopia (Belfort, M., Maley, G. F. & Maley, F. Proceedings of the National Academy of Sciences, 80, 1858 (1983)). Esta complementación se usa en el presente documento como un medio de mantenimiento del plásmido (eliminando la necesidad de un esquema de selección de antibióticos más convencional para mantener el número de copias del plásmido).

10 El genotipo de la cepa E680 se da a continuación. E680 incorpora todos los cambios descritos anteriormente y es una cepa hospedadora adecuada para la producción de oligosacáridos que contienen N-acetilglucosamina.

F'402 proA+B+, PlacIq-lacY, Δ (lacI-lacZ)158, Δ lacA398/araC, Δ gpt-mhpC, Δ thyA::(2.8RBS lacZ+, KAN), rpoS+, rph+, ampC::(Ptrp T7g10 RBS- λ cl+, CAT)

15 E796 es una cepa similar a E680 y porta una mutación thyA (timidilato sintetasa), introducida eliminando casi por completo el gen thyA y reemplazándolo por un gen de E. coli lacZ+ insertado, funcional, pero sin promotor, pero que lleva el sitio de unión del ribosoma 0.8 (SEQ ID NO: 11) [Δ thyA::(0.8RBS lacZ+, KAN)]. El genotipo de la cepa E796 se da a continuación. E796 incorpora todos los cambios descritos anteriormente y es una cepa hospedadora adecuada para la producción de oligosacáridos que contienen N-acetilglucosamina.

20 F'402 proA+B+, PlacIq-lacY, Δ (lacI-lacZ)158, Δ lacA398/araC, Δ gpt-mhpC, Δ thyA::(2.8RBS lacZ+, KAN), rpoS+, rph+, ampC::(Ptrp T7g10 RBS- λ cl+, CAT)

25 E866 es una cepa similar a E796 y es útil para la selección de plásmidos duales. E866 también tiene una mutación thyA (timidilato sintetasa), introducida al eliminar casi por completo el gen thyA y reemplazarlo por un gen de E. coli lacZ+ funcional insertado, de tipo natural, pero sin promotor, y que lleva el sitio de unión al ribosoma 0.8 (SEQ ID NO: 11) [Δ thyA::(0.8RBS lacZ+)]. Además de la eliminación de thyA, E866 también conlleva una eliminación del gen purA. El genotipo de la cepa E866 se da a continuación. E866 incorpora todos los cambios descritos anteriormente y es una cepa hospedadora adecuada para la producción de oligosacáridos que contienen N-acetilglucosamina.

F'402 proA+B+, PlacIq-lacY, Δ (lacI-lacZ)158, Δ lacA398/araC, Δ gpt-mhpC, Δ thyA::(0.8RBS lacZ+), rpoS+, rph+, ampC::(Ptrp T7g10 RBS- λ cl+, CAT), Δ purA727::KAN

30 Ejemplo 2. Producción de oligosacáridos de leche humana que contienen N-acetilglucosamina en E. coli: Lacto-N-tetraosa (LNT) y Lacto-N-neotetraosa (LNnT)

35 El primer paso en la síntesis (a partir de un precursor de lactosa) tanto de Lacto-N-tetraosa (LNT) como de Lacto-N-neotetraosa (LNnT) es la adición de un residuo de β (1,3)-N-acetilglucosamina a la lactosa, utilizando una β (1,3)-N-acetilglucosaminiltransferasa heteróloga para formar Lacto-N-triosa 2 (LNT2). El plásmido pG292 (ColE1, thyA+, bla+, P_L-lgtA) (SEQ ID NO: 2, FIG. 5) porta el gen lgtA β (1,3)-N-acetilglucosaminiltransferasa de N. meningitidis y puede dirigir la producción de LNT2 en la cepa de E. coli E680 en condiciones de cultivo apropiadas. pG221 (ColE1, thyA+, bla+, P_L-lgtA-wbgO) (SEQ ID NO: 3, FIG. 6) es un derivado de pG292 que transporta (dispuesto como un operón) tanto el gen de lgtA β (1,3)-N-acetilglucosaminiltransferasa de N. meningitidis como el gen de wbgO β (1,3)-galactosiltransferasa de E. coli O55:H7. pG221 dirige la producción de LNT en la cepa de E. coli E680 en condiciones de cultivo apropiadas. pG222 (ColE1, thyA+, bla+, P_L-lgtA-4GalT) (SEQ ID NO: 4, FIG. 7) es un derivado de pG292 que transporta (dispuesto como un operón) tanto el gen de lgtA β (1,3)-N-acetilglucosaminiltransferasa de N. meningitidis como el gen 4GalT (jhp0765) β (1,4)-galactosiltransferasa de H. pylori. pG222 dirige la producción de LNnT en la cepa de E. coli E680 en condiciones de cultivo apropiadas.

45 La adición de triptófano al medio de crecimiento que contiene lactosa de cultivos de cualquiera de las cepas derivadas de E680 transformadas con los plásmidos pG292, pG221 o pG222 conduce, para cada combinación particular de E680/plásmido, a la activación del triptófano del hospedador de E. coli, utilización del represor TrpR, la subsiguiente represión de P_{trpB} y la consiguiente disminución de los niveles citoplásmicos de cl, lo que resulta en una des-represión de P_L, expresión de lgtA, lgtA+wbgO, o lgtA+4GalT, respectivamente, y la producción de LNT2, LNT o LNnT, respectivamente.

50 Para la producción de LNT2, LNT o LNnT en cultivos de laboratorio a pequeña escala (<100 ml), las cepas se cultivaron a 30°C en un medio selectivo que carecía tanto de timidina como de triptófano hasta la fase exponencial temprana (p. ej., sales M9, glucosa al 0,5%, 0,4% de casaminoácidos). Luego se agregó lactosa a una concentración final de 0,5 o 1%, junto con triptófano (200 μ M final) para inducir la expresión de las respectivas glicosiltransferasas, impulsadas por el promotor P_L. Al final del período de inducción (~ 24 h), se realizó un análisis de TLC en partes alícuotas del medio de cultivo libre de células. La Fig. 8 ilustra esquemáticamente las reacciones enzimáticas necesarias para producir, a partir de la lactosa, a través del trisacárido intermedio lacto-N-triosa 2 (LNT2), los dos oligosacáridos de la leche humana; Lacto-N-tetraosa (LNT) y Lacto-N-neotetraosa (LNnT). Se presenta un cromatograma de capa fina (a la izquierda) de muestras de medio de cultivo tomadas de cultivos de E.

coli a pequeña escala y que demuestran la síntesis de LNT2, LNT y LNnT (utilizando cultivos de E680 inducidos que contienen lactosa transformados con pG292, pG221 o pG222, respectivamente). Se presenta un segundo cromatograma de capa fina (a la derecha) de muestras de medio de cultivo tomadas de un cultivo de biorreactor E680/pG222 de 15 L de *E. coli* y que demuestra la síntesis de LNnT (así como el hMOS de mayor peso molecular, Lacto-N-neohexaosa, LNnH).

Aunque los resultados anteriores demuestran claramente cómo es posible sintetizar oligosacáridos que contienen GlcNAc (es decir, LNT2, LNT y LNnT) en *E. coli* modificada, la FIG. 14 ilustra un problema serio al tratar de usar la reserva de UDP-GlcNAc de *E. coli* durante tales síntesis. En la Fig. 14 se cultivaron cuatro cultivos separados de E680, transformados con pG292, en presencia y en ausencia de lactosa, y con la expresión de LgtA, se indujo y no se indujo por adición de triptófano. Se puede ver claramente que se produce una lisis celular masiva en los cultivos donde está presente la lactosa, es decir, en aquellos cultivos en los que LgtA reduce la acumulación celular de UDP-GlcNAc agregando GlcNAc a la lactosa (y haciendo LNT2). Al hacerlo, la UDP-GlcNAc se desvía de la biosíntesis de la pared celular hacia la biosíntesis de hMOS y resulta en lisis celular. Esta lisis puede controlarse fácilmente no solo por la caída precipitada en la densidad del cultivo como se ve en la figura, sino también por la aparición de ADN en el medio de cultivo.

Ejemplo 3. El aumento de la acumulación de UDP-GlcNAc celular previene la lisis celular durante la biosíntesis de LNnT en *E. coli* modificada

Para examinar el impacto de la mejora de la acumulación celular de UDP-GlcNAc de *E. coli* durante la síntesis de hMOS que contiene N-acetilglucosamina, se construyó el plásmido p15A de replicón pG356 (Figura 19 y SEQ ID NO: 12). pG356 transporta un replicón p15A (compatible con los replicones ColE1), marcadores seleccionables *purA* y *ampC*, y un operón sintético (bajo el control del promotor pL) que lleva el *glmS* de *E. coli* (que codifica L-glutamina:D-fructosa-6-fosfato aminotransferasa) y *nagC* (que codifica el activador/represor transcripcional bifuncional de los operones *glm* y *nag*). Cuando pL está activo en las cepas que llevan el plásmido pG356, aumenta la acumulación de UDP-GlcNAc. La cepa E796 (ver ejemplo 1) se transformó con pG222 (FIG. 7), y la cepa E866 (ver ejemplo 1) se transformó con pG222 (FIG. 7) y pG356 (FIG. 19). (Las cepas E796 y E866 son isogénicas excepto por la mutación *purA* encontrada en E866 que se usa para la retención del plásmido pG356). Se realizaron corridas de fermentación idénticas de 1.5 L en cada una de las cepas transformadas. Se siguió la densidad óptica de los cultivos y la biosíntesis de LNnT, junto con los parámetros de fermentación estándar. Como puede verse en la fig. 18, el cultivo E796/pG222 produjo LNnT, pero se lisó cuando la densidad celular alcanzó 75 OD600, y alcanzó una densidad celular final al final de la fermentación de solo 50 OD600. En contraste (Figura 19) con el cultivo E866/pG222+pG356 (donde la expresión de los genes *glmS* y *bagC* mejora la acumulación intracelular de UDP-GlcNAc), también se produjo LNnT, pero sin lisis celular observada. En este cultivo, la densidad celular al final de la fermentación alcanzó 108 OD600, más del doble de la densidad alcanzada para E796/pG222.

Ejemplo 4. Producción de 6'-sialil-lactosa (6'-SL) por *E. coli* diseñada por ingeniería (Δ nanRATEK)

Para la producción de 6'-sialil-lactosa, se diseñó *Escherichia coli* GI724 (ATCC55151) con un conjunto de mutaciones que causaban la acumulación citoplásmica del precursor de lactosa no acetilado y prevenían la degradación del ácido N-acetil-5-neuramínico (Figura 3). En particular, se eliminaron los genes *lacZ* (β -galactosidasa) y *lacA* (lactosa acetil transferasa) del operón *lac*, dejando el represor *LacIq* y la permeasa *LacY* completamente funcionales. La permeasa *LacY* puede ser activada por promotores débiles (por ejemplo, *lac8*) o fuertes (por ejemplo, *Ptac*). El operón *nan* completo (*nanRATEK*; genes estructurales y reguladores involucrados en la degradación del ácido neuramínico) se eliminó en este ejemplo. Las manipulaciones del genoma de *E. coli* se lograron utilizando una combinación de técnicas genéticas moleculares estándar, específicamente la recombinación de λ -Red, los intercambios de alelos con vectores de suicidio de selección positiva y las transducciones P1 (Figura 3). El genotipo huésped de la cepa E781, adecuado para la producción de hMOS sialilados, se presenta a continuación:

ampC::(*P*_{trp- λ cl+}), *lacIq lacPL8*, Δ nanRATEK471, Δ lacZ690, Δ lacA 745

Para producir 6'-sialil-lactosa, la acumulación celular de UDP-GlcNAc se debe convertir en el precursor activado de azúcar-nucleótido, CMP-NeuAc, que, a su vez, puede funcionar como una molécula donante para un aceptor de azúcar (es decir, lactosa) en una reacción catalizada por sialiltransferasa (figura 3). Para este propósito, se expresaron de forma constitutiva tres genes de *Campylobacter jejuni* ATCC43438, que codificaban i) UDP-N-acetilglucosamina 2-epimerasa (*NeuC*), ii) N-acetilneuraminato sintasa (*NeuB*), y iii) N-Acetilneuraminato citidilil-transferasa (*NeuA*), en la cepa de *E. coli* diseñada anteriormente, junto con un gen que codificaba una α (2,6) sialiltransferasa de *Photobacterium* spp JT-ISH-224 (SEQ ID NO: 21 Acceso de la proteína del Genbank BAF92026). Los genes *neu* se expresaron a partir de un vector plasmídico de bajo número de copias (pG317, Figura 9, SEQ ID NO: 5) que llevaba un promotor *lac* constitutivo (pBBR1 ori, *cat+*, *Plac*), mientras que el gen de la α (2,6)sialiltransferasa fue expresado a partir de un vector plasmídico de alto número de copias (pG315, Figura 10, SEQ ID NO: 6) que llevaba un promotor *lac* constitutivo (ColE1 ori, *bla+*, *Plac*). Para evitar la síntesis de productos secundarios, la expresión relativa para el gen de la α (2,6)sialiltransferasa en comparación con los genes *neu* se modula mediante la ingeniería de diferentes sitios de unión ribosómica (RBS) que proporcionan diversos grados de eficiencia de la traducción cadena arriba del gen de la α (2,6)sialiltransferasa. Las cepas diseñadas se cultivaron a

alta densidad en fermentadores a escala piloto utilizando una estrategia de lote a lote de alimentación. La Fig. 11 es un análisis de TLC de sobrenadantes de cultivos de dos fermentaciones de este tipo, con muestras a la izquierda de la figura tomadas de una fermentación de una cepa que contiene pG315 (y, por lo tanto, que lleva el RBS presentado en la SEQ ID NO: 7 frente al gen de $\alpha(2,6)$ sialiltransferasa en el vector). Las muestras a la derecha de la figura se toman de una fermentación de una cepa que contiene una variante cercana de pG315 (pG345, FIG. 12, SEQ ID NO: 9, que lleva el RBS más débil presentado en la SEQ ID NO: 8 delante del gen de $\alpha(2,6)$ sialiltransferasa y reemplazando el RBS presentado en la SEQ ID NO: 7). En ambos casos, el precursor de lactosa se añadió a una densidad celular de 50 OD₆₀₀ y la conversión eficiente a productos finales se logró en las 48 horas posteriores a la adición de lactosa. El rendimiento final de 6'SL se incrementó cuando se utilizó el plásmido con el RBS más débil cadena arriba del gen de la $\alpha(2,6)$ sialiltransferasa y, además, el nivel del producto secundario KDO-lactosa se redujo significativamente al usar este RBS más débil. La identidad de la 6'-SL purificada mediante cromatografía en columna de carbón activado se confirmó mediante espectrometría de masas ESI y RMN.

Ejemplo 5. Producción de 6'-sialil-lactosa (6'-SL) por *E. coli* diseñada. (Δ nanA, Δ nanATE)

Para la producción de 6 'sialil-lactosa, se diseñó *Escherichia coli* GI724 (ATCC55151) con un conjunto de mutaciones que causaban la acumulación citoplásmica del precursor de lactosa no acetilada y que prevenían la degradación del ácido N-acetil-5-neuramínico (Figura 13). En particular, se eliminaron los genes lacZ (β -galactosidasa) y lacA (lactosa acetil transferasa) del operón lac, dejando el represor LacI y la permeasa LacY completamente funcionales. La permeasa LacY puede ser activada por promotores débiles (por ejemplo, lac8) o fuertes (por ejemplo, Ptac). Si bien todo el operón nan (nanRATEK; genes estructurales y reguladores implicados en la degradación del ácido neuramínico) se puede eliminar para suprimir el catabolismo del ácido neuramínico como en el Ejemplo 4, también son adecuadas eliminaciones menores que abarcan solo los genes nanA, o nanA, nanT y nanE, o nanA y nanE. En todos los casos en los que se mutó el gen nanE, los últimos 104 bp del gen nanE se dejaron intactos para permitir la transcripción/traducción imperturbada de nanK cadena abajo, aunque son posibles otras longitudes de la secuencia nanE residual. Las manipulaciones del genoma de *E. coli* se lograron utilizando una combinación de técnicas genéticas moleculares estándar, específicamente la recombinación de lambda-Red, los intercambios de alelos con vectores suicidas de selección positiva y las transducciones P1 (Figura 13). Los genotipos hospedadores de las cepas E971, E1017 y E1018, adecuados para la producción de hMOS sialilados con diversos rendimientos y purezas, se presentan a continuación:

ampC::(Ptrp- λ cl+), lacIq lacPL8, Δ nanA::kanR, Δ lacZ690, Δ lacA::scar,

ampC::(Ptrp- λ cl+), lacIq lacPL8, Δ nanATE::kanR::nanK+, Δ lacZ690, Δ lacA::scar y

ampC::(Ptrp- λ cl+), lacIq lacPL8, Δ nanATE::cicat::nanK+, lacZ690 Δ lacA::scar, respectivamente.

Para producir 6'-sialil-lactosa, la acumulación celular UDP-GlcNAc se debe convertir en el precursor activado de azúcar-nucleótido, CMP-NeuAc, que, a su vez, puede funcionar como una molécula donante para un aceptor de azúcar (es decir, lactosa) en una reacción catalizada por sialiltransferasa (figura 13). Para este propósito, se expresaron de forma constitutiva tres genes de *Campylobacter jejuni* ATCC43438, que codificaban i) UDP-N-acetilglucosamina 2-epimerasa (NeuC), ii) N-acetilneuraminato sintasa (NeuB), y iii) N-Acetilneuraminato citidililtransferasa (NeuA), en la cepa de *E. coli* diseñada anteriormente, junto con un gen que codificaba una $\alpha(2,6)$ sialiltransferasa de *Photobacterium* spp JT-ISH-224. Los genes neu se expresaron a partir de un vector plasmídico de bajo número de copias (pG317, Figura 9, SEQ ID NO: 5) que llevaba un promotor lac constitutivo (pBBR1 ori, cat+, Plac), mientras que el gen de la $\alpha(2,6)$ sialiltransferasa fue expresado desde el RBS débil de la SEQ ID NO: 8 en un vector plasmídico de alto número de copias (pG345, Figura 12, SEQ ID NO: 9) que lleva un promotor lac constitutivo (ColE1 ori, bla +, Plac). Las cepas diseñadas se cultivaron a alta densidad en fermentadores a escala piloto utilizando una estrategia de lote a lote de alimentación. La Fig. 14 es un análisis de TLC de gránulos de cultivo o sobrenadantes de tres de tales fermentaciones. El panel A muestra la producción y acumulación de 6'SL en las células de tres fondos genéticos (solo se muestran las mutaciones nan relevantes para las cepas E971, E1017 y E1018), el panel B y C muestran la producción y acumulación de 6'SL en el medio extracelular (sobrenadantes) en las cepas E971, E1017 y E1018 (solo se muestran las mutaciones nan relevantes) con rendimientos volumétricos máximos estimados de 15 g por litro de sobrenadante. En todos los casos, el precursor de lactosa se agregó a una densidad celular de 40 OD₆₀₀ y la conversión en estado estacionario a productos finales se logró en aproximadamente 90 horas a partir de la adición de lactosa (EFT es el tiempo de fermentación transcurrido).

Las diversas secuencias presentadas en este documento se enumeran a continuación.

SEQ ID NO: 1

>E680_thyA::2.8RBS_lacZ cepa de *Escherichia coli*

GCAGCGGAACCTACAAGGCACCATAACGTCGCCCTCCCTGATAACGCTGATACTGTGGTCCG

CGGTTATGCCAGTTGGCATCTTCACGTAAATAGAGCAAATAGTCCCAGCGCTGGCTGGCG
 GTTTGCCATAGCCGTTGCGACTGCTGCCAGTATTGCCAGCCATAGAGTCCACTTGCGCTT
 AGCATGACCAAAATCAGCATCGCGACCAGCGTTTCAATCAGCGTATAACCACGTTGTGTT
 TTCATGCCGGCAGTATGGAGCGAGGAGAAAAAAGACGAGGGCCAGTTTCTATTTCTTCG
 GCGCATCTTCCGGACTATTTACGCCGTTGCAGGACGTTGCAAAATTTCCGGGAAGGCGTCT
 CGAAGAATTTAACGGAGGGTAAAAAAACCGACGCACACTGGCGTCGGCTCTGGCAGGATG
 TTTTCGTAATTAGATAGCCACCGGCGCTTTattaaacctactATGACCATGATTACGGATT
 CACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCTGGCGTTACCCAACCTAATC
 GCCTTGCGACACATCCCCCTTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCCGATC
 GCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCTTTGCCTGGTTTCCGGCAC
 CAGAAGCGGTGCCGGAAGCTGGCTGGAGTGCGATCTTCTGAGGCCGATACTGTGCTCG
 TCCCCTCAAACCTGGCAGATGCACGGTTACGATGCGCCCATCTACACCAACGTGACCTATC
 CCATTACGGTCAATCCGCCGTTTGTTCACCGGAGAATCCGACGGGTTGTTACTCGCTCA
 CATTAAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCAATTATTTTTGATGGCG
 TTAACCTCGGCGTTTCATCTGTGGTCAACGGGCGCTGGGTCGGTTACGGCCAGGACAGTC
 GTTTGCCGCTCTGAATTTGACCTGAGCGCATTTTTACGCGCCGGAGAAAAACCGCTCGCGG
 TGATGGTGCTGCGCTGGAGTGACGGCAGTTATCTGGAAAGATCAGGATATGTGGCGGATGA
 GCGGCATTTTCCGTGACGTCCTGTGCTGCATAAAACCGACTACACAAATCAGCGATTTCC
 ATGTTGCCACTCGCTTTAATGATGATTTAGCCGCGCTGTACTGGAGGCTGAAGTTCAGA
 TGTGCGGCGAGTTGCGTGACTACCTACGGGTAACAGTTTCTTTATGGCAGGGTGAAACGC
 AGGTCGCCAGCGGCACCGCGCTTTCCGGCGGTGAAATTATCGATGAGCGTGGTGGTTATG
 CCGATCGCGTCACACTACGTCGTAACGTCGAAAACCCGAAACTGTGGAGCGCCGAAATCC
 CGAATCTCTATCGTGCGGTGGTTGAACTGCACACCGCCGACGGCACGCTGATTGAAGCAG
 AAGCCTGCGATGTCGGTTTCCGCGAGGTGCGGATTGAAAATGGTCTGCTGCTGCTGAACG
 GCAAGCCGTTGCTGATTTCGAGGCGTTAACCGTCACGAGCATCATCTCTGCATGGTCAGG
 TCATGGATGAGCAGACGATGGTGCAGGATATCCTGCTGATGAAGCAGAACAACCTTAAACG
 CCGTGCGCTGTTTCGCAATTATCCGAACCATCCGCTGTGGTACACGCTGTGCGACCGCTACG
 GCCTGTATGTGGTGGATGAAGCCAATATTGAAACCCACGGCATGGTGCCAATGAATCGTC
 TGACCGATGATCCCGCTGGCTACCGGCGATGAGCGAACCGGTAACGCGAATGGTGCAGC
 GCGACTCGTAATCACCCGAGTGTGATCATCTGGTCTGCTGGGGAATGAATCAGGCCACGGC
 CTAATCACGACGCGCTGTATCGCTGGATCAAACTGTGTCGATCCTTCCCGCCCGGTGCAGT
 ATGAAGGCGGCGGAGCCGACACCACGGCCACCGATATTATTTGCCCGATGTACGCGCGG
 TGGATGAAGACCAGCCCTTCCCGGCTGTGCCGAAATGGTCCATCAAAAAATGGCTTTTCGC
 TACCTGGAGAGACGCGCCCGCTGATCCTTTGCGAATACGCCACGCGATGGGTAACAGTC
 TTGGCGGTTTTCGCTAAATACTGGCAGGCGTTTTCGTCAGTATCCCCGTTTACAGGGCGGCT
 TCGTCTGGGACTGGGTGGATCAGTCGCTGATTAATATGATGAAAACGGCAACCCGTGGT
 CGGCTTACGGCGGTGATTTTGGCGATACGCCGAACGATCGCCAGTCTGTATGAACGGTC
 TGGTCTTTGCCGACCGCACGCCGATCCAGCGCTGACGGAAGCAAAACACCAGCAGCAGT
 TTTTCCAGTTCGGTTTATCCGGGCAAACCATCGAAGTGACCAGCGAATACTGTTCCGTC
 ATAGCGATAACGAGCTCCTGCACTGGATGGTGGCGCTGGATGGTAAGCCGCTGGCAAGCG
 GTGAAGTGCTCTGGATGTGCTCCACAAGGTAACAGTTGATTGAACTGCTGAACTAC
 CGCAGCCGGAGAGCGCCGGGCAACTCTGGCTCACAGTACGCGTAGTGCAACCGAACCGGA
 CCGCATGGTGCAGAAGCCGGGCACATCAGCGCCTGGCAGCAGTGGCGTCTGGCGGAAAACC
 TCAGTGTGACGCTCCCCCGCGCTCCCACGCCATCCCGCATCTGACCACCAGCGAAATGG
 ATTTTTGCATCGAGCTGGGTAATAAGCGTTGGCAATTTAACCGCCAGTCAGGCTTTCTTT
 CACAGATGTGGATTGGCGATAAAAAACAACCTGtTGACCGCGCTGCGCGATCAGTTCACCC
 GTGCACCGCTGGATAACGACATTGGCGTAAGTGAAGCGACCCGATTGACCTAACGCCT
 GGGTCAACGCTGGAAGGCGGCGGGCCATTACCAGGCCGAAGCAGCGTTGTTGCAGTGCA
 CGGCAGATACACTTGCTGATGCGGTGCTGATTACGACCGCTCACGCGTGGCAGCATCAGG
 GGAAAACCTTATTTATCAGCCGGAACCTACCGGATTGATGGTAGTGGTCAAATGGCGA
 TTACCGTTGATGTTGAAGTGGCGAGCGATACACCGCATCCGGCGCGGATTGGCTGAACT
 GCCAGCTGGCGCAGGTAGCAGAGCGGGTAAACTGGCTCGGATTAGGGCCGCAAGAAAACCT
 ATCCCGACCGCTTACTGCCGCTGTTTTGACCGCTGGGATCTGCCATTGTCAGACATGT
 ATACCCCGTACGCTTCCCGAGCGAAAACGGTCTGCGCTGCGGGACGCGCGAATTGAATT
 ATGGCCACACCAGTGGCGCGGCGACTTCCAGTTC AACATCAGCCGCTACAGTCAACAGC
 AACTGATGGAAAACAGCCATCGCCATCTGCTGCACGCGGAAGAAGGCACATGGCTGAATA
 TCGACGGTTTTCCATATGGGGATTGGTGGCGACGACTCCTGGAGCCCGTCAGTATCGGGCG
 AATTCCAGCTGAGCGCCGGTCTGCTACCATACCAGTTGGTCTGGTGTCAAAAATAAGCGG
 CCGCTTATGTAGGCTGGAGCTGCTTCGAAGTTCTATACTTTCTAGAGAATAGGAACTT
 CGGAATAGGAACTTCAAGATCCCCCTTATTAGAAGAACTCGTCAAGAAGGCGATAGAAGGC
 GATGCGCTGCGAATCGGGAGCGGCGATACCGTAAAGCACGAGGAAGCGGTGAGCCCATTC

ES 2 700 274 T3

GCCGCCAAGCTCTTCAGCAATATCACGGGTAGCCAACGCTATGTCCTGATAGCGGTCCGC
CACACCCAGCCGGCCACAGTCGATGAATCCTGAAAAGCGGCCATTTCCACCATGATATT
CGGCAAGCAGGCATCGCCATGGGTACGACGAGATCCTCGCCGTCGGGCATGCGCGCCTT
GAGCCTGGCGAACAGTTTCGGCTGGCGCGAGCCCCTGATGCTCTTCGTCCAGATCATCCTG
ATCGACAAGACCGGCTTCCATCCGAGTACGTGCTCGCTCGATGCGATGTTTCGCTTGGTG
GTCGAATGGGCAGGTAGCCGGATCAAGCGTATGCAGCCGCCGCATTGCATCAGCCATGAT
GGATACTTTCTCGGCAGGAGCAAGGTGAGATGACAGGAGATCCTGCCCCGGCACTTCGCC
CAATAGCAGCCAGTCCCTTCCCGCTTCAGTGACAACGTGAGCACAGCTGCGCAAGGAAC
GCCCCTCGTGGCCAGCCACGATAGCCGCGCTGCCTCGTCTGACAGTTCATTACAGGGCACC
GGACAGGTTCGGTCTTGACAAAAAGAACCAGGGCGCCCCTGCGCTGACAGCCGGAACACGGC
GGCATCAGAGCAGCCGATTGTCTGTTGTGCCAGTCATAGCCGAATAGCCTCTCCACCCA
AGCGGCCGGAGAACCCTGCGTGCAATCCATCTTGTTCAATCATGCGAAACGATCCTCATCC
TGCTCTTGATCAGATCTTGATCCCTGCGCCATCAGATCCTTGCGGGCAAGAAAGCCAT
CCAGTTTACTTTGCAGGGCTTCCCAACCTTACCAGAGGGCGCCCAGCTGGCAATTCCGG
TTCGCTTGCTGTCCATAAAACCGCCAGTCTAGCTATCGCCATGTAAGCCCACTGCAAGC
TACCTGCTTTCTCTTTCGCGCTTGCCTTTCCCTTGTCAGATAGCCAGTAGCTGACATT
CATCCGGGGTCAGCACCGTTTCTGCGGACTGGCTTTCTACGTGTTCCGCTTCCCTTAGCA
GCCCTTGCGCCCTGAGTGCTTGCGGCAGCGTGAGCTTCAAAGCGCTCTGAAGTTCCCTAT
ACTTTCTAGAGAATAGGAACCTCGAACTGCAGGTCGACGGATCCCCGGAATCATGGTTCC
TCAGGAAACGTGTTGCTGTGGGCTGCGACGATAATGCCAGACCATCATGATCACACCCGC
GACAATCATCGGGATGGAAGAATTTGCCCATGCTGATGTAAGTGCACCCAGGCACCGGT
AACTGCGCGTCGGGCTGGCGGAAAACTCAACAATGATGCGAAACGCGCCGTAACCAAT
CAGGAACAAACCTGAGACAGCTCCCATTTGGGCGTGGTTTACGAATATACAGGTTGAGGAT
AATAAACAGCACACACCTTCCAGCAGCAGCTCGTAAAGCTGTGATGGGTGGCGCGGCAG
CACACCGTAAGTGTGAAAAATGGATTGCCACTGCGGGTTGGTTTGCAGCAGCAAAATATC
TTCTGTACGGGAGCCAGGGAACAGCATGGCAAACGGGAAGTTTCGGGTCAACGCGGCCCA
CAATTCACCGTTAATAAAGTTGCCAGACGCCCGGCACCAAGACCAAACGGAATGAGTGG
TGCGATAAAATCAGAGACCTGGAAGAAGGAACGTTTAGTACGGCGGGCGAAGATAATCAT
CACCACGATAACGCCAATCAGGCCGCGGTGGAAAGACATGCCGCGTCCCAGACACGGAA
CAGATACAGCGGATCGGCCATAAACTGCGGGAAATTGTAGAACAGAACATAACCAATACG
TCCCCGAGGAAGACGCCGAGGAAGCCCGCATAGAGTAAGTTTTCAACTTCATTTTTGGT
CCAGCCGCTGCCCGGACGATTCGCCCGTCTGTGTGCCAGCCACATTGCAAAAATGAAACC
CACCAGATACATCAGGCCGTACCAGTGAAGCGCCACGGGTCCTATTGAGAAAATGACCGG
ATCAAACCTCGGAAAATGCAGATAGCTACTGGTCATCTGTCACCACAAGTTCTTGTTATT
TCGCTGAAAGAGAACAGCGATTGAAATGCGCGCCGACAGGTTTCAGGCGCTCCAAAGGTGC
GAATAATAGCACAAAGGGGACCTGGCTGGTTGCCGATACCGTTAAAAGATATGTATA

SEQ ID NO: 2

>pG292, secuencia completa.

ES 2 700 274 T3

TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCA
CAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTG
TTGGCGGGTGTTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGC
ACCATATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAggcg
ccTCCTCAACCTGTATATTCGTAAACCACGCCCAATGGGAGCTGTCTCAGGTTTGTTCCT
GATTGGTTACGGCGCGTTTCGCATCATTGTTGAGTTTTTCCGCCAGCCCGACGCGCAGTT
TACCGGTGCCTGGGTGCAGTACATCAGCATGGGGCAAATTCCTTCCATCCCGATGATTGT
CGCGGGTGTGATCATGATGGTCTGGGCATATCGTCGCAGCCCACAGCAACACGTTTCCTG
AGGAACCATGAAACAGTATTTAGAAGTATGATGCAAAAAGTGCTCGACGAAGGCACACAGAA
AAACGACCGTACCGGAACCGGAACGCTTTCATTTTTGGTTCATCAGATGCGTTTTAACCT
GCAAGATGGATTCCCGCTGGTGACAATAACGTTGCCACCTGCGTTCATCATCCATGA
ACTGCTGTGGTTTTCTGCAGGGCGACACTAACATTGCTTATCTACACGAAAACAATGTCAC
CATCTGGGACGAATGGGCCGATGAAAACGGCGACCTCGGGCCAGTGTATGGTAAACAGTG
GCGCGCCTGGCCAACGCCAGATGGTTCGTCATATTGACCAGATCACTACGGTACTGAACCA
GCTGAAAAACGACCCGGATTTCGCGCCGATTATTGTTTTAGCGTGGAACGTAGGCGAACT
GGATAAAATGGCGCTGGCACCGTGCCATGCATTCTTCCAGTTCTATGTGGCAGACGGCAA
ACTCTCTTGCCAGCTTTATCAGCGCTCCTGTGACGTCTTCCCTCGGCCTGCCGTTCAACAT
TGCCAGCTACGCGTTATTGGTGCATATGATGGCGCAGCAGTGCGATCTGGAAGTGGGTGA
TTTTGTCTGGACCGGTGGCGACACGCATCTGTACAGCAACCATATGGATCAAACCTCATCT
GCAATTAAGCCGCGAACC GCGTCCGCTGCCGAAGTTGATTATCAAACGTAAACCCGAATC
CATCTTCGACTACCGTTTCGAAGACTTTGAGATTGAAGGCTACGATCCGCATCCGGGCAT
TAAAGCGCCGGTGGCTATCTAATTACGAAACATCCTGCCAGAGCCGACGCCAGTGTGCGT

CGGTTTTTTTACCCCTCCGTTAAATTCCTTCGAGACGCCTTCCCGAAAggcgcccATTCCGCCAT
 TCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGC
 TGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAAGTTGGGTAACGCCAGGGTTTTCCAGT
 CACGACGTTGTAACGACGCGCCAGTGCCAAGCTTACTGCTCACAAGAAAAAAGGCACGT
 CATCTGACGTGCCTTTTTTATTTGTACTACCCTGTACGATTACTGCAGGTCGACTCTAGA
 TGCATGCTCGAGTCAACGGTTTTTTCAGCAATCGGTGCAAAATGCCGAAGTATTGCCTCAA
 GGTAAACAGCCGCCGCATCCTGCCGTCTGCCGCAAAATCCAGCCACGCGCCGGCGGGCAG
 CGTGTCCGTCCGTTTGAAGCATTGGTACAAAAACCGCGGGCGCGTTCAAAAATCTTCTTC
 CGGCAAAATGTTTTCTCCAGCAATTCATACGCTACTGCTTTTATTTGGCGGTATCAAGGCT
 GTCGAACCGGGTTTTTAAAACCCATAGACTGCAAAAAATCGTTTTCTGGCGGTTTTTTGGAT
 GCCTTGCGCGATTTTCGTGTTGGCGGATGCTGTATTTGGATGAAACCTGATTGGCGTGAAG
 GCGGTATTTGACCAAGGCTTCGGGATAATAAGCCAGCCTGCCAATTTGCTGACATCGTA
 CCAAAATTTGGTAATCTTCCGCCAATCCCGCTCGGTGTTGTAACGCAAACCGCCGTCAT
 GACGCTGCGCCTCATAATCATCGTGTGTTGTTGTATGGGGTTGCCGAAAGGGAAAAAGTC
 GGCATGTCTTCGTGTCGGGTCGGTTTTTTTCCAAATTTTCCGCTGTTTCGTGGTGCCGCGC
 CAGCCGGTTGCCGTCTTTTTCTTCCGACAAAACTTCCAGCCACGCACCCATCGCGATGAT
 GCTGCGGTCTTTTTCCATCTCACCCACGATTTTCTCAATCCAGTCGGGGGGCGCAATATC
 GTCTGCATCGGTGCGCGCAATATAATCCCCCCCCCCCCCGACTTTGCCAATTCATCCAG
 CCCGATGTTTTAAAGAGGGAATCAGACCGGAATTCGCGCGCTGCGCGAGGATGCGGATGCG
 GCCGTCTGTTCTTGGAAACGCTGGGCAATGGCAAGCGTACCGTCCGTCGAGCCGTCATC
 GACAATCAAAAATATCCAAGTTGCGCCAAGTTTGATTACGACGCGCGCTAATGATTGGGC
 GAAATATTTTTCTACGTTGTAGGCGCAAATCAATACGCTGACTAAAGGCTGCAATTTATF
 CTCCCGATAGGCACGATGCCGTCTGAAGGCTTCAGACGGCATAATGtatatctccttcttg
 aaTTCTAACAATTGATTGAATGTATGCAAAATAATGCATACACCATAGGTGTGGTTAAT
 TTGATGCCCTTTTTTTCAGGGCTGGAATGTGTAAGAGCGGGGTTATTTATGCTGTTGTTTTT
 TTGTTACTCGGGAAGGGCTTTACCTCTTCCGCATAAACGCTTCCATCAGCGTTTATAGTT
 AAAAAAATCTTTCGGAACTGGTTTTGCGCTTACCCCAACCAACAGGGGATTTGCTGCTTT
 CCATTGAGCCTGTTTCTCTGCGCGACGTTTCGCGGCGCGGTGTTTGTGCATCCATCTGGAT
 TCTCCTGTACGTTAGCTTTGGTGGTGTGTGGCAGTTGTAGTCTGAACGAAAAACCCCGC
 CGATTGGCACATTGGCAGCTAATCCGGAATCGCACTTACGGCCAATGCTTCGTTTCGTAT
 CACACACCCCAAAGCCTTCTGCTTTGAATGCTGCCCTTCTTCAGGGCTTAATTTTTAAGA
 GCGTCACCTTCAITGGTGGTCAGTGCCTGCTGATGCTCAGTATCACCGCCAGTGGT
 ATTTATGTCAACACCGCCAGAGATAAATTTATCACCGCAGATGGTTATCTGTATGTTTTTT
 ATATGAATTTATTTTTTTCAGGGGGGCATTGTTTGGTAGGTGAGAGATCAATCTGCATT
 AATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCATTTGGGCGCTCTTCCGCTTCT
 CGCTCACTGACTCGCTGCGCTCGGTGCTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAA
 AGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAA
 AAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGC
 TCCGCCCCCTGACGAGCATCAGAAAAATCGACGCTCAAGTCAGAGGTGGCGAAAACCCGA
 CAGGACTATAAAGATACCAGGCGTTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTC
 CGACCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTT
 CTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCGCTCCAAGCTGGGCT
 GTGTGCACGAACCCCGTTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTG
 AGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTA
 GCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCT
 AACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAA
 GAGTTGGTAGCTCTTGATCCGGCAAAACAAACACCGCTGGTAGCGGTGGTTTTTTTTGTTT
 GCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTA
 CGGGGTCTGACGCTCAGTGAACGAAAACTCACGTAAAGGATTTTGGTCATGAGATTAT
 CAAAAAGGATCTTACCTAGATCCTTTTTAAATTAATAAATGAAGTTTTAAATCAATCTAAA
 GTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCT
 CAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTA
 CGATACGGGAGGGCTTACCATCTGGCCCCAGTGTGCAATGATAACCGCGAGACCCACGCT
 CACCGGCTCCAGATTTATCAGCAATAAACAGCCAGCCGGAAGGGCCGAGCGCAGAAAGTG
 GTCCTGCAACTTTATCCGCTCCATCCAGTCTATTAATTTGTTGCCGGGAAGCTAGAGTAA
 GTAGTTCGCCAGTTAATAGTTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGT
 CACGCTCGTCTGTTGGTATGGCTTCATTACGCTCCGCTTCCCAACGATCAAGGCGAGTTA
 CATGATCCCCATGTTGTGCAAAAAAGCGGTTAGTCTCTTCGGTCTCCGATCGTTGTCA
 GAAGTAAGTTGGCGCAGTGTATCACTCACTGATTGCTTATGGCAGCACTGCATAAATCTCTTA
 CTGTCAATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGGTACTCAACCAAGTCATTTCT
 GAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGCGTCAATACGGGATAAATACCG

ES 2 700 274 T3

CGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAAC
TCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCA
GATCTTCAGCATCTTTTACTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGC
AAAATGCCGCAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTT
CCTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACAT
ATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAA
AGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAAATAG
GGCGTATCACGAGGCCTTTTCGTC

SEQ ID NO: 3

>pG221, secuencia completa.

TCGCGCGTTTCGGTIGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCA
 CAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTG
 TTGGCGGGTGTTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGC
 ACCATATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAggcg
 ccTCCTCAACCTGTATATTCGTAAACCACGCCCAATGGGAGCTGTCTCAGGTTTGTTCCT
 GATTGGTTACGGCGCGTTCGCATCATTGTTGAGTTTTTCCGCCAGCCCAGCGGCAGTT
 TACCGGTGCCTGGGTGCAGTACATCAGCATGGGGCAAATTCTTCCATCCCGATGATTGT
 CGCGGGTGTGATCATGATGGTCTGGGCATATCGTCGCAGCCCACAGCAACACGTTTCCTG
 AGGAACCATGAAACAGTATTTAGAAGTATGCAAAAAGTGTCTGACGAAGGCACACAGAA
 AAACGACCGTACCGGAACCGGAACGCTTTCCATTTTTGGTCATCAGATGCGTTTTAACCT
 GCAAGATGGATTCGCGTGGTGACAATAAACGTTGCCACCTGCGTTCCATCATCCATGA
 ACTGCTGTGGTTTTCTGCAGGGCGACACTAACATTGCTTATCTACACGAAAACAATGTCAC
 CATCTGGGACGAATGGGCCGATGAAAACGGCGACCTCGGGCCAGTGTATGGTAAACAGTG
 GCGCGCTTGCCAACGCCAGATGGTGCATATGACCAGATCACTACGGTACTGAACCA
 GCTGAAAAACGACCCGGATTTCGCGCCGATTATTGTTTCAGCGTGGAACGTAGGCGAACT
 GGATAAAATGGCGCTGGCACCGTGCCATGCATTCTTCCAGTTCTATGTGGCAGACGGCAA
 ACTCTCTTGCCAGCTTTATCAGCGCTCCTGTGACGTCTTCTCGGCCTGCCGTTCAACAT
 TGCCAGCTACGCGTTATTGGTGCATATGATGGCGCAGCAGTGCATCTGGAAGTGGGTGA
 TTTTGTCTGGACCGGTGGCGACACGCATCTGTACAGCAACCATATGGATCAAACCTCATCT
 GCAATTAAGCCGCGAACCGCGTCCGCTGCCGAAGTTGATTATCAAACGTAAACCCGAATC
 CATCTTCGACTACCGTTTCGAAGACTTTGAGATTGAAGGCTACGATCCGCATCCGGGCAT
 TAAAGCGCCGGTGGCTATCTAATTACGAAACATCCTGCCAGAGCCGACGCCAGTGTGCGT
 CGGTTTTTTTACCCTCCGTTAAATTTCTTCGAGACGCCTTCCCGAaggcgccATTCCGCAT
 TCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGC
 TGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCAGT
 CACGACGTTGTAACGACGGCCAGTGCCAAGCTTACTGCTCACAAGAAAAAAGGCACGT
 CATCTGACGTGCCTTTTTTATTTGTAATACTACCCTGTACGATTACTGCAGGTCGACTCTAGA
 TGCATGCTCGAGTTATTATTTAATATATTTACAATAGATGAAGGACGCAATCGTACGGAT
 ACCGCCGAACAGGTAGTTAATGTTACCGGTCAGGAAGAAGCACTTCATTTTGATAACCAG
 GTCGTTAACCATCACCATGTACAGGTTTTTTTTTTGCGGTAGACTGACCTTCGTGCAGGCG
 GTAGTAGAACAGGTATTCGGCAGGTTTTTGGAACTTGATTTTTGCCAGGCTCAGACGGTT
 CCACAGCTCGTAATCTTCGGAGTAGTTAGAAAACATATAACCACCGATGCTCGCGATGAC
 TTTTTTACGAAACATTACGCTCGGGTGAACAATACAACACTTATACGGCAGGTTTTTAAAC
 GATGTCAGGTTCTCTTCGGGCAGTTTGGTCTTGGTGAATTCACGACCTTGTGTCGTCAT
 AAAGATTGCGTTGGTACCCACAACATCTACGTACGGATTGTTCTTCAGGAAGTCAACCTG
 TTTAGTAAAACGGTCCGGGTGAGAGATGTCGTCAGAGTCCATACGGGCAATAAATTCGCC
 GTTGCTCAGGTCGATCGTTTTGTTTACGGGAGTACGGCAGGTAAGCGATGTTAGTGCAGGAT
 CAGTTTGATTTTGTGTTAACTTTGTGTTTTAGTTTATAGAAGTCGTCAGTGCAGCA
 GTTCGCAACGATGATGATTTTGAAGCTGCTGAAGGTCTGAGACAGGATGCTGTTGATCGC
 TTCGTCCAGAAAAGGGTTTTTCTTGTAAACAGGCAGGATAACGCTCACAACCGGGTGGGT
 AGATTCGCGGATTCCGCTTCATCGATGATCATATGTATATCTCCTTCTTCTCGAGTCAA
 CGGTTTTTTCAGCAATCGGTGCAAAATGCCGAAGTATTGCCTCAAGGTAAACAGCCGCCGC
 ATCCTGCCGTCGCCGCAAAATCCAGCCACGCGCCGGCGGGCAGCGTGTCCGTCCGTTTG
 AAGCATTGGTACAAAACCGGCGGGCGCGTTCAAAATCTTCTTCCGGCAAAATGTTTCTCC
 AGCAATTCATACGCTACTGCTTTTATTTGGCGGTATTCAGGCTGTCGAACCGGGTTTTA
 AAACCCATAGACTGCAAAAATCGTTTCTGGCGGTTTTTTGGATGCCTTGCAGGATTTCG
 TGTTGGCGGATGCTGTATTTGGATGAAACCTGATTGGCGTGAAGGCGGTATTTGACCAAG
 GCTTCGGGATAATAAGCCAGCCTGCCAATTTGCTGACATCGTACCAAAATTTGGTAATCT
 TCCGCCAATCCCGCTCGGTGTTGTAACGCAACCGCCGTCAATGACGCTGCGCCTCATA
 ATCATCGTGTGTTGTGTATGGGGTTGCCGAAAGGGAAAAAGTCCGCAATGTCTTCGTGT

CGGGTCGGTTTTTTTCCAAATTTTGCCGTGTTTCGTGGTGGCGGCCAGCCGGTTGCCGTCC
TTTTCTTCCGACAAAACCTTCCAGCCACGCACCCATCGCGATGATGCTGCGGTCTTTTTCC
ATCTCACCCACGATTTTCTCAATCCAGTCGGGGGGCGCAATATCGTCTGCATCGGTGCGC
GCAATATATTTTTCCCCCCCCCCCCGACTTTGCCAATTCATCCAGCCCGATGTTTAAAGAG
GGAATCAGACCCGAATTGCGCGGGCTGCGCGAGGATGCGGATGCGGCCGTCTGTCTTGG
AAACGCTGGGCAATGGCAAGCGTACCGTCCGTCGAGCCGTCATCGACAATCAAAAATATCC
AAGTTGCGCCAAGTTTGATTACGACGGCGGCTAATGATTGGGCGAAAATTTTTTCTACG
TTGTAGGCGCAAATCAATACGCTGACTAAAGGCTGCAATTTATTCTCCCGATAGGCACGA
TGCCGTCTGAAGGCTTCAGACGGCATAIGtatatctccttcttgaatTCTAACAAITGAT
TGAATGTATGCAAATAAATGCATACCCATAGGTGTGGTTAAATTTGATGCCCTTTTTCA
GGGCTGGAATGTGTAAGAGCGGGTTATTTATGCTGTTGTTTTTTTTGTTACTCGGGAAGG
GCTTTACCTCTTCCGCATAAACGCTTCCATCAGCGTTTATAGTTAAAAAAATCTTTCGGA
ACTGGTTTTGCGCTTACCCCAACCAACAGGGGATTTGCTGCTTTCATTGAGCCTGTTTC
TCTGCGGACGTTTCGCGGCGGCGTGTGTTGTGCATCCATCTGGATTCTCCTGTCAGTTAGC
TTTGGTGGTGTGTGGCAGTTGTAGTCCTGAACGAAAACCCCCGCGATTGGCACATTGGC
AGCTAATCCGGAATCGCACTTACGGCCAATGCTTCGTTTCGTATCACACACCCCAAAGCC
TTCTGCTTTGAATGCTGCCCTTCTTCAGGGCTTAATTTTTAAGAGCGTCACCTTCATGGT
GGTCAGTGCCTCTGCTGATGTGCTCAGTATCACCGCCAGTGGTATTTATGTCAACACCG
CCAGAGATAATTTATCACCGCAGATGGTTATCTGTATGTTTTTTATATGAATTTATTTTT
TGCAGGGGGGCATTGTTGGTAGGTGAGAGATCAATTCTGCATTAATGAATCGGCCAACG
CGCGGGGAGAGGCGGTTTGGTATTTGGGCGCTCTTCCGCTTCTCTGCTCACTGACTCGCT
GCGCTCGGTCGTTCCGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTT
ATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGC
CAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGA
GCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAAACCCGACAGGACTATAAAGATA
CCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCCTGCCGCTTAC
CGGATACCTGTCCGCTTCTTCCCTTCGGGAAGCGTGGCGTTTTCTCATAGCTCACGCTG
TAGGTATCTCAGTTCGGTGTAGGTGTTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCC
CGTTCAGCCCAGCCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAG
ACACGACTTATCGCCACTGGCAGCAGCCACTGGTAAACAGGATTAGCAGAGCGAGGTATGT
AGGCGGTGCTACAGAGTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGT
ATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAGAGTTGGTAGCTCTTG
ATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTTGCAAGCAGCAGATTAC
GCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCA
GTGGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTAC
CTAGATCCTTTTTAAATTAATAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAAC
TTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATT
TCGTTTATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTT
ACCATCTGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTT
ATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCTGCAACTTTATC
CGCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCCGCCAGTTAA
TAGTTTTCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCTGTTGG
TATGGCTTCATTACGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTT
GTGCAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGTTGTGCAAGTAAGTTGGCCGC
AGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTTCTTACTGTCATGCCATCCGT
AAGATGCTTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCG
GCGACCGAGTTGCTCTTGCCCCGGCGTCAATACGGGATAAATACCGCGCCACATAGCAGAAC
TTTAAAGTGCTCATCATTGGAACGTTTCTTCGGGGCGAAAACCTCAAGGATCTTACC
GCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTACGATCTTT
TACTTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGG
AATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCTTTTTTCAATATTATGAAG
CATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAA
ACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGCTAAGAAACCAT
TATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTC

SEQ ID NO: 4

>pG222, secuencia completa.

TCGCGCGTTTTCGGTGATGACGGTGA AACCTCTGACACATGCAGCTCCCGGAGACGGTCA
 CAGCTTGTCTGTAAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTG
 TTGGCGGGTGTGGGGCTGGCTTA ACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGC
 ACCATATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAggcg
 ccTCCTCAACCTGTATATTTCGTA AACCCACGCCCAATGGGAGCTGTCTCAGGTTTTGTTCT
 GATTGGTTACGGCGCGTTTTCGCATCATTGTTGAGTTTTCCGCCAGCCCGACGCGCAGTT
 TACCGGTGCCTGGGTGCAGTACATCAGCATGGGGCAAATTCCTTCCATCCCAGATGATTGT
 CGCGGGTGTGATCATGATGGTCTGGGCATATCGTCGCAGCCCACAGCAACACGTTTTCTG
 AGGAACCATGAAACAGTATTTAGA ACTGATGCAAAAAAGTGCTCGACGAAGGCACACAGAA
 AAACGACCGTACCGGAACCGGAACGCTTTCCATTTTTGGTCATCAGATGCGTTTTAACCT
 GCAAGATGGATTCCCCTGGTGCACAACTAAACGTTGCCACCTGCGTTCATCATCCATGA
 ACTGCTGTGGTTTTCTGCAGGGCGACACTAACATIGCTTATCTACACGAAAAACAATGTAC
 CATCTGGGACGAATGGGCCGATGAAAACGGCGACCTCGGGCCAGTGTATGGTAAACAGTG
 GCGCGCCTGGCCAACGCCAGATGGTGTCTATATGACCAGATCACTACGGTACTGAACCA
 GCTGAAAAACGACCCGGATTTCGCGCCGATTATGTTTTAGCGTGGAAACGTAGGCGAACT
 GGATAAAATGGCGCTGGCACCGTGCATGCATTCTCCAGTTCATGTGGCAGACGGCAA
 ACTCTCTTGGCAGCTTTATCAGCGCTCCTGTGACGCTTCCTCGGCCTGCCGTTCAACAT
 TGCCAGCTACGCGTTATTGGTGCATATGATGGCGCAGCAGTGCATCTGGAAGTGGGTGA
 TTTTGTCTGGACCGGTGGCGACACGCATCTGTACAGCAACCATATGGATCAAACCTCATCT
 GCAATTAAGCCGCGAACCCGCTCCGCTGCCGAAGTTGATTATCAAACGTAAACCCGAATC
 CATCTTCGACTACCGTTTTCGAAGACTTTGAGATGAAGGCTACGATCCGATCCGGGCAT
 TAAAGCCCGGTGGCTATCTAATTACGAAACATCCTGCCAGAGCCGACGCCAGTGTGCGT
 CGGTTTTTTTACCCTCCGTTAAATTCCTTCGAGACGCCTTCCCGAAggcgccATTCCGCAT
 TCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGC
 TGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAAACGCCAGGGTTTTCCAGT
 CACGACGTTGTAAAACGACGGCCAGTGCCAAGCTTACTGCTCACAAGAAAAAGGCACGT
 CATCTGACGTGCCTTTTTTATTTGTACTACCCTGTACGATTACTGCAGGTCGACTCTAGA
 TGCATGctcgagTTATACAAACCTGCCAATATTTCAAATATTTAAAATGGAGTTCCTCAT
 TAAGGCGATTTTAGGGCTATAAGGTTCTTCTIIICGTGCTATCGTAGAGATTGCTCATC
 ATCAGCGATCACAAAAGGTTGTAACACCAGATTTTTCAGCCATGGATAAAAAGTAGCGTC
 CATTATCGTATCCACAGGAACAACCCATTTTCGGCTGCATTTCAAAAAAACTTTGGCAAT
 CTTAGGCGTGTACATAGCCTTGAGTCCCCACCCCTTCGCTATAAGCTTTAATGATCCC
 CACACGCTCTTTGTATCTCGTGGTTTTTATGGCTCAATGGCTCACTTTTTACTGTCATC
 ATACAATAAATGCATCAAGCGGATATAGCCTAACCTTGGATGTGTTTTTCTAAAAAATC
 CAAGCCCTCTTTAAATCCTCTTTCAAGGTTATATCGTCTTCTAAAATACAGATCGCTTC
 ATTGAGTTCATATGCATTTTTTCCACAAGGAATAATGACTCGCATAGCACCCAAGCTCCCC
 CAAGCTCATAAACTTCGCATGGTATTTTTAAAGCGTAATAAACTTAGAAAACCTCACTGAT
 GAGATTGGTTGTAATCCCCATGTCTTTGATGTTTTGCGTGTGAAATAAGGGTGTAAATG
 CTTTTTCACTAAGGGGTGCAACCCGCCTTCAAAGTTTTAGAAATAATCGCATCAAAAAT
 TTGCGCTTGGTGGTGGGTGGCATTGATGCTATIGAGTAAAGTTGTGGTGTCTCTAAAAAC
 TAAACCAAATGTATCGCACACTTTTTGATTTAAAGAAATGGCAAAAACACGCATATGtat
 atctctctctctCTCGAGTCAACGGTTTTTCAGCAATCGGTGCAAAAATGCCGAAGTATTGC
 CTCAAGGTAACAGCCCGCGCATCTCGCCGCTGCGCGAAAAATCCAGCCACGCGCCGGCG
 GGCAGCGTGTCCGCTCCGTTGAAGCATTGGTACAAAAACCGCGGGGCGGTTCAAAAATCT
 TCTTCCGGCAAAATGTTTTCTCCAGCAATTCATACGCTACTGCTTTTATTTGGCGGTATTCA
 AGGCTGTGCAACCGGGTTTTTAAACCCATAGACTGCAAAAAATCGTTTCTGGCGGTTTTT
 TGGATGCCTTGCAGGATTTCTGTTGGCGGATGCTGTATTTGGATGAAACCTGATTGGCG
 TGAAGGCGGTATTTGACCAAGGCTTCGGGATAATAAGCCAGCCTGCCAATTTGCTGACA
 TCGTACCAAAATTTGTAATCTTCCGCCAAATCCCGCTCGGTGTGTAACGCAAACCGCCG
 TCAATGACGCTGCGCCTCATAATCATCGTGTGTTGTGTATGGGGTTGCCGAAAGGGAAA
 AAGTCGGCAATGTCTTCGTGTGGGTGGTTTTTCCAAAATTTGCCGTGTTCGTGGTGC
 CGCGCCAGCCGGTTGCCGTCTTTTTCTTCCGACAAAACTTCCAGCCACGCACCCATCGCG
 ATGATGCTGCGGTCTTTTTCCATCTCACCCACGATTTTCTCAATCCAGTCGGGGGCGGCA
 ATATCGTCTGCATCGGTGCGCGCAATATATCCCCCCCCCCCCCGACTTTGCCAATTCA
 TCCAGCCCGATGTTTTAAAGAGGGAATCAGACCGGAATGCGCGGCTGCGCGAGGATGCGG
 ATGCGGCCGCTCTGTCTTGGAAACGCTGGGCAATGGCAAGCGTACCGTCCGTCGAGCCG
 TCATCGACAATCAAAATATCCAAGTTGCGCCAAGTTTGATTACGACGGCGGCTAATGAT
 TGGGCGAAATATTTTTCTACGTTGTAGGCGCAAATCAATACGCTGACTAAAGGCTGCAAT
 TTATTCTCCCGATAGGCACGATGCCGTCTGAAGGCTTCAGACGGCATATGtatatctct

tcttgaaTTCTAACAATTGATTGAATGTATGCAAATAAATGCATACACCATAGGTGTGGT
 TTAATTTGATGCCCTTTTTTCAGGGCTGGAATGTGTAAGAGCGGGGTTATTTATGCTGTTG
 TTTTTTGTACTCGGGAAGGGCTTTACCTCTTCCGCATAAACGCTTCCATCAGCGTTTA
 TAGTTAAAAAATCTTTCGGAAGTGGTTTTGCGCTTACCCCAACCAACAGGGGATTTGCT
 GCTTTCATTGAGCCTGTTTCTCTGCGCGACGTTTCGCGGCGGCGTGTGTTGTGCATCCATC
 TGGATTCTCCTGTCAGTTAGCTTTGGTGGTGTGTGGCAGTTGTAGTCTGAACGAAAACC
 CCCCAGGATTGGCACATTGGCAGCTAATCCGGAATCGCACTTACGGCCAATGCTTCGTTT
 CGTATCACACACCCCAAAGCCTTCTGCTTTGAATGCTGCCCTTCTTCAGGGCTTAATTTT
 TAAGAGCGTACCTTTCATGGTGGTTCAGTGCCTCTGCTGATGTGCTCAGTATCACCGCCA
 GTGGTATTTATGTCAACACCGCCAGAGATAATTTATCACCGCAGATGGTTATCTGTATGT
 TTTTTATATGAATTTATTTTTTGCAGGGGGGCAATGTTTTGGTAGGTGAGAGATCAATTTCT
 GCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTTGCGTATTGGGCGCTCTCCGC
 TTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCCGGCTGCGGCGAGCGGTATCAGCTCA
 CTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAATGTG
 AGCAAAGGCCAGCAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCA
 TAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAA
 CCGACAGGACTATAAAGATACCAGGCGTTTTCCCCCTGGAAGCTCCCTCGTGCCTCTCC
 TGTTCCGACCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGC
 GCTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCGCTCCAAGCT
 GGGCTGTGTGCACGAACCCCCGTTTACGCCCCGACCGCTGCGCCTTATCCGGTAACTATCG
 TCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAG
 GATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTA
 CGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGG
 AAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTT
 TGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTT
 TTCTACGGGGTCTGACGCTCAGTGGAACGAAAACCTCACGTTAAGGGATTTGGTTCATGAG
 ATTATCAAAAAGGATCTTACCTAGATCCTTTAAATTAATAAATGAAGTTTTAAATCAAT
 CTAAAGTATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACC
 TATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGAT
 AACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGTGCAATGATACCGCGAGACCC
 ACGTCAACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAG
 AAGTGGTCTCGCAACTTTATCCGCTCCATCCAGTCTATTAATGTTGCCGGGAAGCTAG
 AGTAAGTAGTTCGCCAGTTAATAGTTTGCAGCAACGTTGTTGCCATTGCTACAGGCATCGT
 GGTGTCACGCTCGTCTGTTGGTATGGCTTCAATCAGCTCCGGTTCCCAACGATCAAGGCG
 AGTTACATGATCCCCCATGTTGTGCAAAAAGCGGTTAGCTCCTTCGGTCCCTCCGATCGT
 TGTCAGAAGTAAGTTGGCCGCAAGTGTATCACTCATGGTTATGGCAGCACTGCATAATTC
 TCTTACTGTGATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGTGACTCAACCAAGTC
 ATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAA
 TACCGCGCCACATAGCAGAACTTTAAAAGTGTCTCATCATTGGAAAACGTTCTTCGGGGCG
 AAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTTCGATGTAACCCACTCGTGCACC
 CAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAG
 GCAAAATGCCGCAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTT
 CCTTTTTCAATATTATTGAAGCATTATCAGGGTTATTGTCTCATGAGCGGATACATATT
 TGAATGTATTTAGAAAAATAAACAAATAGGGGTTCGCGCACATTTCCCCGAAAAGTGCC
 ACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCAC
 GAGGCCCTTTCGTC

SEQ ID NO: 5

>pG317, secuencia completa.

ES 2 700 274 T3

GTACCCAGCTTTTGTTCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCA
TAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGA
AGCATAAAGTGTAAGCCTGGGGTGCCATAATGAGTGAGCTAACTCACATTAATTGCGTTG
CGCTCACTGCCCCGCTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGC
CAACGCGCGGGGAGAGGCGGTTTGCATATTGGGCGCATGCATAAAAACTGTTGTAATTCA
TTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGC
GGCATCAGCACCTTGTGCGCTTGCATATAATATTGCCCATGGTGAAAACGGGGCGAAG
AAGTTGTCCATATTGGCCACGTTTAAATCAAACTGGTGAACTCACCCAGGGATTGGCT
GAGACGAAAACATATTCTCAATAAACCTTTAGGGAAATAGGCCAGGTTTTTACCCTAA
CACGCCACATCTTGCGAATATATGTGTAGAACTGCCGAAAATCGTCGTGGTATTCACTC
CAGAGCGATGAAAACGTTTTAGTTTGTCTCATGGAAAACGGTGTAACAAGGGTGAACACTA
TCCCATATCACCAGCTCACCGTCTTTTATTGCCATACGGAATTCGGATGAGCATTTCATC

AGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAAACCTTGTGCTTATTTTTCTTTACGGTC
 TTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGAC
 TGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCA
 GTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAAT
 ACGCCCGGTAGTGATCTTATTTTATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCA
 ACGTCTCATTTTTGCCCCAAAAGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGA
 TTTATTTATTCTGCGAAGTGATCTTCCGTACAGGTATTTATTGGAAGACGAAAGGGCCT
 CGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGG
 TGGCACTTTTCGGGGAAAATGTGCGCGCCCGCTTCCTGCTGGCGCTGGGCCTGTTTCTGG
 CGCTGGACTTCCCGCTGTTCCGTACAGCAGCTTTTCGCCCACGGCCTTGATGATCGCGGGC
 GCCTTGGCCTGCATATCCCGATTCAACGGCCCCAGGGCGTCCAGAACGGGCTTCAGGGCGC
 TCCCGAAGGTCTCGGGCCGTCTCTTGGGCTTGATCGGCCTTCTTGGCATCTCACGCGCT
 CCTGCGGGCGCCTGTAGGGCAGGCTCATAACCCCTGCCGAACCGCTTTTGTGAGCCGGTGC
 GCCACGGCTTCCGGCGTCTCAACGCGCTTTGAGATTCCAGCTTTTCGGCCAATCCCTGC
 GGTGCATAGGCGCGTGGCTCGACCGCTTGGCGGCTGATGGTGACGTGGCCACTGGTGGC
 CGTCCAGGGCCTCGTAGAACGCCTGAATGCGCGTGTGACGTGCCCTGCTGCCCTCGATG
 CCCCCTTGCAGCCCTAGATCGGCCACAGCGGCCGCAACGTGGTCTGGTCCGCGGGTCACT
 TCGCCTTTGTTGCCGATGAACTCCTTGGCCGACAGCCTGCCGTCTGCGTCAGCGGCACC
 ACGAACCGGGTCACTGTCGGGGTGGTTTCGTACGGTGGATGCTGGCCGTACGATGCGA
 TCCGCCCGTACTTGTCCGCCAGCCACTTGTGCGCTTCTCGAAGAACGCGCCTGCTGT
 TCTTGGCTGGCCGACTTCCACCATTCCGGGCTGGCCGTATGACGTACTCGACCCCAAC
 ACAGCGTCTTGCGCCGCTTCTTGGCAGCAACTCGCGCAGTCGGCCCATCGCTTCATCG
 GTGCTGCTGGCCGCCAGTGTCTGCTTCTTGGCGTCTGCTGGCGTCAGCGTTGGCGCTC
 TCGCGCTCGCGGTAGGCGTGTGAGACTGGCCGCCACGTTGCCCATTTTCGCCAGCTTC
 TTGCATCGCATGATCGCGTATGCCGCCATGCCCTGCCCTTCCCTTTTGGTGTCCAACCGGC
 TCGACGGGGGCGAGCGCAAGGCGGTGCCCTCCGGCGGGCCACTCAATGCTTGAGTATACTCA
 CTAGACTTTGCTTCGAAAAGTCGTGACCGCCTACGGCGGCTGCGGGCCCTACGGGCTTG
 CTCTCCGGGCTTCGCCCTGCGCGGTGCTGCGCTCCCTTGCCAGCCCGTGGATATGTGGA
 CGATGGCCCGAGCGGCCACCGGCTGGCTCGCTTCGCTCGGCCCGTGGACAACCCCTGCTG
 GACAAGCTGATGGACAGGCTGCGCCTGCCACGAGCTTGACCACAGGGATTGCCACCGG
 CTACCCAGCCTTCGACCACATACCCACCGGCTCCAACCTGCGCGGCCCTGCGGCCCTTGC
 ATCAATTTTTTAAATTTCTCTGGGGAAAAGCCTCCGGCCTGCGGCCCTGCGCGCTTCGCT
 TGCCGGTTGGACACCAAGTGGAAAGCGGGTCAAGGCTCGCGCAGCGACCGCGCAGCGGCT
 TGGCCTTGACCGCCTGGAACGACCCAAGCCTATGCGAGTGGGGCAGTCGAAGGCGAAG
 CCCGCCCGCTGCCCCCGAGCCTCACGGCGGCGAGTGGGGGGTTCGAAGGGGGCAGCG
 CCACCTTGGGCAAGGCCGAAGGCCGCGCAGTCGATCAACAAGCCCGGAGGGGCCACTTT
 TTGCCGGAGGGGGAGCCGCGCCGAAGGCGTGGGGGAACCCCGAGGGGTGCCCTTCTTTG
 GGCACCAAAGAAGTATAGATATAGGGCGAAATGCGAAAGACTTAAAAATCAACAACCTTAAAA
 AAGGGGGGTACGCAACAGCTCATTGCGGCACCCCCGCAATAGCTCATTGCGTAGGTTAA
 AGAAAACTGTAATTGACTGCCACTTTTACGCAACGCATAATTGTTGTGTCGCGCTGCCGAA
 AAGTTGAGCTGATTGCGCATGGTGCCGCAACCGTGGCGCACCCCTACCGCATGGAGATAA
 GCATGGCCACGCAGTCCAGAGAAAATCGGCATTCGAAGCCAAGAACAAGCCCGGTCACTGGG
 TGCAACCGAACGCAAAAGCGCATGAGGCGTGGGCCGGGCTTATTGCGAGGAAACCCACGG
 CGGCAATGCTGCTGCATCACCTCGTGGCGCAGATGGGCCACCAGAACGCCGTGGTGGTCA
 GCCAGAAGACTTTCCAAGCTCATCGGACGTTCTTTGCGGACGGTCCAATACGCAGTCA
 AGGACTTGGTGGCCGAGCGCTGGATCTCCGTGCTGAAGCTCAACGGCCCCGGCACCGTGT
 CGGCCTACGTGGTCAATGACCGCGTGGCGTGGGGCCAGCCCCGCGACCAAGTTGCGCCTGT
 CGGTGTTTCACTGCCCGCTGGTGGTTGATCACGACGACCAGGACGAATCGCTGTTGGGGC
 ATGGCGACCTGCGCCGCATCCCAGCCCTGTATCCGGGCGAGCAGCAACTACCGACCGGCC
 CCGGCGAGGAGCCGCCAGCCAGCCCGCATTCGGGCATGGAACCAGACCTGCCAGCCT
 TGACCGAAACGAGGAATGGGAACGGCGGGCAGCAGCGCCTGCCGATGCCGATGAGC
 CGTGTCTTCTGGACGATGGCGAGCCGTTGGAGCCCGACACGGGTACGCTGCCGCGCC
 GGTAGCACTTGGGTTGCGCAGCAACCCGTAAGTGCCTGTTCCAGACTATCGGCTGTAGC
 CGCTCGCCGCCCTATACCTTGTCTGCCCTCCCCCGCTTGCCTCGCGGTGCATGGAGCCGG
 GCCACCTCGACCTGAATGGAAAGCCGGCGGCACCTCGCTAACGGATTACCGTTTTTATCA
 GGCTCTGGGAGGCAGAATAAATGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCT
 GAGCAAACTGGCCTCAGGCATTTGAGAAGCACACGGTCACTGCTTCCGGTAGTCAATA
 AACCGGTAAACCAGCAATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGA
 CCGGGTCAATTTGCTTTCGAATTTCTGCCATTATCCGCTTATTATCACTTATTACGGC
 GTAGCACAGGCGTTAAGGGCACCAATAACTGCCTTAAAAAAATACGCCCCGCCCTGC
 CACTCATCGCAGTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTAACCGC

AATTTTAAACAAAATATTAACGCTTACAATTTCCATTTCGCCATTTCAGGCTGCGCAACTGTT
 GGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAAGGGGGATGTG
 CTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCAGTACACGACGTTGTAAAACGA
 CGGCCAGTGAGCGCGTAATACGACTCACTATAGGGCGAATTGGAGCTCCACCGCGGTG
 GCGGCCGCTCTAGAACTAGTGGATCCCCGGGCTGCAGGAATTCGATATCAAGCTTATCG
 ATACCGTCGACCTCGAGTTAAGTCTCTAATCGATTGTTTTCCAATGGAATGGTTATAAAA
 TCTTTGGTTTTTAGTCTTGAAAATCTTCTAGGATTTCTATGTAAGTTTTTGTATAAATA
 TTATATTGCTTTAATAAAATTTAATATATTTTTATTGCATTTTAAGGTTATTTTTCCATA
 TCTGTTCAACCTTTTTTAAATCCTCCAAACAGTCAATATCTAAACTTGAGCTTTCGTCCA
 TTAAAAATGCTTGGTTTTGCTTTGTAAAAAGCTAGGATTGTTTTAAAAATCTTTTTATCT
 TTAAAAATATAAATGCAACATTGCTCATATAAGTTTTAGGCAATTTTGCCTGGCATAA
 AAGGATATTCATCATTACAAATCCCTGCTAAATCGCCACAATCATTACAAACAAAGGCTT
 TTAGAATTTTATTATCACATTTCGCTTACGCTAATTAGGGCATTTCGATTGCTATTTTTAT
 AAAGATTAAGGCTTCATTAATATGAATATTTGTTCTTAGCGGTGAAGTGGGTTGTAAAA
 AAACCTACATCTTCATAATCTTTATAAAATTTAGAGCATGTAACAGCACTTTATCGCTTG
 TGGTATCATCTTGTGCAAGGCTAATTGGGCGTTTTAAAAATCAACATTTGACTTTTTG
 CATAATTTAAAAATTCATCACTATCACTGCTTACAACAACCTTTACTAATGCTTTTAGCAT
 TTAGTGCAGCTTTGATCGTGTAGTAAATTAAGGTTTATTGTTTAATAAAAACCAAATTTT
 TATTTTTAATACCTTTGAGCCACCACGAGCAGGGATTATTGCTAAGCTCATTTTATATC
 CTTAAAAACTTTTTGTGTGCTGAGTTTAAAAAATCTCCGCTTTGTAAATATTCAAAAA
 TAATTTTGAGCTATCTAAATCTCTAACTTAGCGCTAAATAAATCTTGTTTTTTATGAAT
 AGTGTAAATAGCTTTTAGTATTTTATCACTATTTGCATTAACCTTTTAGTGTATTTTCAAT
 GCCAAGTCTTCCATTTTGTCTTGAGCCAACTAAAAATCCCTGCTGTTTTTAAGTATAAGGC
 CTCTTTTAAAAATACAACCTGAAATTACCTATTATAAAATCAGCATTTTTTTAACAAAGTTAT
 AAAACTCAAAATCTAAGCGATGGAAAAAGCTTAAATCTAGGGTTATTTTTAAACTCTTC
 ATAGCTTTGCAAGATTAATTCAAAACCTAAATCATTATTTGGATAAATAACAATATAATT
 TTTATTACTTTGTATCAGTGTCTTTACTAAATTTGTCTGCTTGATTTTTAATGCTAGTAAT
 TTCAGTTGTAACAGGATGAAACATAAGCAAAGCGTAGTTTTTATAATTTATATCATAATA
 TTTTTTGTCTTCGCTAAGTGAATTTTATTATCGTTTTAAAAGTTCTAAATCAGGCGAACC
 TAGATAAAAAATAGATTTTTTATCTTCTCCAAGCTGCATTAACGCCTTTTTTGCAAACTC
 ATCATTTACTAAAATGAATATGAGCTAGTTTTGATATAGCGTGGCGTAAGCTATCGTCAAT
 AGTTCTGAAAATCTCTCCGCTTCAATAIGCGCTACTAAGATATTATTTAATGCTCCAAC
 AATAGCTGCTGCTAAAGGCTCAATTCATCTCCATGTACTACGATTAATCAGGTTTTAG
 CTCATTTGCATACCTTGAAAATCCATCAATTGTAGTAGCTAAAGCCTTATCAGTTTGATA
 ATATTTATCATAATTTATAAATTCATAAATATTTTTAAAGCCATTTTTATAAAGTTCTTT
 AACTGTATAGCCAAAATTTTTACTTAAGTGCATTCCTGTTGCAAAGATGTAAAGTTCAAA
 TTCGCTTGAGTTTTGCACCCTGTACATTAAGATTTAATCTTAGAATAATCAGCCCTAGA
 GCCTGTTATAAAAAGGATTTTTTTTACGCAAAAATCCTCATAGCTTAACTGAGCATCATT
 TCTATATCTCTTAATGCTTTTTTGCCTAAAATATTTTCAAATTCAGCCGCCTAATTCCA
 CCAAGTCCAGGTCTTTTAAACCCAAATATTATCCATAGATAAAACTTCGCCTTTTTTAATA
 TCTTTAATGCTAACTACACTTGCAAAGGCAAAAATCAATTGTAACCTGTTCTTGTTTAGCC
 GCTTTTTTACTTTTCAATATTTCTCTTATTATAGCCATTTGCTCACTTTGTATAAATTAGC
 TCTTTTAAAGCCTTTGTATCCATAGAACAACCTATAATCAGGGCCACTTCTATGCATACTA
 TCAGTAAAATGTCTTTCAAGCACACAAGCTCCAAGTACAACCTGCACCTAAACACGCAAGA
 TTATCTGTTGTGTGGTTCGCTTAAGCCTACCATAACAAGAAAATCTTTTTTTAACTCAAGC
 ATAGCGTTAATCTTACAAGATTATGCGGGGTTGGGTAAAGATTGGTCGTGTGCATTA
 ACAAAGGAATTTTCAATGCTAATAAGATTTTTTACAGTTGGTTTTTACTTTCAATACTA
 TTCATTCTGTGCTAACTATCATAGGCTTTTTTAAAGGCTGCTATGTGTTAATAAGCGGA
 TAATTATTACACTCACCTGAACCAATCTTAAAGCACTAACTCCCATATCTTCTAAGCGG
 TTCGCACCTGCACGAGAAAAGGTGTGCTAAGATAAACAAGACCTAATTTTTCTGTGTAT
 TCTTTAAGTGTAGCTCATCTTTATAATCCAAAGCACATTTTTTGCATAATCTCATAAATG
 CTTATTTTTGCATTACCAGGAATTACTTTTTTAGCGGCCTTACTCATCTCATCTTCAACA
 ATATGAGTTTGATGCTTTATAATCTTAGCACCTGCGCTAAAGGCTGCATCTACCATAATT
 TTAGCTAGTTCTAAACTGCCATTATGATTAATGCCTATTTTCAGGTACGACTAAGGGTGTCT
 TTTCTTCACTTATGATTATATTTTGTATTTTTATTCTTTTCAATTTATTTTCTCTCTTAG

ES 2 700 274 T3

SEQ ID NO: 6

>pG315, secuencia completa

```
CTAAATTGTAAGCGTTAATATTTTGTAAAATTCGCGTTAAATTTTGTAAATCAGCTC  
ATTTTTAACCAATAGGCCGAAATCGGCAAATCCCTTATAAATCAAAGAATAGACCGA  
GATAGGGTTGAGTGTTGTTCCAGTTTGAACAAGAGTCCACTATTAAAGAACGTGGACTC
```

CAACGTCAAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACC
 CTAATCAAGTTTTTTTTGGGGTCGAGGTGCCGTAAGCACTAAAATCGGAACCCCTAAAAGGGAG
 CCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAAGGAAGGGAAAGAA
 AGCGAAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTACGCTGCGCGTAACCAC
 CACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTCGCCATTCAGGCTGCG
 CAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAAGG
 GGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCAGTCACGACGTTG
 TAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGGAGCTCCA
 CCGCGGTGGCGGCCGCTCTAGAAGTAGTGGATCCCTAGACTGCAATACAAAACACCTGTTT
 CACAATTTGGCAGATCAGCCAAAAAAGTACATTTCTCTTTTACAATACCTAGTTTTTA
 TCATTACTTGAACATAAAGGACTTCTCAAAGCAGTTTTACGATCAGTTATAGTTTTCTGTGCG
 ATGTAAAAACTATAAAATTTAATTTTTTTCAGCTGGTATCGTGAAATATAAAGAGCTCGCTA
 TACCAGCAACTGCATCAGGAAGCATATCTGTATCATCAAAACTTCAAATGATATTTTTG
 ATGGAATATCAACCATTTGAAGGATAGTTTTGCATTATTAATGTATTAATGATACCGCCAC
 CAGGGTGACCTTTGAAGAACAATACTATAACTATTGCCTAAAATAATGTGGGCTCGATTCA
 TAATTGCATTATTAATGACATTAATTTGTTGTTTTCGCATAATACTCTCTTTTCATGGTTAC
 CAGCCCATACAGTCGTACCTGTAAACACAAAAGTTTGGTAAATTAGATGAATTATATTCAT
 TTTGTAATTTTTGTTTGTCAAAAATTAACAATCGATAAGAATAATTTCTTGTGTTTGTCTAT
 TGAATTTTTTGAACCATCCCATTTGCATTTGCTTTAAACTATCACCAATATAGTCTCGTA
 ACTCATGTAATGATGGTTCTAAAGTTAAATAATCTTTTCTAAAAAATGGTAGTTAGCTG
 GATATAGTTTTTCCAGTTATAAACAGATGATGTTCTGTATTTGAAGTGTCTTCATTGA
 TACCATTAATGACATCCTCAAGATAAICTTTACCAATTTTTAAATTAICTGTTTTAATTA
 ATGTATCTCTCCAGTTATATAAAATTTACATATTTCTGCTGAACCATCATCATATAAAATCTA
 TATTTGTTACCGTAACGTTATTAACGAATTTAATTTCTTTTAGTATTGGCACTAAATTA
 CAAATGAATGAGCAGTGTAGAGCTAAGTTAACATTTCAATCTATGCTTTGTTTGTGCTT
 GCTTAACAATTTCTGTACTAAGTCAGCTGGTGTATGGTTATTTATCAATGCAAACGATG
 TAATATTTAACTCTTTCATTTGCTCATCAGTCGGAACATTTCTCCCCAAGCTATATATC
 TTTGTGCTGTAGGATTTCTTCTTCCGATTTAATAATATCCATTAGCTGTGAAGAGTTG
 GAAGATGATGATGATCAACATAAACCTCTAAAGATGGAGCCACTACGTTAATGTTACTT
 TTGTTATATATTTTTACCTTTTATTACTAACACCATTAAAATCAAAGCAGTACTTTTTCA
 CGTCAICTAATCGTGGCGCCACTACAGATAATGATATTGACTCTTTATTTTGTCTGTGTA
 ATAGTTGTTGCGTACCACAAGTTTTGTACCCAAGAGTGTTTTGTAAAAAGAGATGTTTGATT
 GATTAATTTGGCTCTAAATTAACATACTCCTCATCAATAATAGTTTTATTAATATCATTTTT
 TAATAATAGATTGTGATTTTTCTTCTGACATgggtctgttctctcCTCGAGGGGGGGCCCG
 GTACCCAGCTTTTGTTCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCA
 TAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAATTCACACAACATACGAGCCGGA
 AGCATAAAGTGTAAGCCCTGGGGTGCCATAAGAGTGAGCTAACTCACATTAATTGCGTTG
 CGCTCACTGCCCCGCTTTCCAGTCGGGAAACCTGTCTGTGCCAGCTGCATTAATGAATCGGC
 CAACGCGCGGGGAGAGGCGGTTTTGCGTATTGGGCGCTCTTCCGCTTCTCTGCTCACTGAC
 TCGCTGCGCTCGGTCGTTCCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATA
 CGGTTATCCACAGAATCAGGGGATAACCGAGGAAAGAACATGTGAGCAAAGGCCAGCAA
 AAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCT
 GACGACATCACAAAACTGACGCTCAAGTCAGAGGTGGCGAAAACCCGACAGGACTATAA
 AGATAACCAGGCGTTTTCCCTTGGAAAGCTCCCTCGTGCCTCTCTGTTCGGACCTCGCG
 CTTACCGGATACCTGTCCGCTTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCA
 CGCTGTAGGTATCTCAGTTCCGGTGTAGGTGCTTTCGCTCCAAGCTGGGCTGTGTGCACGAA
 CCCCCGTTTCCAGCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCG
 GTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGG
 TATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGG
 ACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGC
 TCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAG
 ATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGAC
 GCTCAGTGGAAACGAAAACTCACGTTAAGGGATTTTTGGTCATGAGATTATCAAAAAAGGATC
 TTCACCTAGATCCTTTTTAAATTAAAAAATGAAGTTTTAAATCAATCTAAAAGTATATATGAG
 TAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGT
 CTATTTCTGTTTATCCATAGTTGCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAG
 GGCTTACCATCTGGCCCCAGTGTGCAATGATACCCGAGACCCACGCTCACCGGCTCCA
 GATTTATCAGCAATAAACAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCTCTGCAACT
 TTATCCGCTCCATCCAGTCTATTAATTTGTTGCCGGAAAGCTAGAGTAAGTAGTTCCGCA
 GTTAATAGTTTGCACAACTGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTTCG
 TTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCC

ES 2 700 274 T3

ATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCCTCCGATCGTTGTCAGAAGTAAGTTG
GCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCA
TCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGT
ATGCGGCGACCGAGTTGCTCTTGCCCGGGCGTCAATACGGGATAATACCGCGCCACATAGC
AGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATC
TTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCA
TCTTTTACTTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAA
AAGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTAT
TGAAGCATTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAA
AATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

SEQ ID NO: 7

CTCGAGgaggaaacagaccATG

SEQ ID NO: 8

5 CTCGAGgaaagaggggacaaactagATG

SEQ ID NO: 9

>pG345, secuencia completa

CTAAATTGTAAGCGTTAATATTTTGTAAAATTCGCGTTAAATTTTTGTAAATCAGCTC
 ATTTTTTAACCAATAGGCCGAAATCGGCCAAAATCCCTTATAAATCAAAAAGAATAGACCGA
 GATAGGGTTGAGTGTGTTCAGTTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTC
 CAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCACTACGTGAACCATCACC
 CTAATCAAGTTTTTTGGGGTTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAG
 CCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAAGGAAGGGAAGAA
 AGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTACGCTGCGCGTAACCAC
 CACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCCATTTCGCCATTACAGGCTGCG
 CAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGG
 GGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTG
 TAAAACGACGGCCAGTGAGCGCGCGTAATACGACTCACTATAGGGCGAATTGGAGCTCCA
 CCGCGGTGGCGGCCGCTCTAGAAGTAGTGGATCCCTAGACTGCAATACAAACACCTGTTT
 CACAATTTGGCAGATCAGCCCCAAAAAGTACATTTCTTTCTTTTACAATACCTAGTTTTTA
 TCATTACTTGAACATAAGGACTTCTCAAAGCAGTTTCACGATCAGTTATAGTTTTCTGTGC
 ATGTA AAAACTATAAATTTAATTTTTTTCAGCTGGTATCGTGAAATATAAAGAGCTCGCTA
 TACCAGCAACTGCATCAGGAAGCATATCTGTATCATCAAACTTCAAATGATATTTTTG
 ATGGAATATCAACCATTGAAGGATAGTTTTGCATTATTAATGTATTAATGATACCGCCAC
 CAGGGTGACCTTTGAAGAACAATCATAACTATGCTTAAATAATGTGGGCTCGATTCA
 TAATTGCATTATTAATGACATTAATTTGTTGTTTCGCATAATACTCTCTTTCATGGTTAC
 CAGCCCATACAGTCGTACCTGTAAACACAAAGTTGGTAAATTAGATGAATTATATTCAT
 TTTGTAATTTTTGTTGTCAAATTAACAATCGATAAGAATAATTTCTTGTGTTGCTAT
 TGAATTTTTTGAAACCATCCCATTGCATTTGCTTAAACTATCACCAATATAGTCTCGTA
 ACTCATGTAATGATGGTTCTAAAGTTAAATAATCTTTTTCTTAAAAAATGGTAGTTAGCTG
 GATATAGTTTTTGGCAGTTATAAACAGATGATGTTCTGTATTTGAAGTGTCTTCATTGA
 TACCATTAATGACATCCTCAAGATAATCTTTACCAATTTTTAAATTATCTGTTTTATTTA
 ATGTATCTCTCCAGTTATATAAATTTACATATTCTGCTGAACCATCATCATATAAATCTA
 TATTTGTTACCGTAACGTTATTAACGAATTTAATTTCTTTTAGTATTGGCACTAAATTAT
 CAAATGAATGAGCAGTGTAGAGCTAAGTTTAAACATTCAATCTATGCTTTGTTGTGCTT
 GCTTAACAATTTCTTGTACTAAGTCAGCTGGTGTATGGTTATTTATCAATGCAAACGATG
 TAATATTTAACTCTTTCATTTGCTCATCAGTCGGAACCTATCTCCCCAAGCTATATATC
 TTTGTGCTGTAGGATTTCTTCTTCCGATTTAATAATATCCATTAGCTGCTGAAGAGTTG
 GAAGAGATGCATGATCAACATAAACCTCTAAAGATGGAGCCACTACGTTTAAATGTTACTT
 TTGTTATATATTTTTACCTTTATTACTAACACCATTAAAATCAAAGCAGTACTTTTCAT
 CGTCATCTAATCGTGGCGCCACTACAGATAATGATATTGACTCTTTATTTTGTCTGTTA
 ATAGTTGTTGCGTACCACAAGTTTGTACCCAAGAGTGTTTTGTA AAAAGAGATGTTTGATT
 GATTAATTGGCTCTAAATTAACATACTCCTCATCAATAATAGTTTTATTAATATCATTTT
 TAATAATAGATTGTGTATTTTTCTTCTGACATctagtttgtccctctttcCTCGAGGGGG
 GGCCCGGTACCCAGCTTTTGTCCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCA
 TGGTCATAGCTGTTTCTGTGTGAAATGTTATCCGCTCACAATTCACACAACATAACGA
 GCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATT

ES 2 700 274 T3

GCGTTGCGCTCACTGCCCGCTTTCAGTCGGGAAACCTGTTCGTGCCAGCTGCATTAATGA
ATCGGCCAACGCGCGGGGAGAGGCGGTTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTC
ACTGACTCGCTGCGCTCGGTCGTTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCG
GTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGC
CAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGC
CCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGA
CTATAAAGATAACCAGGCGTTTTCCCCCTGGAAGCTCCCTCGTGGCTCTCCIGTTCCGACC
CTGCCGCTTACCGGATACCTGTCCGCCTTTCCTCCCTTCGGGAAGCGTGGCGCTTTCAT
AGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTTCGTTCCGCTCCAAGCTGGGCTGTGTG
CACGAACCCCCCGTTACAGCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCC
AACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGA
GCGAGGTATGTAGGCGGTGCTACAGAGTTCCTGAAGTGGTGGCCTAACTACGGCTACACT
AGAAGGACAGTATTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTT
GGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTTGCAAG
CAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGG
TCTGACGCTCAGTGGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAA
AGGATCTTACCTAGATCCTTTTTAAATTAATAAATGAAGTTTTAAATCAATCTAAAGTATA
TATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCG
ATCTGTCATTTTCGTTTCAATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATA
CGGGAGGGCTTACCATCTGGCCCCAGTGTGCAATGATACCGCGAGACCCACGCTCACCG
GCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCT
GCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGT
TCGCCAGTTAATAGTTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGC
TCGTCGTTTTGGTATGGCTTCATTCAGTCCGGTTCCTAACGATCAAGGCGAGTTACATGA
TCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCCTCCGATCGTTGTGAGAAGT
AAGTTGGCCGAGTGTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTC
ATGCCATCCGTAAGATGCTTTTTCTGTGACTGGTGTGACTCAACCAAGTCATTCTGAGAA
TAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAAATACCGCGCCA
CATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACCTCTCA
AGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCT
TCAGCATCTTTTACTTTTACCAGCGTTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCC
GCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCACTCTTCCTTTTTCAA
TATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTGAATGTATT
TAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

SEQ ID NO: 10

CTTTattaaacctactATG

SEQ ID NO: 11

5 CTTTtctcaacctactATG

SEQ ID NO: 12

>pEC3'-(T7)GImS-(T7)NagC-purA_(pG356)

ES 2 700 274 T3

TCGCGCGTTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCA
CAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTG
TTGGCGGGTGTTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGC
ACCATATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCG
CCactagtGTTGAGGAAAACGATTGGCTGAACAAAAAACAGACTGATCGAGGTCATTTTT
GAGTGCAAAAAGTGCTGTAACCTGAAAAAGCGATGGTAGAATCCATTTTTAAGCAAACG
GTGATTTTGAAAAATGGGTAACAACGTCGTCGTAACGGGCTAAATATGTTGTACGCTACCAGG
TAAAGGTAAGATCGTCGATCTTCTGACTGAACGGGCTAAATATGTTGTACGCTACCAGG
CGGTCACAACGCAGGCCATACTCTCGTAATCAACGGTGAAAAAACCGTTCTCCATCTTAT
TCCATCAGGTATTCTCCGCGAGAATGTAACCAGCATCATCGGTAACGGTGTTGTGCTGTC
TCCGGCCGCGCTGATGAAAGAGATGAAAGAAGTGAAGACCGTGGCATCCCCGTTTCGTGA
GCGTCTGCTGCTGTCTGAAGCATGTCCGCTGATCCTTGATTATCACGTTGCGCTGGATAA
CGCGCGTGAGAAAGCGCGTGGCGCGAAAAGCGATCGGCACCACCGGTCGTGGTATCGGGCC
TGCTTATGAAGATAAAGTAGCACGTCGCGGTCTGCGTGTGGCGACTTTTTCGACAAAGA
AACCTTCGCTGAAAAACTGAAAGAAGTGATGGAATATCACAACCTCCAGTTGGTTAACTA

CAGAAGCGATAAAGTTTTTCGCCCATCCCCAGGCCAATCACCAGCGGACTACCAGAACGTG
 CCGCCAGCAGGGTATCCGGGTGACGGGAGTCCATGATCACTGTACCGTACGCACCACGCA
 GCTGCGGGATAGCACGCAGAACGGCCTCACGCAGAGTCCCGCCTTGTTTCAGCTCCCAGT
 TCACCAGATGGGCAATCACTTCGGTGTCCGGTTTCAGAAACGAAGGTATAGCCACGCGCTT
 TTAGCTCTTCACGCAGCGGTTTCATGGTTTTTCGATGATGCCGTTATGCACCACCACAATGT
 GTTCAGAAACATGCGGATGCGCATTCACCTTCTGAAGGTTACCGTGGGTCGCCCAGCGAG
 TGTGAGCAATACCAGTGCCGCCATGCAGAGGATGTTCTTCCGCTGCCTGTGCCAGCATCT
 GGACTTTACCGAGGCGACGCAGGCGGGTCATATGACCTTCTGCATCAACAACGGCCAGAC
 CGGCAGAGTCATATCCGCGGTATTCAGACGACGTAAACCTTCAAGAAGGATTTCTGCTA
 CATCACGTTGCGGATCGCGCCAACAATTCACACATATGtatatctccttcttgaaTTC
 TAACAATTGATTGAATGTATGCAAATAAATGCATACACCATAGGTGTGGTTTTAATTTGAT
 GCCTTTTTTCAGGGCTGGAATGTGTAAGAGCGGGGTTATTTATGCTGTTGTTTTTTTGT
 ACTCGGAAGGGCTTTACCTCTTCCGCATAAACGCTTCCATCAGCGTTTATAGTTAAAA
 AATCTTTTCGGAACGGTTTTGCGCTTACCCCAACCAACAGGGGATTTGCTGCTTTCCATT
 GAGCCTGTTTTCTCTGCGCGACGTTCCGCGCGGCGTGTGTTGTGCATCCATCTGGATTCTCC
 TGTGAGTTAGCTTTGGTGGTGTGTGGCAGTTGTAGTCTGAACGAAAACCCCCGCGATT
 GGCACATTGGCAGCTAATCCGGAATCGCACTTACGGCCAATGCTTCGTTTCGTATCACAC
 ACCCAAAGCCTTCTGCTTTGAATGCTGCCCTTCTTCAGGGCTTAATTTTTAAGAGCGTC
 ACCTTCATGGTGGTCAGTGCCTCTGCTGATGTGCTCAGTATCACCGCCAGTGGTATTTA
 TGTCAACACCGCCAGAGATAATTTATCACCAGAGATGGTTATCTGTATGTTTTTTATATG
 AATTTATTTTTTGCAGGGGGGCATGTTTTGGTAGGTGAGAGATCAATCTGCATTAATGA
 ATCGGCCAACCGCGCGGGGAGAGGCGGTTTTGCGTATTTGGGCGCTCTTCCGCTGCTAGCGGA
 GTGTATACTGGCTTACTATGTTGGCACTGATGAGGGTGTGAGTGAAGTGTTCATGTGGC
 AGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATATGTGATACAGGATATATTCGCT
 TCCTCGCTCACTGACTCGCTACGCTCGGTCGTTTCGACTGCGGCGAGCGGAAATGGCTTAC
 GAACGGGGCGGAGATTTCTTGGAAAGTGCAGGAAGATACTTAACAGGGGAAGTGAAGGG
 CCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCCCCTGACAAGCATCACGAAATCTGAC
 GCTCAAAATCAGTGGTGGCGAAACCCGACAGGACTATAAAGATAACCAGGCGTTTCCCCCTG
 GCGGCTCCCTCGTGCCTCTCTCTGTTTCTGCTTTCGGTTTACCGGTGTCATTCCGCTGT
 TATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCGCGGTAGGCAGTTCCGCTCC
 AAGCTGGACTGTATGCACGAACCCCCGTTTTCAGTCCGACCGCTGCGCCTTATCCGGTAAC
 TATCGTCTTGAGTCCAACCCGAAAGACATGAAAAGCACCCTGGCAGCAGCCACTGGT
 AATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGTTAAGGCTAAACTGAAAGGAC
 AAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCA
 GAGAACCTTCGAAAACCGCCCTGCAAGGCGGTTTTTTCGTTTTTCAGAGCAAGAGATTAC
 GCGCAGACCAAAAACGATCTCAAGAAGATCATCTTATTAATCAGATAAAAATATTTCTAGGC
 ggccgcGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTT
 CACCTAGATCCTTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAAGTATATATGAGTA
 AACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAAGGACCTATCTCAGCGATCTGTCT
 ATTTGCTTTCATCCATAGTTGCTGACTCCCCGTCGTTGATAGATAACTACGATACGGGAGGG
 CTTACCATCTGGCCCCAGTGTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGA
 TTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCTGCAACTTT
 ATCCGCTCCATCCAGTCTAATTAATGTTGCGGGGAAGCTAGAGTAAGTAGTTCGCCAGT
 TAATAGTTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCTGTT
 TGGTATGGCTTCATTACGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCAT
 GTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGTTGTCAGAAGTAAGTTGGC
 CGCAGTGTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTATGCCATC
 CGTAAGATGCTTTTCTGTGACTGGTGAAGTACTCAACCAAGTCAATCTGAGAATAGTGTAT
 GCGGCGACCGAGTTGCTCTTGCCTGGCGTCAATACGGGATAATACCGCGCCACATAGCAG
 AACTTTAAAAGTGTCTCATCATTGAAAACGTTCTTCGGGGCGAAAACCTCAAGGATCTT
 ACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACCTGATCTTCAGCATC
 TTTTACTTTTACCAGCGTTTTCTGGGTGAGCAAAAACAGGAAGGCAAAAATGCCGCAAAAA
 GGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCTTTTTCAATATTATTG
 AAGCATTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGATTTAGAAAAA
 TAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAAC
 CATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCTGTC

SEQ ID NO: 13

>neuC_N-acetylglucosamine-6-phosphate-2-epimerase_GI_15193223_in_pG317

MKKILFITGSRADYSKIKSLMYRVQNSSEFELYIFATGMHLSKNFGYTVKELYKNGFKNI
YEFINYDKYYQTDKALATTIDGFSRYANELKPD LIVVHGDRIEPLAAAI V GALNNILVAH
IEGGEISGTIDDSL RHAI SKLAHIHLV NDEF AKRRLMQLGEDEKSI F I IGSPDLELLNDN
KISLSEAKKYDINYENYALLMFHPVTTEITSIKNQADNLVKALIQSNKNYIVIYPNNDL
GFELILQSYEEFKNNPRFKLFP SLRF EYFITLLKNADF I IGNSSC ILKEALY LKTAGILV
GSRQNGRLGNENTLKV NANSDEILK AINTIHKKQDLFSAKLEILDSSK LFF EYLQSGDFF
KLSTQKVFKDIK

SEQ ID NO: 14

5 >neuB_sialic_acid_synthase_GI_15193222_in_pG317

MKEIKIQNI I ISEEKAPLVVPEIGINHNGSLELAKIMVDAAFSAGAKI IKHQTHIVEDEM
SKAAKVIPGNAKISIYEIMQKCALDYKDELALKEYTEKLGVLVYLPSTPFSRAGANRLEDM
GVSFAFKIGSGECNNYPLIKHIAAFKKPMIVSTGMNSIESIKPTVKILLDNEIPFVLMHTT
NLYPTPHNLVRLNAMLELKKEFSCMVGLSDHTTDNLACLGAVVLGACVLERHFTDSMHR
GPDIVCSMDTKALKELIQSEQMAIIRGNNE SKKAQEQVTIDFAFASVVS IKDIKKE
VLSMDNIWVKRPG LGGISA AEFENILGKKALRDIENDAQLSYEDFA

SEQ ID NO: 15

>neuA_CMP-Neu5Ac_synthase_GI_15193224_in_pG317

MSLAIIPARGGSKGIKKNLVLNNKPLIYYTIKAALNAKSI SKVVVSSDSDEILNYAKS
QNVDILKRPI SLAQDDTTSDK VLLHALKFYKDYEDVVFLQPT SPLRTNIHINEAFNLYKN
SNANALISVSECDNKILKAFVCNDCGLAGICNDEYPFMPRQKLPKTYMSNGAIYILKIK
EFLNNSFLQSKTKHFLMDESSSLDIDCLEDLKKVEQIWKK

10 SEQ ID NO: 16

>AAF42258 lacto-N-neotetraosa biosíntesis glicosilo transferasa LgtA [Neisseria meningitidis MC58].

MPSEAFRRHRAYRENKLQPLVSVLICAYNVEKYFAQSLAAVVNQ TWRNLDILIVDDGSTD
GTLAIAQRFQEQDGRIRILAQPRNSGLIPSLNIGLDELAKSGGGGEYIARTDADDIAAPD
WIEKIVGEMEKDRSIIAMGAWLEVLSEEKDNRLARHHEHGKIWKKPTRHEDIADFFPFG
NPIHNNTMIMRRSVIDGGLRYNTERDWAEDYQFWYDVSKLGRLAYYPEALVKYRLHANQV
SSKYSIRQHEIAQGIQKTARNDFLQSMGFKTRFDSLEYRQIKAVAYELLEKHLPEEDFER
ARRFLYQCFKRTDTLPAGAWLDFAADGRMRRLFTLRQYFGILHRLKLNK

SEQ ID NO: 17

>NP_207619 epítipo lipopoligosacárido 5G8 proteína asociada a la biosíntesis Lex2B [Helicobacter pylori 26695].

MRVFAISLNQKVCDFGLVFRDITLLNSINATHHQAQIFDAIYSKTFEGGLHPLVKKHL
HPYFITQNIKDMGITTNLISEVSKFYALKYHAKFMSLGELG CYASHYSLWEKCIELNEA
ICILEDDITLKEDFKEGLDFLEKHIQELGYIRLMHLLYDASVKSEPLSHKNHEIQERVGI
IKAYSEGVTQGYVITPKIAKVFLKCSRKVVVVDTIMDATFIHG VKNLVLQPFVIADDE
QISTIARKEEPYSPKIALMRELHF KYLKYWQFV

15

SEQ ID NO: 18

> E.coli_WbgO_YP_003500090 supuesta glicosil-transferasa WbgO [cepa de Escherichia coli O55: H7 CB9615].

MIIDEAESAESTEPVSVILPVNKNPFLDEAINSILSQTFSSFEI IIVANCTDDDFYNE
LKHKVNDKIKLIRTNIA YLPYSLNK AIDL SNGEF IARMDSDDI SHPDRFTKQVDFLKNNP
YVDVVG TNAIFIDDKGREINKTKLPEENLDIVKNL PYKCCIVHPSVMFRKKVIASIGGYM

ES 2 700 274 T3

FSNYSEDYELWNRLSLAKIKFQNLPEYLFYRRLHEGQSTAKKNLYMVMVNDLVIKMKCFE
LTGNINYLFGGIRTIASFIYCKYIK

SEQ ID NO: 19

>BAA35319 regulador doble transcripcional de unión al ADN nagC [cepa de Escherichia coli K-12 sustr. W3110].

MTPGGQAQIGNVDLVKQLNSAAVYRLIDQYGPISRIQIAEQSQLAPASVTKITRQLIERG
LIKEVDQASTGGRRASIVTETRNFHAIGVRLGRHDATITLFDLSSKVLAEHYPLPER
TQQTLEHALLNAIAQFIDSYQRKLRELIAISVILPGLVDPDSGKIHYMPHIQVENWGLVE
ALEERFKVTCFVGHDIRSLALAEHYFGASQDCEDSILVRVHRGTGAGIISNGRIFIGRNG
NVGEIGHIQVEPLGERCHCGNFGCLETIAANAAIEQRVNLNLLKQGYQSRVPLDDCTIKTI
CKAANKGDSLASEVIEYVGRHLGKTIATAINLNFNPQKIVIAGEITEADKVLLPAIESCIN
TQALKAFRTNLPVVRSELDHRSAIGAFALVKRAMLNGLLQHLLN

5 SEQ ID NO: 20

>NP_418185 L-glutamina: D-fructosa-6-fosfato aminotransferasa glmS [cepa de Escherichia coli K-12 sustr. MG1655].

MCGIVGAIQRDVAEILLEGLRRLEYRGYDSAGLAVVDAEGHMTRLRRLGKVQMLAQAAE
EHPLHGGTGIAHTRWATHGEPSEVNAHPHVSEHIVVHNGI IENHEPLREELKARGYTFV
SETDTEVIAHLVNWELKQGGTLREAVLRAIPQLRGAYGTVIMDSRHPDTLLAARSGSPLV
IGLGMGENFIASDQLALLPVTRRFIFLEEGDIAEITRRSVNIFDKTGAEVKRQDIESNLQ
YDAGDKGIYRHYMQKEIYEQPNAIKNTLTGRI SHGQVDLSELGPNADLLEL SKVEHIQILA
CGTSYNSGMVSRYWFESLAGIPCDVEIASEFRYRKSAVRRNSLMI TLSQSGETADTLAGL
RLSKELGYLGLSLAICNVPGSSLVRESDLALMTNAGTEIGVASTKAFTTQTLTVLLMLVAKL
SRLKGLDASIEHDIVHGLQALPSRIEQMLSQDKRIEALAEDFS DKHHALFLGRGDQYPIA
LEGALKLKEISYIHAEEAYAAGELKHGPLALIDADMPVIVVAPNNELLEKLSNIEEVRAR
GGQLYVFADQDAGFVSSDNMHI IEMPHVEEVIAPIFYTVPLQLLAYHVALIKGTDVDQPR
NLAKSVTVE

SEQ ID NO: 21

10 > BAF92026 beta-galactósido alfa-2,6-sialiltransferasa [Photobacterium sp. JT-ISH-224].

MKNFLLLTLILLTACNNSEENTQSI I I KNDINKT I I DEEYVNLEPINQSNISFTKHSWVQT
CGTQQLLQTEQNKESISLSVVAPRLDDDEKYCFDFNGVSNKGEKYITKVTLNVVAPSLEVY
VDHASLPTLQQLMDI IKSEENPTAQRYSIAWGRIVPTDEQMKELNITSFALINNHTPADL
VQEIVKQAQTKHRLNVKLSSENTAHSFDNLVPIKELNSFNNTVTNIDLYDDGSAEYVNL
YNWRDTLNLKTDNLKIGKDYLEDVINGINEDT SNTGTSSVYNWQKLYPANYHFLRKDYLT
EPLHELDRDYIGDSLKQMQWDGFKKFNKQQLFLSIVNFDKQKLQNEYNSNLNPNFVFT
GTTVWAGNHEREYYAKQQINVINNAINNESSPHYLGNSYDLFFKGHPPGGIINTLIMQNY
SMVDIPSKI SFEVLMMDMLPDAVAGIASSLYFTIPAEEKIKFIVFTSTETITDRETALRS
PLVQVMIKLGIVKEENVLFWADLPNCETGVCI AV

A continuación, se proporciona la secuencia de ADN en formato Genbank de la nueva configuración de genes diseñados en el locus thyA de Escherichia coli en cepas utilizadas para producir oligosacáridos que contienen N-acetilglucosamina.

15	LOCUS	E680_thyA::2.8RBS_lacZ 5877 bp ADN lineal BCT 04 MAR. 2013
	DEFINICIÓN	cepa de Escherichia coli K-12 sustr. MG1655, genoma completo.
	ACCESO	NC_000913
	VERSIÓN	NC_000913.2 GI: 49175990
	PALABRAS CLAVE	.
20	FUENTE	cepa de Escherichia coli K-12 sustr. MG1655 (desconocido)

ES 2 700 274 T3

	ORGANISMO	cepa de Escherichia coli K-12 sustr. MG1655
		Bacterias; Proteobacterias; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
	REFERENCIA	1 (bases 1 a 4639675)
5	AUTORES	Riley, M., Abe, T., Arnaud, MB, Berlyn, MK, Blattner, FR, Chaudhuri, RR, Glasner, JD, Horiuchi, T., Keseler, IM, Kosuge, T., Mori, H., Perna, NT, Plunkett, G. III, Rudd, K.E., Serres, M.H., Thomas, G.H., Thomson, N.R., Wishart, D. y Wanner, B.L.
	TÍTULO	Escherichia coli K-12: a cooperatively developed annotation snapshot--2005
10	REVISTA	Nucleic Acids Res. 34 (1), 1-9 (2006)
	PUBMED	16397293
	OBSERVACIÓN	Estado de la publicación: solo en línea
	REFERENCIA	2 (bases 1 a 4639675)
15	AUTORES	Blattner, FR, Plunkett, G. III, Bloch, CA, Perna, NT, Burland, V., Riley, M., Collado-Vides, J., Glasner, JD, Rode, CK, Mayhew, GF, Gregor, J., Davis, NW, Kirkpatrick, HA, Goeden, MA, Rose, DJ, Mau, B. y Shao, Y.
	TÍTULO	La secuencia completa del genoma de Escherichia coli K-12
	REVISTA	Science 277 (5331), 1453-1474 (1997)
	PUBMED	9278503
20	REFERENCIA	3 (bases 1 a 4639675)
	AUTORES	Arnaud, M., Berlyn, MKB, Blattner, FR, Galperin, MY, Glasner, JD, Horiuchi, T., Kosuge, T., Mori, H., Perna, NT, Plunkett, G. III, Riley, M., Rudd, K.E., Serres, M.H., Thomas, G.H. y Wanner, B.L.
	TÍTULO	Workshop on Annotation of Escherichia coli K-12
25	REVISTA	Sin publicar
	OBSERVACIÓN	Woods Hole, Massachusetts, el 14-18 Noviembre de 2003 (correcciones de secuencia)
	REFERENCIA	4 (bases 1 a 4639675)
30	AUTORES	Glasner, JD, Perna, NT, Plunkett, G. III, Anderson, B.D., Bockhorst, J., Hu, J.C., Riley, M., Rudd, K.E. y Serres, M.H.
	TÍTULO	ASAP: Escherichia coli K-12 strain MG1655 version m56
	REVISTA	Sin publicar
	OBSERVACIÓN	descarga ASAP del 10 de junio de 2004 (actualizaciones de anotación)
	REFERENCIA	5 (bases 1 a 4639675)
35	AUTORES	Hayashi, K., Morooka, N., Mori, H. y Horiuchi, T.
	TÍTULO	Una comparación de secuencias más precisa entre los genomas de las cepas de Escherichia coli K12 W3110 y MG1655
	REVISTA	Sin publicar
	OBSERVACIÓN	accesiones GenBank AG613214 a AG613378 (correcciones de secuencia)
40	REFERENCIA	6 (bases 1 a 4639675)
	AUTORES	Perna, N. T.

ES 2 700 274 T3

	TÍTULO	Escherichia coli K-12 MG1655 yqjK-rfaE intergenic region, genomic sequence correction
	REVISIÓN	Sin publicar
	OBSERVACIÓN	accesión GenBank AY605712 (correcciones de secuencia)
5	REFERENCIA	7 (bases 1 a 4639675)
	AUTORES	Rudd, KE
	TÍTULO	A manual approach to accurate translation start site annotation: an E. coli K-12 case study
	REVISTA	Sin publicar
10	REFERENCIA	8 (bases 1 a 4639675)
	CONSRM	Proyecto de genoma NCBI
	TÍTULO	Presentación directa
	REVISTA	Enviado (04-MAR-2013) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, EE.UU.
15	REFERENCIA	9 (bases 1 a 4639675)
	AUTORES	Rudd, KE
	TÍTULO	Presentación directa
	REVISTA	Enviado (06-FEB-2013) Departamento de Bioquímica y Biología Molecular, Universidad de Miami Miller School of Medicine, 118 Gautier Bldg., Miami, FL 33136, EE.UU.
20	OBSERVACIÓN	Actualización de la secuencia por el remitente
	REFERENCIA	10 (bases 1 a 4639675)
	AUTORES	Rudd, KE
	TÍTULO	Presentación directa
25	REVISTA	Enviado (24-ABR-2007) Departamento de Bioquímica y Biología Molecular, Universidad de Miami Miller School of Medicine, 118 Gautier Bldg., Miami, FL 33136, EE.UU.
	OBSERVACIÓN	Actualización de anotación de ecogene.org como una colaboración de base de datos múltiple
30	REFERENCIA	11 (bases 1 a 4639675)
	AUTORES	Plunkett, G. III.
	TÍTULO	Presentación directa
	REVISTA	Presentado (07-FEB-2006) Laboratorio de Genética, Universidad de Wisconsin, 425G Henry Mall, Madison, WI 53706-1580, EE.UU.
35	OBSERVACIÓN	Actualizaciones de proteínas por el remitente
	REFERENCIA	12 (bases 1 a 4639675)
	AUTORES	Plunkett, G . III.
	TÍTULO	Presentación directa
40	REVISTA	Presentado (10-JUN-2004) Laboratorio de Genética, Universidad de Wisconsin, 425G Henry Mall, Madison, WI 53706-1580, EE.UU.
	OBSERVACIÓN	Actualización de secuencia por el remitente

ES 2 700 274 T3

5	REFERENCIA	13 (bases 1 a 4639675)
	AUTORES	Plunkett, G . III.
	TÍTULO	Presentación directa
	REVISTA	Presentado (13-OCT-1998) Laboratorio de Genética, Universidad de Wisconsin, 425G Henry Mall, Madison, WI 53706-1580, EE. UU.
	REFERENCIA	14 (bases 1 a 4639675)
	AUTORES	Blattner, F.R. y Plunkett, G. III.
	TÍTULO	Presentación directa
10	REVISTA	Presentado (02-SEP-1997) Laboratorio de Genética, Universidad de Wisconsin, 425G Henry Mall, Madison, WI 53706-1580, EE. UU.
	REFERENCIA	15 (bases 1 a 4639675)
	AUTORES	Blattner, FR y Plunkett, G. III.
	TÍTULO	Presentación directa
15	REVISTA	Presentado (16-JAN-1997) Laboratorio de Genética, Universidad de Wisconsin, 425G Henry Mall, Madison, WI 53706-1580, EE. UU.
20	COMENTARIO	REFSEQ PROVISIONAL: Este registro aún no ha sido sujeto a final revisión de NCBI. La secuencia de referencia es idéntica a U00096. El 24 de junio de 2004, esta versión de secuencia reemplazó gi: 16127994. Las actualizaciones de anotación actuales de U00096 se derivan de EcoGene http://ecogene.org . Las sugerencias para las actualizaciones se pueden enviar al Dr. Kenneth Rudd (krudd@miami.edu). Estas actualizaciones se están generando a partir de una colaboración que también incluye ASAP/ERIC, el Coli Genetic Stock Center, EcoliHub, EcoCyc, RegulonDB y UniProtKB/Swiss-Prot.
		COMPLETO: longitud completa.
25	CARACTERÍSTICAS	Ubicación/Calificadores
	Gen	complemento (<1..245) /gene = "ppdA" /locus_tag = "b2826" /gene_synonym = "ECK2822; JW2794"
30		/db_xref = "EcoGene: EG12081" /db_xref = "GeneID: 945393"
	CDS	complemento (<1..245) /gene = "ppdA" /locus_tag = "b2826" /gene_synonym = "ECK2822; JW2794" /function = "enzima putativa; No clasificado" /GO_component = "GO: 0009289 - pilus" /GO_process = "GO: 0009101 - proceso biosintético de glicoproteína" /note = "proteína A dependiente de la peptidasa prepilina"
35		
40		/codon_start = 1 /transl_table = 11

ES 2 700 274 T3

/product = "proteína hipotética"
 /protein_id = "NP_417303.1"
 /db_xref = "GI: 16130730"
 /db_xref = "ASAP: ABE-0009266"
 5 /db_xref = "UniProtKB/Swiss: P33554"
 /db_xref = "EcoGene: EG12081"
 /db_xref = "GeneID: 945393"
 /translation="MKTQRGYTLIETLVAMLILVMLSASGLYGWQYWQSQRLWQTAS
 QARDYLLYLREDANWHNRDHSISVIREGTLWCLVSSAAGANTCHGSSPLVFPVPRWPEV
 EMSDLTPSLAFFGLRNTAWAGHIRFKNSTGEWWLVVSPWGRRLRLCQQGETEGCL" (SEQ ID NO: 22)

fuente enlace (<1..449,4852 .. >5877)
 10 /organism = "cepa de Escherichia coli K-12 sustr. MG1655"
 /mol_type = "ADN genómico"
 /strain = "K-12 "
 /sub_strain = "MG1655"
 /db_xref = "Taxón: 511145"

15 cebador 346..366
 /note = cagtcagtcaggcgccTTCGGGAAGGCGTCTCGAAGA (SEQ ID NO: 23)
 /label = 0268-THYA-R

característica mixta complemento (388..394)
 /feature_type = "Bucle de horquilla"
 20 /label = Terminador

cebador 400..449
 /note=GGCGTCGGCTCTGGCAGGATGTTTCGTAATTAGATAGCCACCGGCGCTTTtag
 GaaacctactATGACCATGATTACGGATTAC (SEQ ID NO: 24)
 /label="homología del cebador 50bp thyA 3"

cebador 400..483
 /note=GGCGTCGGCTCTGGCAGGATGTTTCGTAATTAGATAGCCACCGGCGCTTTat
 taaacctactATGACCATGATTACGGATTAC (SEQ ID NO: 25)
 25 /label=1389-thyAKANlacZ-R-2-8

cebador 400..483
 /note=GGCGTCGGCTCTGGCAGGATGTTTCGTAATTAGATAGCCACCGGCGCTTTct
 tCaacctactATGACCATGATTACGGATTAC (SEQ ID NO: 26)
 /label=1516-thyAKANlacZ-R-0-8

cebador 400..483
 /note=GGCGTCGGCTCTGGCAGGATGTTTCGTAATTAGATAGCCACCGGCGCTTTtag
 GaaacctactATGACCATGATTACGGATTAC (SEQ ID NO: 27)
 30

ES 2 700 274 T3

/label="1041-thyAKANlacZ-R (4-8)"
 característica mixta complemento(401..407)
 /feature_type="Bucle de horquilla "
 /label=Terminador
 5 Cebador 405..472
 /note=CGGCTCTGGCAGGATGTTTCGTAATTAGATAGCCACCGGCGCTTTaTTaaac (SEQ ID NO: 28)
 ctactATGACCATGAT
 /label = 1394-2/8-F
 gen complemento (unión (429..449,4852..4854))
 /gene = "ThyA"
 10 CDS complemento (unión (429..449,4852..4854))
 /gene = "thyA"
 /note = "ECK2823: JW2795: b2827"
 /codon_start = 1
 /transl_table = 11
 15 /product = "timidilato sintetasa"
 /protein_id = "BAE76896 .1"
 /db_xref = "GI: 85675643"
 /translation="MKQYLELMQKVLDEGTQKNDRTGTGTLISIFGHQMRFNLQDGFPL
 VITKRCHLRSIIHELLWFLQGDNIAYLHENNVTIWDEWADENGDLGPVYGKQWRAP
 TPDGRHIDQITTVLNQLKNDPDSRRIIVSAWNVGELDKMALAPCHAFFQFYVADGKLS
 CQLYQRSCDVFLGLPFNIASYALLVHMMAQQCDLEVGDFVWTGGDTHLYSNHMDQTHL
 QLSREPRPLPKLI IKRKPE SIFDYRFEDFEIEGYDPHPGIKAPVAI " (SEQ
 ID NO: 43)
 RBS 450..461
 20 /label="2.8 RBS"
 fuente 450..3536
 /organism =" Escherichia coli W3110"
 "/mol_type = "ADN genómico"
 /strain = "K-12"
 25 /sub_strain = "W3110"
 /db_xref = "taxón: 316407"
 /note = "sinónimo: cepa de Escherichia coli K12 substr. W3110"
 característica mixta 450..4851
 /feature_type = Inserción
 30 /note = "se origina a partir de KanR-lacZRBS (E403)"
 /label = Insertar

ES 2 700 274 T3

	característica mixta	449^450	<i>/feature type = "sitio de variación de RBS"</i> <i>/label = "C en 0/8"</i>
5	característica mixta	450..453	<i>/feature_type = "sitio de variación de RBS"</i> <i>/label = "CTTC en 0/8"</i>
	característica mixta	451..452	<i>/feature_type = "sitio de variación de RBS"</i> <i>/label = "GG en 4/8"</i>
10	característica mixta	451..452	<i>/feature_type = "sitio de variación de RBS"</i> <i>/label = "TT en 2/8"</i>
	CDS	462 .. 3536	<i>/gene = "lacZ"</i> <i>/note = "ECK0341: JW0335: b0344"</i> <i>/codon_start = 1</i> <i>/transl_table = 11</i> <i>/producto = "beta-D-galactosidasa"</i> <i>/protein_id= "BAE76126.1"</i>
15			<i>/db_xref = "GI: 85674486"</i>
20			

ES 2 700 274 T3

/translation="MTMITDSLAVVLQRRDWDENPGVTQLNRLAAHPPFASWRNSEEAR
TDRPSQQLRSLNGEWRFAWFPAPEAVPESWLECDLPEADTVVVP SNWQMHGYDAPIYT
NVTYPITVNPFFVPTENPTGCYSLTFNVDES WLQEGQTRI IFDGVNSAFHLWCNGRWV
GYGQDSRLPSEFDLSAFLRAGENRLAVMVLRWSDGSYLEDDQDMWRMSGIFRDVSLLHK
PTTQISDFHVATR FNDDFSRAVLEAEVQMC GELRDYLRVTVSLWQGETQVASGTAPFG
GEIIDERGGYADRVT LRLNVENPKLWSAEIPNLYRAVVELHTADGTLIEAEACDVGFR
EVRIENGLLLLLNGKPLLIRGVNRHEHHP LHGQVMDEQTMVQDILLMKQNNFNAVRC SH
YPNHPLWYTLCDRYGLYVVDEANIETHGMVPMNRLTDDPRWLPAMSERVTRMVQRDRN
HPSV I IWSLGNESGHGANHDALYRWIKSVDP SRPVQYEGGGADTTATDIICPMYARVD
EDQPFPAVPKWSIKKWL SLPGETRPLILCEYAHAMGNSLGGFAKYWQAFRQYPRLOGG
FVWDWVDQSLIKYDENGNPWSAYGGDFGDTPNDRQFCMNGLVFADRTPHPALTEAKHQ
QQFFQFRLSGQTI EVTSEYLF RHSDNELLHWMVALDGKPLASGEVPLDVAPQGKQLIE
LPELPQ PESAGQLWLTVRVVQPNATAWSEAGHISAWQQWRLAENLSVTLPAASHAIPH
LTTSEMDFCIELGNKRWQFN RQSGFLSQMWIGDKKQLLTPLRDQFTRAPLDNDIGVSE
ATRIDPNAWVERWKAAGHYQAE AALLQCTADTLADAVLIITTAHAWQHQGKTLFISRKT
YRIDGSGQMAITVDVEVASDTPH PARIGLNCQLAQVAERVNWLGLGPQENYPDRLTAA
CFDRWDLPLSDMYTPYVFPSENGLR CGTRELNYGPHQWRGDFQFNISRYSQQQLMETS
HRHLLHAE EGTWLNIDGFHMGIGGDDSWSPSVSAEFQLSAGRYHYQLVWCQK"
(SEQ ID NO: 29)

		/label = "lacZ+ CDS natural"
	cebador	complemento (1325..1345)
5		/note = TTCAGACGTAGTGTGACGCGA
		/label = 1042-thyAlacZcheck
	cebador	2754..2776
		/note = TTTCTTTCACAGATGTGGATTGG
		/label = "1395-mid lacZ-F"
10	cebador	complemento (2779..2801)
		/note = CGGCGTCAGCAGTTGTTTTTAT
		/label = "1396-mid lacZ-R"
	mutación	2793
		/label = "C en MG1655 lacZ (cambio silencioso)"
15	marca	complemento (3549..3567)

ES 2 700 274 T3

```

                                     /label = "KD13 secuencia de marcas descendente"
fuente                               3549..4851
                                     /organism = "Plásmido plantilla pKD13"
                                     /mol_type = "ADN genómico"
5                                     /db_xref = "taxón: 170493"
cebador                              3549..3568
                                     /label = "0339 Plw-P2b"
unidad de repetición                 3568..3579
                                     /label = "sitio FLP"
10  característica mixta              complemento (3568..3601)
                                     /feature type = "sitio FRT"
                                     /label = "sitio FRT de 34bp"
nota                                 complemento (3568..4789)
                                     /label = "región suprimida tras la introducción de pCP20"
15  unidad de repetición              complemento (3590..3601)
                                     /label = "sitio de Flp"
característica mixta                 complemento (3602..3615)
                                     /feature type = "sitio FRT"
                                     /note = "sitio FRT natural"
20                                     /label = "sitio FRT ascendente"
unidad de repetición                 complemento (3604..3615)
                                     /label = "sitio Flp"
característica mixta                 complemento (3628..4422)
                                     /feature type = "CDS (resistencia KAN)"
25                                     /note = "resistencia a la kanamicina"
                                     /codon_start = 1
                                     /transl_table = 11
                                     /producto = "Tn5 neomicina fosfotransferasa "
                                     /protein_id="AAL02037.1"
30                                     /db_xref = "GI: 15554336"

/translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGR
PVLVFKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDL
LSSHLAPAEEKVSIADAMRRLHTLDPATCPFDPHQAKHRIERARTRMEAGLVDQDDLDE
EHQGLAPAELEFARLKARMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRY
QDIALATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF" (SEQ
ID NO: 30)

```

ES 2 700 274 T3

	cebador	complemento (3677..3696) /label = "0389 KD13 K4"
	unión de cebador	3791..3810 /label = "sitio de cebado común kt"
5	cebador	3791..3810 /label = "cebador 0344 Wanner Kt"
	mutación	3811 /label = "A en wt (cambio silencioso)"
10	cebador	complemento (4242..4261) /label = "cebador 0343 Wanner K2"
	unión de cebador	4261..4280 /label = "sitio de cebado común k2"
	unión de cebador	4352..4371 /label = "sitio de cebado común k1"
15	cebador	4352..4371 /label = "cebador 0342 Wanner K1"
	unidad de repetición	4790..4801 /label = "sitio FLP"
20	marca	complemento (4790..4851) /label = "marca ascendente KD13"
	característica mixta	complemento (4790..4823) /feature type = "sitio FRT" /label = "sitio FRT de 34bp"
25	unidad de repetición	complemento (4812..4823) /label = "sitio Flp"
	cebador	complemento (4832..4851) /label = "0338 P4w-P1b"
	cebador	complemento (4832..4901)
	/note=TCTGGGCATATCGTCGCGAGCCACAGCAACACGTTTCCTGAGGAACCATGAT TCCGGGGATCCGTCGACC (SEQ ID NO: 31)	
30	sitio	/label=1040-thyAKANlacZ-F complemento (4858..4863) /site_type = "sitio de enlace" /label = "thyA RBS"
35	gen	complemento (4861..5736) /gene = "lgt"
	CDS	complemento (4861..5736)

ES 2 700 274 T3

/gene = "lgt"
/note = "ECK2824: JW2796: b2828"
/codon_start = 1
/transl_table = 11
/product = "fosfatidilglicerol-prolipoproteína diacilgliceril transferasa"
/protein_id="BAE76897.1"
/db_xref = "GI: 85675644"

5

/translation="MTSSYLHFPEFDPVIFSIGPVALHWYGLMYLVGFIFAMWLATRR
ANRPGSGWTKNEVENLLYAGFLGVFLGGRIGYVLFYNFPQFMADPLYLFRVWDGGMSF
HGGLIGVIVVMIIFARRTKRSFFQVSDFIAPLIPFGLGAGRLGNFINGELWGRVDPNF
PFAMLFPGSRTEIDILLQTNPQWQSIFDTYGVLP RHPSQLYELLLEGVVLFIIILNLYI
RKPRPMGAVSGLFLIGYGAFRIIVEFFRQPDAQFTGAWVQYISMGQILSIPMIVAGVI
MMVWAYRRSPQQHVS" (SEQ ID NO: 32)

- promotor complemento (4957..4962)
/label = "thyA WEAK -10"
- promotor complemento (4978..4983)
/label = "thyA -35"
- cebador complemento (5076..5099)
/note = cagtcagtcaggcgccTTCCTCAACCTGTATATTCGTAAAC (SEQ ID NO: 33)
/label = 0267-THYA-F
- Sitio complemento (5739..5744)
/site type = "sitio de enlace"
/label = "lgt RBS"
- promotor complemento (5823..5828)
/label = "lgt -10 (fuerte)"

10

15

20

ORIGEN

1 GCAGCGGAAC TCACAAGGCA CCATAACGTC CCCTCCCTGA TAACGCTGAT ACTGIGGTCG
61 CGGTTATGCC AGTTGGCATC TTCACGTAAG TAGAGCAAAT AGTCCC CGCG CTGGCTGGCG
121 GTTIGCCATA GCCGTTGCGA CTGCTGCCAG TATTGCCAGC CATAGAGTCC ACTTGCCTT
181 AGCATGACCA AAATCAGCAT CGCGACCAGC GTTTC AATCA GCGTATAACC ACGTTGTGTT
241 TTCATGCCCG CAGTATGGAG CGAGGAGAAA AAAAGACGAG GGCCAGTTTC TATTCTTCG
301 GCGCATCTTC CGGACTATTT ACGCCGTTGC AGGACGTTGC AAAATTTTCGG GAAGGCGICT
361 CGAAGAATTT AACGGAGGGT AAAAAACCG ACGCACACIG GCGTCGGCTC TGGCAGGATG
421 TTTCGTAATT AGATAGCCAC CGGCGCTTTa ttaaacctac tATGACCATG ATTACGGATT
481 CACTGGCCGT CGTTTTACAA CGTCGTGACT GGGAAAACCC TGGCGTTACC CAACTAATC
541 GCCTTG CAGC ACATCCCCCT TTCGCCAGCT GGC GTAATAG CGAAGAGGCC CGCACCGATC
601 GCCCTTCCCA ACAGTTGCGC AGCCTGAATG GCGAATGGCG CTTTGCCTGG TTTCCGGCAC
661 CAGAAGCGGT GCCGGAAAGC TGGCTGGAGT GCGATCTTCC TGAGGCCGAT ACTGTCGTCG
721 TCCCTCAAA CTGGCAGATG CACGGTTACG ATGCGCCCAT CTACACCAAC GTGACCTATC

ES 2 700 274 T3

781 CCATTACGGT CAATCCGCCG TTTGTTCCCA CGGAGAATCC GACGGGTTGT TACTCGCTCA
 841 CATTTAATGT TGATGAAAGC TGGCTACAGG AAGGCCAGAC GCGAATTATT TTTGATGGCG
 901 TTAACTCGGC GTTTCATCTG TGGTGCAACG GCGCTGGGT CGGTTACGGC CAGGACAGTC
 961 GTTIIGCCGTC TGAATTTGAC CTGAGCGCAT TTTTACGCGC CGGAGAAAAC CGCCTCGCGG
 1021 TGATGGTGCT GCGCTGGAGT GACGGCAGTT ATCTGGAAAG TCAGGATATG TGGCGGATGA
 1081 GCGGCATTTT CCGTGACGTC TCGTTGCTGC ATAAACCGAC TACACAAATC AGCGATTTCC
 1141 ATGTTGCCAC TCGCTTTAAT GATGATTTCA GCCGCGCTGT ACTGGAGGCT GAAGTTCAGA
 1201 TGTGCGGCGA GTTGCCTGAC TACCTACGGG TAACAGTTTC TTTATGGCAG GGTGAAACGC
 1261 AGGTCGCCAG CGGCACCGCG CCTTTCGGCG GTGAAATTAT CGATGAGCGT GGTGGTTATG
 1321 CCGATCGCGT CACACTACGT CTGAACGTCG AAAACCCGAA ACTGTGGAGC GCCGAAATCC
 1381 CGAATCCTA TCGTGGGGT GTTGAACGTC ACACCGCCGA CGGCACGCTG ATTGAAGCAG
 1441 AAGCCTGCGA TGTCGGTTTC CGCGAGGTGC GGATTGAAAA TGGTCTGCTG CTGCTGAACG
 1501 GCAAGCCGTT GCTGATTCTG GCGGTTAAAC GTCACGAGCA TCATCCTCTG CATGGTCAGG
 1561 TCATGGATGA GCAGACGATG GTGCAGGATA TCCTGCTGAT GAAGCAGAAC AACTTTAACG
 1621 CCGTGGCTG TTCGCATTAT CCGAACCATC CGCTGTGGTA CACGCTGTGC GACCCTACG
 1681 GCCTGTATGT GGTGGATGAA GCCAATATTG AAACCCACGG CATGGTCCA ATGAATCGTC
 1741 TGACCGATGA TCCGCGCTGG CTACCGGCGA TGAGCGAACG CGTAACCGCA ATGGTGCAGC
 1801 GCGATCGTAA TCACCCGAGT GTGATCATCT GGTGCTGGG GAATGAATCA GGCCACGGCG
 1861 CTAATCACGA CGCGCTGTAT CGCTGGATCA AATCTGTCTG TCCITCCCGC CCGGTGCGAGT
 1921 ATGAAGGCGG CGGAGCCGAC ACCACGGCCA CCGATATTAT TTGCCCGATG TACGCGCGCG
 1981 TGGATGAAGA CCAGCCCTTC CCGGCTGTGC CGAAATGGTC CATCAAAAAA TGGCTTTCGC
 2041 TACCTGGAGA GACGCGCCCG CTGATCCTTT GCGAATACGC CCACGCGATG GGTAACAGTC
 2101 TTGGCGGTTT CGCTAAATAC TGGCAGGCGT TTCGTCAGTA TCCCGTTTA CAGGGCGGCT
 2161 TCGTCTGGGA CTGGGTGGAT CAGTCGCTGA TTAAATATGA TGAAAACGGC AACCCGTGGT
 2221 CCGCTTACGG CCGTGATTTT GCGGATACGC CGAACGATCG CCAGTTCTGT ATGAACGGTC
 2281 TGGTCTTTCG CGACCGCACG CCGCATCCAG CGCTGACGGA AGCAAAAACAC CAGCAGCAGT
 2341 TTTTCCAGTT CCGTTTATCC GGGCAAAACA TCGAAGTGAC CAGCGAATAC CTGTTCCGTC
 2401 ATAGCGATAA CGAGCTCCTG CACTGGATGG TGGCGCTGGA TGGTAAGCCG CTGGCAAGCG
 2461 GTGAAGTGCC TCTGGATGTC GCTCCACAAG GTAAACAGTT GATTGAACTG CCTGAACTAC
 2521 CGCAGCCGGA GAGCGCCGGG CAACTCTGGC TCACAGTACG CGTAGTGAA CCGAACCGCA
 2581 CCGCATGGTC AGAAGCCGGG CACATCAGCG CCTGGCAGCA GTGGCGCTG CCGGAAAACC
 2641 TCAGTGTGAC GCITCCCGCC GCGTCCACG CCATCCCGCA TCTGACCACC AGCGAAATGG
 2701 ATTTTTGCAT CGAGCTGGGT AATAAGCGTT GGCAATTTAA CCGCCAGTCA GGCTTCTTT
 2761 CACAGATGTG GATTGGCGAT AAAAAACAAC TGtTGACGCC GCTGCGCGAT CAGTTTACCC
 2821 GTGCACCGCT GGATAACGAC ATTGGCGTAA GTGAAGCGAC CCGCATTGAC CCTAACGCCT
 2881 GGGTCGAACG CTGGAAGGCG GCGGGCCATT ACCAGGCCGA AGCAGCGTTG TTGCAGTGCA
 2941 CCGCAGATAC ACTTGCTGAT GCGGTGCTGA TTACGACCGC TCACGCGTGG CAGCATCAGG
 3001 GGAAAACCTT ATTTATCAGC CGGAAAACCT ACCGGATTGA TGGTAGTGGT CAAATGGCGA
 3061 TTACCGTTGA TGTTGAAGTG GCGAGCGATA CACCGCATCC GCGCGGATT GGCCTGAACT
 3121 GCCAGCTGGC GCAGGTAGCA GAGCGGGTAA ACTGGCTCGG ATTAGGGCCG CAAGAAAACCT
 3181 ATCCCGACCG CCTTACTGCC GCCTGTTTTG ACCGCTGGGA TCTGCCATTG TCAGACATGT
 3241 ATACCCCGTA CGTCTTCCCG AGCGAAAACG GTCTGCGCTG CCGGACGCGC GAATTGAATT
 3301 ATGGCCACCA CCAGTGGCGC GGCGACTTCC AGTTCAACAT CAGCCGCTAC AGTCAACAGC
 3361 AACTGATGGA AACCAGCCAT CGCCATCTGC TGCACGCGGA AGAAGGCACA TGGTGAATA
 3421 TCGACGGTTT CCATATGGGG ATTGGTGGCG ACGACTCCTG GAGCCCGTCA GTATCGGCGG
 3481 AATTCCAGCT GAGCGCCGGT CGTACCATT ACCAGTTGGT CTGGTGTCAA AAATAAGCCG
 3541 CCGCTTATG TAGGCTGGAG CTGCTTCGAA GTTCTTATAC TTTCTAGAGA ATAGGAACTT
 3601 CGGAATAGGA ACTTCAAGAT CCCCTTATTA GAAGAACTCG TCAAGAAAGC GATAGAAGGC
 3661 GATGCGCTGC GAATCGGGAG CGGCGATACC GTAAAAGCAG AGGAAGCGGT CAGCCCATT
 3721 GCCGCCAAGC TCTTCAGCAA TATCACGGGT AGCCAACGCT ATGTCTCTGAT AGCGGTCCGC
 3781 CACACCCAGC CGGCCACAGT CGATGAATCC tGAAAAGCGG CCATTTTCCA CCATGATATT
 3841 CGGCAAGCAG GCATCGCCAT GGGTACGAC GAGATCCTCG CCGTGGGCA TCGCGCCCTT
 3901 GAGCTTGGCG AACAGTTCGG CTGGCGCGAG CCCCTGATGC TCTTCGTCCA GATCATCTG
 3961 ATCGACAAGA CCGGCTTCCA ICCGAGTACG TGCTCGCTCG ATGCGAIGTT TCGCTTGGTG
 4021 GTCGAATGGG CAGGTAGCCG GATCAAGCGT ATGCAGCCGC CGCATTGCAT CAGCCATGAT
 4081 GGATACTTTC TCGGCAGGAG CAAGGTGAGA TGACAGGAGA TCTTGCCCCG GCACTTCGCC
 4141 CAATAGCAGC CAGTCCCTTC CCGCTTCACT GACAACGTCG AGCACAGCTG CGCAAGGAAC
 4201 GCCCGTCGTG GCCAGCCACG ATAGCCGCGC TGCTCTGTCC TGCAGTTTAT TCAGGGCACC
 4261 GGACAGGTCG GTCTTGACAA AAAGAACCAG GCGCCCTGCG GCTGACAGCC GGAACACGGC
 4321 GGCATCAGAG CAGCCGATTG TCTGTTGTGC CCAGTCATAG CCGAATAGCC TCTCCACCCA
 4381 AGCGGCCGGA GAACCTGCGT GCAATCCATC TTGTTCAATC ATGCGAAACG ATCCTCATCC

ES 2 700 274 T3

4441 TGTCTCTTGA TCAGATCTTG ATCCCCTGCG CCATCAGATC CTTGGCGGCA AGAAAGCCAT
4501 CCAGTTTACT TTGCAGGGCT TCCCAACCTT ACCAGAGGGC GCCCCAGCTG GCAATTCCGG
4561 TTCGCTTGCT GTCCATAAAA CCGCCCAGTC TAGCTATCGC CATGTAAGCC CACTGCAAGC
4621 TACCTGCTTT CTCTTTGCGC TTGCGTTTTT CCTTGTCCAG ATAGCCCAGT AGCTGACATT
4681 CATCCGGGGT CAGCACCGTT TCTGCGGACT GGCTTTTCTAC GTGTTCCGCT TCCTTTAGCA
4741 GCCCTTGCGC CCTGAGTGCT TGCGGCAGCG TGAGCTTCAA AAGCGTCTG AAGTTCCTAT
4801 ACTTTCTAGA GAATAGGAAC TTCGAACTGC AGGTCGACGG ATCCCCGGAA TCATGGTTCC
4861 TCAGGAAACG TGTTGCTGTG GGCTGCGACG ATATGCCAG ACCATCATGA TCACACCCCG
4921 GACAATCATC GGGATGGAAA GAAITTTGCC CATGCTGATG TACIGCACCC AGGCACCCGT
4981 AACTGCGCG TCGGGCTGGC GGAAAACTC AACAATGATG CGAAACGCGC CGTAACCAAT
5041 CAGGAACAAA CCTGAGACAG CTCCCATTGG GCGTGGTTTA CGAATATACA GGTTGAGGAT
5101 AATAAACAGC ACCACACCTT CCAGCAGCAG CTCGTAAAGC TGTGATGGGT GGCGCGGCAG
5161 CACACCGTAA GTGTCGAAA TGGATTGCCA CTGCGGGTTG GTTTGCAGCA GCAAAATATC
5221 TTCTGTACGG GAGCCAGGGA ACAGCATGGC AAACGGGAAG TTCGGTCAA CGCGGCCCA
5281 CAATTCACCG TTAATAAAGT TGCCCAGACG CCCGGCACCA AGACCAAACG GAATGAGTGG
5341 TGCGATAAAA TCAGAGACCT GGAAGAAGGA ACGTTTTAGTA CGGCGGGCGA AGATAATCAT
5401 CACCACGATA ACGCCAATCA GGCCGCCGTG GAAAGACATG CCGCCGTCCC AGACACGGAA
5461 CAGATACAGC GGATCGGCCA TAAACTGCGG GAAATTGTAG AACAGAACAT AACCAATACG
5521 TCCCCCGAGG AAGACCCGA GGAAGCCCGC ATAGAGTAAG TTTTCAACTT CATTITTTGGT
5581 CCAGCCGCTG CCCGGACGAT TCGCCCGTCG TGTTGCCAGC CACATTGCAA AAATGAAACC
5641 CACCAGATAC ATCAGGCCGT ACCAGTGAAG CGCCACGGGT CCTATTGAGA AAATGACCGG
5701 ATCAAATCC GGAAAATGCA GATAGCTACT GGTCACTGT CACCACAAGT TCTTGTATT
5761 TCGCTGAAAG AGAACAGCGA TTGAAATGCG CGCCGACGGT TTCAGGCGCT CCAAAGGTGC
5821 GAATAATAGC ACAAGGGGAC CTGGCTGGTT GCCGGATACC GTTAAAAGAT ATGTATA
(SEQ ID NO: 34)

Se proporciona a continuación la secuencia de ADN en formato Genbank de la configuración de genes en el locus de *Escherichia coli* nan, y los detalles de los puntos finales de eliminación encontrados en las cepas diseñadas por ingeniería genética E1017 y E1018.

5	LOCUS	W3110_nanRATEKyhchH_region 5861 pb ADN lineal BCT 19FEB. 2009
	DEFINICIÓN	cepa de <i>Escherichia coli</i> K-12 sustr. W3110 cepa K-12.
	ACCESO	AC_000091
	VERSIÓN	AC_000091.1 GI: 89106884
	PALABRAS CLAVE	.
10	FUENTE	cepa de <i>Escherichia coli</i> K-12 sustr. W3110 (desconocido)
	ORGANISMO	cepa de <i>Escherichia coli</i> K-12 sustr. W3110 Bacterias; Proteobacterias; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; <i>Escherichia</i> .
	REFERENCIA	1
15	AUTORES	Riley, M., Abe, T., Arnaud, MB, Berlyn, MK, Blattner, FR, Chaudhuri, RR, Glasner, JD, Horiuchi, T., Keseler, IM, Kosuge, T., Mori, H., Perna, NT, Plunkett, G. III, Rudd, K.E., Serres, M.H., Thomas, G.H., Thomson, N.R., Wishart, D. y Wanner, B.L.
	TÍTULO	<i>Escherichia coli</i> K-12: a cooperatively developed annotation snapshot--2005 –
20	REVISTA	Nucleic Acids Res. 34 (1), 1-9 (2006)
	PUBMED	16397293
	OBSERVACIÓN	Estado de la publicación: solo en línea
	REFERENCIA	2 (bases 1 a 4646332)

ES 2 700 274 T3

AUTORES	Hayashi, K., Morooka, N., Yamamoto, Y., Fujita, K., Isono, K., Choi, S. , Ohtsubo, E., Baba, T., Wanner, BL, Mori, H. y Horiuchi, T.
TÍTULO	Highly accurate genome sequences of Escherichia coli K-12 strains MG1655 and W3110
5 REVISTA	Mol. Syst. Biol. 2, 2006 (2006)
PUBMED	16738553
REFERENCIA	3
10 AUTORES	Yamamoto, Y., Aiba, H., Baba, T., Hayashi, K., Inada, T., Isono, K., Itoh, T., Kimura, S., Kitagawa, M., Makino, K., Miki, T., Mitsushashi, N., Mizobuchi, K., Mori, H., Nakade, S., Nakamura, Y., Nashimoto, H., Oshima, T., Oyama, S., Saito, N., Sampei, G., Satoh, Y., Sivasundaram, S., Tagami, H., Takahashi, H., Takeda, J., Takemoto, K., Uehara, K., Wada, C., Yamagata, S. y Horiuchi, T.
15 TÍTULO	Construction of a contiguous 874-kb sequence of the Escherichia coli -K12 genome corresponding to 50.0-68.8 min on the linkage map and analysis of its sequence features
REVISTA	DNA Res. 4 (2), 91-113 (1997)
PUBMED	9205837
REFERENCIA	4
20 AUTORES	Itoh, T., Aiba, H., Baba, T., Hayashi, K., Inada, T., Isono, K., Kasai, H., Kimura, S., Kitakawa, M., Kitagawa, M., Makino , K., Miki, T., Mizobuchi, K., Mori, H., Mori, T., Motomura, K., Nakade, S., Nakamura, Y., Nashimoto, H., Nishio, Y., Oshima, T., Saito, N., Sampei, G., Seki, Y., Sivasundaram, S., Tagami, H., Takeda, J., Takemoto, K., Wada, C., Yamamoto, Y. y Horiuchi, T.
25 TÍTULO	A 460-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 40.1-50.0 min region on the linkage map
REVISTA	DNA Res. 3 (6), 379-392 (1996)
PUBMED	9097040
REFERENCIA	5
30 AUTORES	Aiba, H., Baba, T., Hayashi, K., Inada, T., Isono, K., Itoh, T., Kasai, H., Kashimoto, K., Kimura, S., Kitakawa, M., Kitagawa , M., Makino, K., Miki, T., Mizobuchi, K., Mori, H., Mori, T., Motomura, K., Nakade, S., Nakamura, Y., Nashimoto, H., Nishio, Y., Oshima, T., Saito, N., Sampei, G., Seki, Y., Sivasundaram, S., Tagami, H., Takeda, J., Takemoto, K., Takeuchi, Y., Wada, C., Yamamoto, Y. y Horiuchi, T.
35 TÍTULO	A 570-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 28.0-40.1 min region on the linkage map
REVISTA	DNA Res. 3 (6), 363-377 (1996)
PUBMED	9097039
REFERENCIA	6
40 AUTORES	Arn, E.A. y Abelson, J.N.
TÍTULO	The 2'-5' RNA ligase of Escherichia coli. Purification, cloning, and genomic disruption
REVISTA	J. Biol. Chem. 271 (49), 31145-31153 (1996)
PUBMED	8940112
45 REFERENCIA	7

ES 2 700 274 T3

5	<p>AUTORES Oshima, T., Aiba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Ikemoto, K., Inada, T., Itoh, T., Kajihara, M., Kanai, K., Kashimoto, K., Kimura, S., Kitagawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nashimoto, H., Nishio, Y., Saito, N., Sampei, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y., Yano, M. y Horiuchi, T.</p> <p>TÍTULO A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map</p> <p>REVISTA DNA Res. 3 (3), 137-155 (1996)</p> <p>PUBMED 8905232</p>
10	<p>REFERENCIA 8</p> <p>AUTORES Fujita, N., Mori, H., Yura, T. e Ishihama, A.</p> <p>TÍTULO Systematic sequencing of the Escherichia coli genome: analysis of the 2.4-4.1 min (110,917-193,643 bp) region</p> <p>REVISTA Nucleic Acids Res. 22 (9), 1637-1639 (1994)</p>
15	<p>PUBMED 8202364</p> <p>REFERENCIA 9</p> <p>AUTORES Janosi, L., Shimizu, I. y Kaji, A.</p> <p>TÍTULO Ribosome recycling factor (ribosome releasing factor) is essential for bacterial growth</p>
20	<p>JOURNAL Proc. Natl Acad Sci. USA 91 (10), 4249-4253 (1994)</p> <p>PUBMED 8183897</p> <p>REFERENCIA 10</p> <p>AUTORES Allikmets, R., Gerrard, B., Court, D. y Dean, M.</p>
25	<p>TÍTULO Cloning and organization of the abc and mdl genes of Escherichia coli: relationship to eukaryotic multidrug resistance</p> <p>JOURNAL Gene 136 (1-2), 231-236 (1993)</p> <p>PUBMED 7904973</p> <p>REFERENCIA 11</p>
30	<p>AUTORES van Heeswijk, WC, Rabenberg, M., Westerhoff, HV y Kahn, D.</p> <p>TÍTULO The genes of the glutamine synthetase adenylylation cascade are not regulated by nitrogen in Escherichia coli</p> <p>REVISTA Mol. Microbiol. 9 (3), 443-457 (1993)</p> <p>PUBMED 8412694</p>
35	<p>REFERENCIA 12</p> <p>AUTORES Zhao, S., Sandt, CH, Feulner, G., Vlazny, DA, Gray, JA y Hill, C.W.</p> <p>TÍTULO Rhs elements of Escherichia coli K-12: complex composites of shared and unique components that have different evolutionary histories</p> <p>REVISTA J. Bacteriol. 175 (10), 2799-2808 (1993)</p> <p>PUBMED 8387990</p>
40	<p>REFERENCIA 13</p> <p>AUTORES Yamada, M., Asaoka, S., Saier, MH Jr. y Yamada, Y.</p>

ES 2 700 274 T3

	TÍTULO	Characterization of the <i>gcd</i> gene from <i>Escherichia coli</i> K-12 W3110 and regulation of its expression
	REVISTA	J. Bacteriol. 175 (2), 568-571 (1993)
	PUBMED	8419307
5	REFERENCIA	14
	AUTORES	Cormack, R.S. y Mackie, G.A.
	TÍTULO	Structural requirements for the processing of <i>Escherichia coli</i> 5 S ribosomal RNA by RNase E in vitro
	REVISTA	J. Mol. Biol. 228 (4), 1078-1090 (1992)
10	PUBMED	1474579
	REFERENCIA	15
	AUTORES	Gervais, F.G. y Drapeau, G.R.
	TÍTULO	Identification, cloning, and characterization of <i>rcsF</i> , a new regulator gene for exopolysaccharide synthesis that suppresses the division mutation <i>ftsZ84</i> in <i>Escherichia coli</i> K-12
15	REVISTA	J. Bacteriol. 174 (24), 8016-8022 (1992)
	PUBMED	1459951
	REFERENCIA	16
	AUTORES	Yamanaka, K., Ogura, T., Niki, H. y Hiraga, S.
20	TÍTULO	Identification and characterization of the <i>smbA</i> gene, a suppressor of the <i>mukB</i> null mutant of <i>Escherichia coli</i>
	JOURNAL	J. Bacteriol. 174 (23), 7517-7526 (1992)
	PUBMED	1447125
	REFERENCIA	17
25	AUTORES	Condon, C., Philips, J., Fu, ZY, Squires, C. y Squires, C.L.
	TÍTULO	Comparison of the expression of the seven ribosomal RNA operons in <i>Escherichia coli</i>
	REVISTA	EMBO J. 11 (11), 4175-4185 (1992)
	PUBMED	1396599
30	REFERENCIA	18
	AUTORES	Arnqvist, A., Olsen, A. , Pfeifer, J., Russell, DG y Normark, s.
	TÍTULO	The <i>Crl</i> protein activates cryptic genes for curli formation and fibronectin binding in <i>Escherichia coli</i> HB101
	REVISTA	Mol. Microbiol. 6 (17), 2443-2452 (1992)
35	PUBMED	1357528
	REFERENCIA	19
	AUTORES	Talarico, TL, Ray, PH , Dev, IK, Merrill, BM y Dallas, W.S.
	TÍTULO	Cloning, sequence analysis, and overexpression of <i>Escherichia coli</i> <i>folK</i> , the gene coding for 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase
40	REVISTA	J. Bacteriol. 174 (18), 5971-5977 (1992)

ES 2 700 274 T3

	PUBMED	1325970
	REFERENCIA	20
	AUTORES	Li, S.J. y Cronan, J.E. Jr.
5	TÍTULO	The genes encoding the two carboxyltransferase subunits of Escherichia coli acetyl-CoA carboxylase
	REVISTA	J. Biol. Chem. 267 (24), 16841-16847 (1992)
	PUBMED	1355089
	REFERENCIA	21
10	AUTORES	Yura, T., Mori, H., Nagai, H., Nagata, T., Ishihama, A., Fujita, N., Isono, K., Mizobuchi, K. y Nakata, A.
	TÍTULO	Systematic sequencing of the Escherichia coli genome: analysis of the 0-2.4 min region
	REVISTA	Nucleic Acids Res. 20 (13), 3305-3308 (1992)
	PUBMED	1630901
15	REFERENCIA	22
	AUTORES	Ghosh, SK, Biswas, SK Paul K. y Das, J.
	TÍTULO	Nucleotide and deduced amino acid sequence of the recA gene of Vibrio cholerae
	REVISTA	Nucleic Acids Res. 20 (2), 372 (1992)
	PUBMED	1741267
20	REFERENCIA	23
	AUTORES	Smallshaw, J.E. y Kelln, R.A.
	TÍTULO	Cloning, nucleotide sequence and expression of the Escherichia coli K-12 pyrH gene encoding UMP kinase
	JOURNAL	Genetics (Life Sci. Adv.) 11, 59-65 (1992)
25	REFERENCIA	24
	AUTORES	O'Neill, G.P., Grygorczyk, R., Adam, M. y Ford-Hutchinson, A.W.
	TÍTULO	The nucleotide sequence of a voltage-gated chloride channel from the electric organ of Torpedo californica
	REVISTA	Biochim . Biofis Acta 1129 (1), 131-134 (1991)
30	PUBMED	1721838
	REFERENCIA	25
	AUTORES	Kajie, S., Ideta , R., Yamato, I. y Anraku, Y.
	TÍTULO	Molecular cloning and DNA sequence of dniR, a gene affecting anaerobic expression of the Escherichia coli hexaheme nitrite reductase
35	JOURNAL	FEMS Microbiol. Lett. 67 (2), 205-211 (1991)
	PUBMED	1663890
	REFERENCIA	26
	AUTORES	Hershfield, MS, Chaffee, S., Koro-Johnson, L., Mary, A., Smith, AA y Short, S.A.
40	TÍTULO	Use of site-directed mutagenesis to enhance the epitope-shielding effect of covalent modification of proteins with polyethylene glycol

ES 2 700 274 T3

	REVISTA	Proc. Natl Acad Sci. U.S.A. 88 (16), 7185-7189 (1991)
	PUBMED	1714590
	REFERENCIA	27
	AUTORES	Shimizu, I. y Kaji, A.
5	TÍTULO	Identification of the promoter region of the ribosome-releasing factor cistron (frr)
	REVISTA	J. Bacteriol. 173 (16), 5181-5187 (1991)
	PUBMED	1860827
	REFERENCIA	28
	AUTORES	Poulsen, LK, Refn, A., Molin, S. y Andersson, P.
10	TÍTULO	The gef gene from Escherichia coli is regulated at the level of translation
	REVISTA	Mol. Microbiol. 5 (7), 1639-1648 (1991)
	PUBMED	1943701
	REFERENCIA	29
	AUTORES	Poulsen, LK, Refn, A., Molin, S. y Andersson, P.
15	TÍTULO	Topographic analysis of the toxic Gef protein from Escherichia coli
	REVISTA	Mol. Microbiol. 5 (7), 1627-1637 (1991)
	PUBMED	1943700
	REFERENCIA	30
	AUTORES	Kawamukai, M., Utsumi, R., Takeda, K., Higashi, A., Matsuda, H., Choi, YL y Komano, T.
20	TÍTULO	Nucleotide sequence and characterization of the sfs1 gene: sfs1 is involved in CRP*-dependent mal gene expression in Escherichia coli
	REVISTA	J. Bacteriol. 173 (8), 2644-2648 (1991)
	PUBMED	2013578
25	REFERENCIA	31
	AUTORES	Hulton, C.S., Higgins, C.F. y Sharp, P.M.
	TÍTULO	ERIC sequences: a novel family of repetitive elements in the genomes of Escherichia coli, Salmonella typhimurium and other enterobacteria
	JOURNAL	Mol. Microbiol. 5 (4), 825-834 (1991)
30	PUBMED	1713281
	REFERENCIA	32
	AUTORES	Munro, AW, Ritchie, GY , Lamb, AJ, Douglas, RM y Booth, I.R.
	TÍTULO	The cloning and DNA sequence of the gene for the glutathione-regulated potassium-efflux system KefC of Escherichia coli
35	REVISTA	Mol. Microbiol. 5 (3), 607-616 (1991)
	PUBMED	2046548
	REFERENCIA	33
	AUTORES	Arigoni, F., Kaminski, PA, Hennecke, H. y Elmerich, C.

	TÍTULO	Nucleotide sequence of the fixABC region of Azorhizobium caulinodans ORS571: similarity of the fixB product with eukaryotic flavoproteins, characterization of fixX, and identification of nifW
	JOURNAL	Mol. Gen. Genet. 225 (3), 514-520 (1991)
5	PUBMED	1850088
	REFERENCIA	34
	AUTORES	Mattick, JS, Anderson, BJ, Cox, PT, Dalrymple, BP, Bills, MM, Hobbs, M. y Egerton, J.R.
10	TÍTULO	Gene sequences and comparison of the fimbrial subunits representative of Bacteroides nodosus serotypes A to I: class I and class II strains
	REVISTA	Mol. Microbiol. 5 (3), 561-573 (1991)
	PUBMED	1675419
	REFERENCIA	35
	AUTORES	Company, M., Arenas, J. y Abelson, J.
15	TÍTULO	Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes
	REVISTA	Nature 349 (6309), 487-493 (1991)
	PUBMED	1992352
	REFERENCIA	36
20	AUTORES	Umeda, M. y Ohtsubo, E.
	TÍTULO	Four types of IS1 with differences in nucleotide sequence reside in the Escherichia coli K-12 chromosome
	REVISTA	Gene 98 (1), 1-5 (1991)
	PUBMED	1849492
25	REFERENCIA	37
	AUTORES	Hirvas, L., Koski, PAG. y Vaara, M.
	TÍTULO	The ompH gene of Yersinia enterocolitica: cloning, sequencing, expression, and comparison with known enterobacterial ompH sequences
	REVISTA	J. Bacteriol. 173 (3), 1223-1229 (1991)
30	PUBMED	1991717
	REFERENCIA	38
	AUTORES	Bouvier, J. y Stragier, P.
	TÍTULO	Nucleotide sequence of the lsp-dapB interval in Escherichia coli
	REVISTA	Nucleic Acids Res. 19 (1), 180 (1991)
35	PUBMED	2011499
	REFERENCIA	39
	AUTORES	Dicker, I.B. y Seetharam, S.
	TÍTULO	Cloning and nucleotide sequence of the firA gene and the firA200(Ts) allele from Escherichia coli
40	REVISTA	J. Bacteriol. 173 (1), 334-344 (1991)

ES 2 700 274 T3

	PUBMED	1987124
	REFERENCIA	40
	AUTORES	Grimm, B., Bull, A. y Breu, V.
5	TÍTULO	Structural genes of glutamate 1-semialdehyde aminotransferase for porphyrin synthesis in a cyanobacterium and <i>Escherichia coli</i>
	JOURNAL	Mol. Gen. Genet. 225 (1), 1-10 (1991)
	PUBMED	1900346
	REFERENCIA	41
	AUTORES	Allen, B.L., Gerlach, G.F. y Clegg, S.
10	TÍTULO	Nucleotide sequence and functions of mrk determinants necessary for expression of type 3 fimbriae in <i>Klebsiella pneumoniae</i>
	JOURNAL	J. Bacteriol. 173 (2), 916-920 (1991)
	PUBMED	1670938
	REFERENCIA	42
15	AUTORES	Chen, H., Lawrence, CB, Bryan, SK y Moses, R.E.
	TÍTULO	Aphidicolin inhibits DNA polymerase II of <i>Escherichia coli</i> , an alpha-like DNA polymerase
	REVISTA	Nucleic Acids Res. 18 (23), 7185-7186 (1990)
	PUBMED	2124684
20	REFERENCIA	43
	AUTORES	Mallonee, D.H., White, W.B. y Hylemon, P.B.
	TÍTULO	Cloning and sequencing of a bile acid-inducible operon from <i>Eubacterium</i> sp. strain VPI 12708
	REVISTA	J. Bacteriol. 172 (12), 7011-7019 (1990)
25	PUBMED	2254270
	REFERENCIA	44
	AUTORES	Young, C., Collins- Emerson, JM, Terzaghi, EA y Scott, D.B.
	TÍTULO	Nucleotide sequence of <i>Rhizobium loti</i> nodI
	REVISTA	Nucleic Acids Res. 18 (22), 6691 (1990)
30	PUBMED	2251131
	REFERENCIA	45
	AUTORES	Chen, H., Sun, Y., Stark, T., Beattie, W. y Moses, R.E.
	TITULO	Nucleotide sequence and deletion analysis of the polB gene of <i>Escherichia coli</i>
	REVISTA	DNA Cell Biol. 9 (9), 631-635 (1990)
35	PUBMED	2261080
	REFERENCIA	46
	AUTORES	Eriani, G., Delarue, M., Poch, O., Gangloff, J. y Moras, D.
	TÍTULO	Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs

ES 2 700 274 T3

	REVISTA	Nature 347 (6289), 203-206 (1990)
	PUBMED	2203971
	REFERENCIA	47
	AUTORES	Showalter, RE y Silverman, M.R.
5	TÍTULO	Nucleotide sequence of a gene, hpt, for hypoxanthine phosphoribosyltransferase from <i>Vibrio harveyi</i>
	REVISTA	Nucleic Acids Res. 18 (15), 4621 (1990)
	PUBMED	2388850
	REFERENCIA	48
10	AUTORES	Martin-Verstraete, I., Debarbouille, M., Klier, A. y Rapoport, G.
	TÍTULO	Levanase operon of <i>Bacillus subtilis</i> includes a fructose-specific phosphotransferase system regulating the expression of the operon
	REVISTA	J. Mol. Biol. 214 (3), 657-671 (1990)
	PUBMED	2117666
15	REFERENCIA	49
	AUTORES	Henrich, B., Monnerjahn, U. y Plapp, R.
	TÍTULO	Peptidase D gene (pepD) of <i>Escherichia coli</i> K-12: nucleotide sequence, transcript mapping, and comparison with other peptidase genes
	REVISTA	J. Bacteriol. 172 (8), 4641-4651 (1990)
20	PUBMED	1695895
	REFERENCIA	50
	AUTORES	Nunn, D., Bergman, S. y Lory, S.
	TÍTULO	Products of three accessory genes, pilB, pilC, and pilD, are required for biogenesis of <i>Pseudomonas aeruginosa</i> pili
25	REVISTA	J. Bacteriol. 172 (6), 2911-2919 (1990)
	PUBMED	1971619
	REFERENCIA	51
	AUTORES	Rosenthal, E.R. y Calvo, J.M.
	TÍTULO	The nucleotide sequence of leuC from <i>Salmonella typhimurium</i>
30	REVISTA	Nucleic Acids Res. 18 (10), 3072 (1990)
	PUBMED	2190189
	REFERENCIA	52
	AUTORES	Kang, P.J. y Craig, E.A.
35	TÍTULO	Identification and characterization of a new <i>Escherichia coli</i> gene that is a dosage-dependent suppressor of a dnaK deletion mutation
	REVISTA	J. Bacteriol. 172 (4), 2055-2064 (1990)
	PUBMED	2180916
	REFERENCIA	53
	AUTORES	Wurgler, S.M. y Richardson, C.C.

ES 2 700 274 T3

	TÍTULO	Structure and regulation of the gene for dGTP triphosphohydrolase from <i>Escherichia coli</i>
	REVISTA	Proc. Natl Acad Sci. USA 87 (7), 2740-2744 (1990)
	PUBMED	2157212
5	REFERENCIA	54
	AUTORES	Schaaff, I., Hohmann , S. y Zimmermann, F.K.
	TÍTULO	Molecular analysis of the structural gene for yeast transaldolase
	REVISTA	Eur. J. Biochem. 188 (3), 597-603 (1990)
	PUBMED	2185015
10	REFERENCIA	55
	AUTORES	Ricca, E. y Calvo, J.M.
	TÍTULO	The nucleotide sequence of <i>leuA</i> from <i>Salmonella typhimurium</i>
	REVISTA	Nucleic Acids Res. 18 (5), 1290 (1990)
	PUBMED	2181403
15	REFERENCIA	56
	AUTORES	Honore, N. y Cole, S.T.
	TÍTULO	Nucleotide sequence of the <i>aroP</i> gene encoding the general aromatic amino acid transport protein of <i>Escherichia coli</i> K-12: homology with yeast transport proteins
	REVISTA	Nucleic Acids Res. 18 (3), 653 (1990)
20	PUBMED	2408019
	REFERENCIA	57
	AUTORES	Angerer, A., Gaisser, S. y Braun, V.
	TÍTULO	Nucleotide sequences of the <i>sfuA</i> , <i>sfuB</i> , and <i>sfuC</i> genes of <i>Serratia marcescens</i> suggest a periplasmic-binding-protein-dependent iron transport mechanism
25	REVISTA	J. Bacteriol. 172 (2), 572-578 (1990)
	PUBMED	2404942
	REFERENCIA	58
	AUTORES	Surin, BP, Watson, JM , Hamilton, WD, Economou, A. y Downie, J.A.
30	TÍTULO	Molecular characterization of the nodulation gene, <i>nodT</i> , from two biovars of <i>Rhizobium leguminosarum</i>
	REVISTA	Mol. Microbiol. 4 (2), 245-252 (1990)
	PUBMED	2338917
	REFERENCIA	59
	AUTORES	Zhou, Z. y Syvanen, M.
35	TÍTULO	Identification and sequence of the <i>drpA</i> gene from <i>Escherichia coli</i>
	REVISTA	J. Bacteriol. 172 (1), 281-286 (1990)
	PUBMED	1688424
	REFERENCIA	60

ES 2 700 274 T3

	AUTORES	Roncero, MI, Jepsen, LP , Stroman, P. y van Heeswijck, R.
	TÍTULO	Characterization of a leuA gene and an ARS element from Mucor circinelloides
	REVISTA	Gene 84 (2), 335-343 (1989)
	PUBMED	2693214
5	REFERENCIA	61
	AUTORES	Ichikawa, S. y Kaji, A.
	TÍTULO	Molecular cloning and expression of ribosome releasing factor
	REVISTA	J. Biol. Chem. 264 (33), 20054-20059 (1989)
	PUBMED	2684966
10	REFERENCIA	62
	AUTORES	Minami-Ishii, N., Taketani, S., Osumi, T. y Hashimoto, T.
	TÍTULO	Molecular cloning and sequence analysis of the cDNA for rat mitochondrial enoyl-CoA hydratase. Structural and evolutionary relationships linked to the bifunctional enzyme of the peroxisomal beta-oxidation system
15	REVISTA	Eur. J. Biochem. 185 (1), 73-78 (1989)
	PUBMED	2806264
	REFERENCIA	63
	AUTORES	Matsubara, Y., Indo, Y., Naito, E., Ozasa, H., Glassberg, R., Vockley, J., Ikeda, Y., Kraus, J. Tanaka, K.
20	TÍTULO	Molecular cloning and nucleotide sequence of cDNAs encoding the precursors of rat long chain acyl-coenzyme A, short chain acyl-coenzyme A, and isovaleryl-coenzyme A dehydrogenases. Sequence homology of four enzymes of the acyl-CoA dehydrogenase family
	REVISTA	J. Biol. Chem. 264 (27), 16321-16331 (1989)
25	PUBMED	2777793
	REFERENCIA	64
	AUTORES	Roa, B.B., Connolly, D.M. y Winkler, M.E.
	TÍTULO	Overlap between pdxA and ksgA in the complex pdxA-ksgA-apaG-apaH operon of Escherichia coli K-12
30	REVISTA	J. Bacteriol. 171 (9), 4767-4777 (1989)
	PUBMED	2670894
	REFERENCIA	65
	AUTORES	Lindquist, S., Galleni, M., Lindberg, F. y Normark, s.
	TÍTULO	Signalling proteins in enterobacterial AmpC beta-lactamase regulation
35	REVISTA	Mol. Microbiol. 3 (8), 1091-1102 (1989)
	REFERENCIA	2691840
	REFERENCIA	66
	AUTORES	Xie, Q.W., Tabor, C.W. y Tabor, H.
40	TÍTULO	Spermidine biosynthesis in Escherichia coli: promoter and termination regions of the speED operon

ES 2 700 274 T3

	REVISTA	J. Bacteriol. 171 (8), 4457-4465 (1989)
	PUBMED	2666401
	REFERENCIA	67
	AUTORES	Sato, S., Nakada, Y. y Shiratsuchi, A.
5	TÍTULO	IS421, a new insertion sequence in Escherichia coli
	REVISTA	FEBS Lett. 249 (1), 21-26 (1989)
	PUBMED	2542093
	REFERENCIA	68
	AUTORES	Liu, J.D. y Parkinson, J.S.
10	TÍTULO	Genetics and sequence analysis of the pcnB locus, an Escherichia coli gene involved in plasmid copy number control
	REVISTA	J. Bacteriol. 171 (3), 1254-1261 (1989)
	PUBMED	2537812
	REFERENCIA	69
15	AUTORES	Henrich, B., Schroeder, U., Frank, RW y Plapp, R.
	TÍTULO	Accurate mapping of the Escherichia coli pepD gene by sequence analysis of its 5' flanking region
	REVISTA	Mol. Gen. Genet. 215 (3), 369-373 (1989)
	PUBMED	2651887
20	REFERENCIA	70
	AUTORES	Lipinska, B., Sharma, S. y Georgopoulos, C.
	TÍTULO	Sequence analysis and regulation of the htrA gene of Escherichia coli: a sigma 32-independent mechanism of heat-inducible transcription
	REVISTA	Nucleic Acids Res. 16 (21), 10053-10067 (1988)
25	PUBMED	3057437
	REFERENCIA	71
	AUTORES	Sung, Y.C. y Fuchs, J.A.
	TITULO	Characterization of the cyn operon in Escherichia coli K12
	REVISTA	J. Biol. Chem. 263 (29), 14769-14775 (1988)
30	PUBMED	3049588
	REFERENCIA	72
	AUTORES	Lozoya, E., Hoffmann, H., Douglas, C., Schulz, W., Scheel, D. y Hahlbrock, K.
	TÍTULO	Primary structures and catalytic properties of isoenzymes encoded by the two 4-coumarate: CoA ligase genes in parsley
35	JOURNAL	Eur. J. Biochem. 176 (3), 661-667 (1988)
	PUBMED	3169018
	REFERENCIA	73
	AUTORES	Andrews, S.C. y Guest, J.R.

ES 2 700 274 T3

	TÍTULO	Nucleotide sequence of the gene encoding the GMP reductase of Escherichia coli K12
	REVISTA	Biochem. J. 255 (1), 35-43 (1988)
	PUBMED	2904262
5	REFERENCIA	74
	AUTORES	Jaiswal, AK, McBride, OW, Adesnik, M. y Nebert, D.W.
	TÍTULO	Human dioxin-inducible cytosolic NAD(P)H:menadione oxidoreductase. cDNA sequence and localization of gene to chromosome 16
	REVISTA	J. Biol. Chem. 263 (27), 13572-13578 (1988)
10	PUBMED	2843525
	REFERENCIA	75
	AUTORES	Karpel, R., Olami, Y., Taglicht, D., Schuldiner, S. y Padan, E.
	TÍTULO	Sequencing of the gene ant which affects the Na ⁺ /H ⁺ antiporter activity in Escherichia coli
15	REVISTA	J. Biol. Chem. 263 (21), 10408-10414 (1988)
	PUBMED	2839489
	REFERENCIA	76
	AUTORES	Mellano, M.A. y Cooksey, D.A.
20	TÍTULO	Nucleotide sequence and organization of copper resistance genes from Pseudomonas syringae pv. tomato
	REVISTA	J. Bacteriol. 170 (6), 2879-2883 (1988)
	PUBMED	3372485
	REFERENCIA	77
	AUTORES	Coleman, J. y Raetz, C.R.
25	TÍTULO	First committed step of lipid A biosynthesis in Escherichia coli: sequence of the lpxA gene
	REVISTA	J. Bacteriol. 170 (3), 1268-1274 (1988)
	PUBMED	3277952
	REFERENCIA	78
30	AUTORES	Gebhard, W., Schreitmuller, T., Hochstrasser, K. y Wachter, E.
	TÍTULO	Complementary DNA and derived amino acid sequence of the precursor of one of the three protein components of the inter-alpha-trypsin inhibitor complex
	REVISTA	FEBS Lett. 229 (1), 63-67 (1988)
	PUBMED	2450046
35	REFERENCIA	79
	AUTORES	Tomasiewicz, H.G. y McHenry, C.S.
	TÍTULO	Sequence analysis of the Escherichia coli dnaE gene
	REVISTA	J. Bacteriol. 169 (12), 5735-5744 (1987)
	PUBMED	3316192

ES 2 700 274 T3

	REFERENCIA	80
	AUTORES	Crowell, D.N., Reznikoff, W.S. y Raetz, C.R.
	TÍTULO	Nucleotide sequence of the Escherichia coli gene for lipid A disaccharide synthase
5	REVISTA	J. Bacteriol. 169 (12), 5727-5734 (1987)
	PUBMED	2824445
	REFERENCIA	81
	AUTORES	Tabor, C.W. y Tabor, H.
10	TÍTULO	The speEspeD operon of Escherichia coli. Formation and processing of a proenzyme form of S-adenosylmethionine decarboxylase
	REVISTA	J. Biol. Chem. 262 (33), 16037-16040 (1987)
	PUBMED	3316212
	REFERENCIA	82
	AUTORES	Nonet, M.L., Marvel, C.C. y Tolan, D.R.
15	TITULO	The hisT-purF region of the Escherichia coli K-12 chromosome. Identification of additional genes of the hisT and purF operons
	REVISTA	J. Biol. Chem. 262 (25), 12209-12217 (1987)
	PUBMED	3040734
	REFERENCIA	83
20	AUTORES	Coulton, JW, Mason, P. y Allatt, D.D.
	TITULO	fhuC and fhuD genes for iron (III)-ferrichrome transport into Escherichia coli K-12
	REVISTA	J. Bacteriol. 169 (8), 3844-3849 (1987)
	PUBMED	3301821
	REFERENCIA	84
25	AUTORES	Horiuchi, T., Nagasawa, T., Takano, K. y Sekiguchi, M.
	TÍTULO	A newly discovered tRNA(IAsp) gene (aspV) of Escherichia coli K12
	REVISTA	Mol. Gen. Genet. 206 (2), 356-357 (1987)
	PUBMED	3295485
	REFERENCIA	85
30	AUTORES	Ben-Bassat, A. Bauer, K., Chang, SY, Myambo, K., Boosman, A. y Chang, S.
	TÍTULO	Processing of the initiation methionine from proteins: properties of the Escherichia coli methionine aminopeptidase and its gene structure
	REVISTA	J. Bacteriol. 169 (2), 751-757 (1987)
	PUBMED	3027045
35	REFERENCIA	86
	AUTORES	Gronger, P., Manian, SS, Reilander, H., O'Connell, M., Priefer, UB y Puhler, A.
	TÍTULO	Organization and partial sequence of a DNA region of the Rhizobium leguminosarum symbiotic plasmid pRL6JI containing the genes fixABC, nifA, nifB and a novel open reading frame

ES 2 700 274 T3

	REVISTA	Nucleic Acids Res. 15 (1), 31-49 (1987)
	PUBMED	3029674
	REFERENCIA	87
	AUTORES	Richardson, KK, Richardson, FC , Crosby, RM, Swenberg, JA y Skopek, T.R.
5	TITULO	DNA base changes and alkylation following in vivo exposure of Escherichia coli to N-methyl-N-nitrosourea or N-ethyl-N-nitrosourea
	REVISTA	Proc. Natl Acad Sci. U.S.A. 84 (2), 344-348 (1987)
	PUBMED	3540961
	REFERENCIA	88
10	AUTORES	Chye, M.L. y Pittard, J.
	TÍTULO	Transcription control of the aroP gene in Escherichia coli K-12 : analysis of operator mutants
	REVISTA	J . Bacteriol. 169 (1), 386-393 (1987)
	PUBMED	3025182
15	REFERENCIA	89
	AUTORES	Blanchin-Roland, S., Blanquet, S., Schmitter, JM y Fayat, G.
	TÍTULO	The gene for Escherichia coli diadenosine tetraphosphatase is located immediately clockwise to folA and forms an operon with ksgA
	JOURNAL	Mol. Gen. Genet. 205 (3), 515-522 (1986)
20	PUBMED	3031429
	REFERENCIA	90
	AUTORES	Takano, K., Nakabeppu, Y., Maki, H., Horiuchi, T. y Sekiguchi, M.
	TÍTULO	Structure and function of dnaQ and mutD mutators of Escherichia coli
	REVISTA	Mol. Gen. Genet. 205 (1), 9-13 (1986)
25	PUBMED	3540531
	REFERENCIA	91
	AUTORES	Mackie, G.A.
	TÍTULO	Structure of the DNA distal to the gene for ribosomal protein S20 in Escherichia coli K12: presence of a strong terminator and an IS1 element
30	REVISTA	Nucleic Acids Res. 14 (17), 6965-6981 (1986)
	PUBMED	2429258
	REFERENCIA	92
	AUTORES	Koster, W. y Braun, V.
	TÍTULO	Iron hydroxamate transport of Escherichia coli: nucleotide sequence of the fluB gene and identification of the protein
35	REVISTA	Mol. Gen. Genet. 204 (3), 435-442 (1986)
	PUBMED	3020380
	REFERENCIA	93
	AUTORES	Breton, R., Sanfacon, H., Papayannopoulos, I., Biemann, K. y Lapointe, J.

ES 2 700 274 T3

	TÍTULO	Glutamyl-tRNA synthetase of Escherichia coli. Isolation and primary structure of the gltX gene and homology with other aminoacyl-tRNA synthetases
	REVISTA	J. Biol. Chem. 261 (23), 10610-10617 (1986)
	PUBMED	3015933
5	REFERENCIA	94
	AUTORES	Birnbaum, M.J., Haspel, H.C. y Rosen, O.M.
	TÍTULO	Cloning and characterization of a cDNA encoding the rat brain glucose-transporter protein
	REVISTA	Proc. Natl Acad Sol. U.S.A. 83 (16), 5784-5788 (1986)
10	PUBMED	3016720
	REFERENCIA	95
	AUTORES	Cox, E.C. y Horner, D.L.
	TÍTULO	DNA sequence and coding properties of mutD(dnaQ) a dominant Escherichia coli mutator gene
15	REVISTA J.	J. Mol. Biol. 190 (1), 113-117 (1986)
	PUBMED	3023634
	REFERENCIA	96
	AUTORES	Ohki, M., Tamura, F., Nishimura, S. y Uchida, H.
20	TÍTULO	Nucleotide sequence of the Escherichia coli dnaJ gene and purification of the gene product
	REVISTA	J. Biol. Chem. 261 (4), 1778-1781 (1986)
	PUBMED	3003084
	REFERENCIA	97
	AUTORES	Coulton, JW, Mason, P ., Cameron, DR, Carmel, G., Jean, R. y Rode, HN
25	TÍTULO	Protein fusions of beta-galactosidase to the ferrichrome-iron receptor of Escherichia coli K-12
	REVISTA	J. Bacteriol. 165 (1), 181-192 (1986)
	PUBMED	3079747
	REFERENCIA	98
30	AUTORES	Lee, N., Gielow, W., Martin, R., Hamilton, E. y Fowler, A.
	TÍTULO	The organization of the araBAD operon of Escherichia coli
	REVISTA	Gene 47 (2-3), 231-244 (1986)
	PUBMED	3549454
	REFERENCIA	99
35	AUTORES	Sekiguchi, T., Ortega-Cesena, J., Nosoh, Y., Ohashi, S., Tsuda, K. y Kanaya, S.
	TITULO	DNA and amino-acid sequences of 3-isopropylmalate dehydrogenase of Bacillus coagulans. Comparison with the enzymes of Saccharomyces cerevisiae and Thermus thermophilus
	REVISTA	Biochim. Biophys. Acta 867, 36-44 (1986)
40	REFERENCIA	100

ES 2 700 274 T3

	AUTORES	Chong, P., Hui, I., Loo, T. y Gillam, S.
	TÍTULO	Structural analysis of a new GC-specific insertion element IS186
	REVISTA	FEBS Lett. 192 (1), 47-52 (1985)
	PUBMED	2996940
5	REFERENCIA	101
	AUTORES	Icho, T., Sparrow, CP y Raetz, C.R.
	TÍTULO	Molecular cloning and sequencing of the gene for CDP-diglyceride synthetase of Escherichia coli
	REVISTA	J. Biol. Chem. 260 (22), 12078-12083 (1985)
10	PUBMED	2995358
	REFERENCIA	102
	AUTORES	Nomura, T., Aiba, H. e Ishihama, A.
	TÍTULO	Transcriptional organization of the convergent overlapping dnaQ-rnh genes of Escherichia coli
15	REVISTA	J. Biol. Chem. 260 (11), 7122-7125 (1985)
	PUBMED	2987244
	REFERENCIA	103
	AUTORES	Kamio, Y., Lin, CK, Regue, M. y Wu, H.C.
20	TÍTULO	Characterization of the ileS-lsp operon in Escherichia coli. Identification of an open reading frame upstream of the ileS gene and potential promoter(s) for the ileS-lsp operon
	REVISTA	J. Biol. Chem. 260 (9), 5616-5620 (1985)
	PUBMED	2985604
	REFERENCIA	104
25	AUTORES	Cowing, DW, Bardwell, JC , Craig, EA, Woolford, C., Hendrix, RW y Gross, CA
	TÍTULO	Consensus sequence for Escherichia coli heat shock gene promoters
	REVISTA	Proc. Natl Acad Sol. USA 82 (9), 2679-2683 (1985)
	PUBMED	3887408
	REFERENCIA	105
30	AUTORES	Broome-Smith, JK, Edelman, A., Yousif, S. y Spratt, B.G.
	TÍTULO	The nucleotide sequences of the ponA and ponB genes encoding penicillin-binding protein 1A and 1B of Escherichia coli K12
	REVISTA	Eur. J. Biochem. 147 (2), 437-446 (1985)
	PUBMED	3882429
35	REFERENCIA	106
	AUTORES	Becerril, B., Valle, F., Merino, E., Riba, L. y Bolívar, F.
	TITULO	Repetitive extragenic palindromic (REP) sequences in the Escherichia coli gdhA gene
	REVISTA	Gene 37 (1-3), 53-62 (1985)

ES 2 700 274 T3

	PUBMED	3902576
	REFERENCIA	107
	AUTORES	Friedberg, D., Rosenthal, ER, Jones, JW y Calvo, J.M.
	TÍTULO	Characterization of the 3' end of the leucine operon of Salmonella typhimurium
5	REVISTA	Mol. Gen. Genet. 199 (3), 486-494 (1985)
	PUBMED	2993799
	REFERENCIA	108
	AUTORES	Bouvier, J., Richaud, C., Richaud, F., Patte, JC y Stragier, p.
	TÍTULO	Nucleotide sequence and expression of the Escherichia coli dapB gene
10	REVISTA	J. Biol. Chem. 259 (23), 14829-14834 (1984)
	PUBMED	6094578
	REFERENCIA	109
	AUTORES	Richaud, C., Richaud, F., Martin, C., Haziza, C. y Patte, J.C.
15	TÍTULO	Regulation of expression and nucleotide sequence of the Escherichia coli dapD gene
	REVISTA	J. Biol. Chem. 259 (23), 14824-14828 (1984)
	PUBMED	6094577
	REFERENCIA	110
	AUTORES	Nuesch, J. y Schumperli, D.
20	TÍTULO	Structural and functional organization of the gpt gene region of Escherichia coli
	REVISTA	Gene 32 (1-2), 243-249 (1984)
	PUBMED	6397401
	REFERENCIA	111
	AUTORES	Jagadeeswaran, P., Ashman, CR, Roberts, S. y Langenberg, J.
25	TÍTULO	Nucleotide sequence and analysis of deletion mutants of the Escherichia coli gpt gene in plasmid pSV2 gpt
	REVISTA	Gene 31 (1-3), 309-313 (1984)
	PUBMED	6396164
	REFERENCIA	112
30	AUTORES	Deutch, AH, Rushlow, KE y Smith, C.J.
	TÍTULO	Analysis of the Escherichia coli proBA locus by DNA and protein sequencing
	REVISTA	Nucleic Acids Res. 12 (15), 6337-6355 (1984)
	PUBMED	6089111
	REFERENCIA	113
35	AUTORES	Bouvier, J., Patte, JC y Stragier, p.
	TÍTULO	Multiple regulatory signals in the control region of the Escherichia coli carAB operon
	REVISTA	Proc. Natl Acad Sci. USA 81 (13), 4139-4143 (1984)

ES 2 700 274 T3

	PUBMED	6377309
	REFERENCIA	114
	AUTORES	Innis, MA, Tokunaga, M., Williams, ME, Loranger, JM, Chang, SY, Chang, S. y Wu, H.C.
5	TÍTULO	Nucleotide sequence of the Escherichia coli prolipoprotein signal peptidase (lsp) gene
	REVISTA	Proc. Natl. Acad. Sci. U.S.A. 81 (12), 3708-3712 (1984)
	PUBMED	6374664
	REFERENCIA	115
10	AUTORES	Bardwell, J.C. y Craig, E.A.
	TÍTULO	Major heat shock gene of Drosophila and the Escherichia coli heat-inducible dnaK gene are homologous
	REVISTA	Proc. Natl. Acad. Sci. U.S.A. 81 (3), 848-852 (1984)
	PUBMED	6322174
15	REFERENCIA	116
	AUTORES	Pratt, D. y Subramani, S.
	TÍTULO	Nucleotide sequence of the Escherichia coli xanthine-guanine phosphoribosyl transferase gene
	REVISTA	Nucleic Acids Res. 11 (24), 8817-8823 (1983)
20	PUBMED	6324103
	REFERENCIA	117
	AUTORES	Richardson, KK, Fostel, J. y Skopek, T.R.
	TÍTULO	Nucleotide sequence of the xanthine guanine phosphoribosyl transferase gene of E. coli
25	REVISTA	Nucleic Acids Res. 11 (24), 8809-8816 (1983)
	PUBMED	6324102
	REFERENCIA	118
	AUTORES	Parsot, C., Cossart, P., Saint-Girons, I. y Cohen, G.N.
30	TÍTULO	Nucleotide sequence of thrC and of the transcription termination región of the threonine operon in Escherichia coli K12
	JOURNAL	Nucleic Acids Res. 11 (21), 7331-7345 (1983)
	PUBMED	6316258
	REFERENCIA	119
	AUTORES	Stephens, PE, Lewis, HM , Darlison, MG y Guest, J.R.
35	TÍTULO	Nucleotide sequence of the lipoamide dehydrogenase gene of Escherichia coli K12
	REVISTA	Eur. J. Biochem. 135 (3), 519-527 (1983)
	PUBMED	6352260
	REFERENCIA	120
40	AUTORES	Stephens, PE, Darlison, MG , Lewis, HM y Guest, J.R.

ES 2 700 274 T3

	TÍTULO	The pyruvate dehydrogenase complex of Escherichia coli K12. Nucleotide sequence encoding the dihydrolipoamide acetyltransferase component
	REVISTA	Eur. J. Biochem. 133 (3), 481-489 (1983)
	PUBMED	6345153
5	REFERENCIA	121
	AUTORES	Stephens, PE, Darlison, MG , Lewis, HM y Guest, J.R.
	TÍTULO	The pyruvate dehydrogenase complex of Escherichia coli K12. Nucleotide sequence encoding the pyruvate dehydrogenase component
	REVISTA	Eur. J. Biochem. 133 (1), 155-162 (1983)
10	PUBMED	6343085
	REFERENCIA	122
	AUTORES	Kanaya, S. y Crouch, R.J.
	TÍTULO	Low levels of RNase H activity in Escherichia coli FB2 rnh result from a single-base change in the structural gene of RNase H
15	REVISTA	J. Bacteriol. 154 (2), 1021-1026 (1983)
	PUBMED	6302075
	REFERENCIA	123
	AUTORES	Overbeeke, N., Bergmans, H., van Mansfeld, F. y Lugtenberg, B.
20	TÍTULO	Complete nucleotide sequence of phoE, the structural gene for the phosphate limitation inducible outer membrane pore protein of Escherichia coli K12
	REVISTA	J. Mol. Biol. 163 (4), 513-532 (1983)
	PUBMED	6341601
	REFERENCIA	124
	AUTORES	Gilson, E., Nikaido, H. y Hofnung, M.
25	TÍTULO	Sequence of the malK gene in E.coli K12
	REVISTA	Nucleic Acids Res. 10 (22), 7449-7458 (1982)
	PUBMED	6296778
	REFERENCIA	125
	AUTORES	Stoner, C.M. y Schleif, R.
30	TÍTULO	Is the amino acid but not the nucleotide sequence of the Escherichia coli araC gene conserved?
	REVISTA	J. Mol. Biol. 154 (4), 649-652 (1982)
	PUBMED	6283093
	REFERENCIA	126
35	AUTORES	An, G., Bendiak, DS, Mamelak, LA y Friesen, J.D.
	TÍTULO	Organization and nucleotide sequence of a new ribosomal operon in Escherichia coli containing the genes for ribosomal protein S2 and elongation factor Ts
	REVISTA	Nucleic Acids Res. 9 (16), 4163-4172 (1981)
	PUBMED	6272196

ES 2 700 274 T3

	REFERENCIA	127
	AUTORES	Mackie, G.A.
	TÍTULO	Nucleotide sequence of the gene for ribosomal protein S20 and its flanking regions
5	REVISTA	J. Biol. Chem. 256 (15), 8177-8182 (1981)
	PUBMED	6267039
	REFERENCIA	128
	AUTORES	Little, J.W., Mount, D.W. y Yanisch-Perron, C.R.
	TÍTULO	Purified <i>lexA</i> protein is a repressor of the <i>recA</i> and <i>lexA</i> genes
10	REVISTA	Proc. Natl. Acad. Sci. U.S.A. 78 (7), 4199-4203 (1981)
	PUBMED	7027255
	REFERENCIA	129
	AUTORES	Mulligan, R.C. y Berg, p.
	TÍTULO	Factors governing the expression of a bacterial gene in mammalian cells
15	REVISTA	Mol. Cell. Biol. 1 (5), 449-459 (1981)
	PUBMED	6100966
	REFERENCIA	130
	AUTORES	Lee, N.L., Gielow, W.O. y Wallace, R.G.
20	TÍTULO	Mechanism of <i>araC</i> autoregulation and the domains of two overlapping promoters, <i>P_c</i> and <i>P_{BAD}</i> , in the L-arabinose regulatory region of <i>Escherichia coli</i>
	REVISTA	Proc. Natl. Acad. Sci. USA 78 (2), 752-756 (1981)
	PUBMED	6262769
	REFERENCIA	131
	AUTORES	Cossart, P., Katinka, M. y Yaniv, M.
25	TÍTULO	Nucleotide sequence of the <i>thrB</i> gene of <i>E. coli</i> , and its two adjacent regions; the <i>thrAB</i> and <i>thrBC</i> junctions
	REVISTA	Nucleic Acids Res. 9 (2), 339-347 (1981)
	PUBMED	6259626
	REFERENCIA	132
30	AUTORES	Miyada, CG, Horwitz, AH, Cass, LG, Timko, J. y Wilcox, G.
	TÍTULO	DNA sequence of the <i>araC</i> regulatory gene from <i>Escherichia coli</i> B/r
	REVISTA	Nucleic Acids Res. 8 (22), 5267-5274 (1980)
	PUBMED	7008027
	REFERENCIA	133
35	AUTORES	Katinka, M., Cossart, P., Sibilli, L., Saint-Girons, I., Chalvignac, MA, Le Bras, G., Cohen, GN y Yaniv, M.
	TÍTULO	Nucleotide sequence of the <i>thrA</i> gene of <i>Escherichia coli</i>
	REVISTA	Proc. Natl. Acad. Sci. USA 77 (10), 5730-5733 (1980)

ES 2 700 274 T3

	PUBMED	7003595
	REFERENCIA	134
	AUTORES	Ogden, S., Haggerty, D., Stoner, CM, Kolodrubetz, D. y Schleif, R.
5	TÍTULO	The Escherichia coli L-arabinose operon: binding sites of the regulatory proteins and a mechanism of positive and negative regulation
	REVISTA	Proc. Natl. Acad. Sci. U.S.A. 77 (6), 3346-3350 (1980)
	PUBMED	6251457
	REFERENCIA	135
	AUTORES	Smith, D.R. y Calvo, J.M.
10	TÍTULO	Nucleotide sequence of the E coli gene coding for dihydrofolate reductase
	REVISTA	Nucleic Acids Res. 8 (10), 2255-2274 (1980)
	PUBMED	6159575
	REFERENCIA	136
	AUTORES	Johnsrud, L.
15	TÍTULO	DNA sequence of the transposable element IS1
	REVISTA	Mol. Gen. Genet. 169 (2), 213-218 (1979)
	PUBMED	375010
	REFERENCIA	137
	AUTORES	Smith, B.R. y Schleif, R.
20	TÍTULO	Nucleotide sequence of the L-arabinose regulatory region of Escherichia coli K12
	REVISTA	J. Biol. Chem. 253 (19), 6931-6933 (1978)
	PUBMED	357433
	REFERENCIA	138
	AUTORES	Greenfield, L., Boone, T. y Wilcox, G.
25	TITULO	DNA sequence of the araBAD promoter in Escherichia coli B/r
	REVISTA	Proc. Natl. Acad. Sci. U.S.A. 75 (10), 4724-4728 (1978)
	PUBMED	368797
	REFERENCIA	139
	AUTORES	Young, R.A. y Steitz, J.A.
30	TÍTULO	Complementary sequences 1700 nucleotides apart form a ribonuclease III cleavage site in Escherichia coli ribosomal precursor RNA
	REVISTA	Proc. Natl. Acad. Sci. U.S.A. 75 (8), 3593-3597 (1978)
	PUBMED	358189
	REFERENCIA	140
35	AUTORES	Ohtsubo, H. y Ohtsubo, E.
	TÍTULO	Nucleotide sequence of an insertion element, IS1
	REVISTA	Proc. Natl. Acad. Sci. USA 75 (2), 615-619 (1978)

ES 2 700 274 T3

	PUBMED	273224
	REFERENCIA	141
	AUTORES	Musso, R., Di Lauro, R., Rosenberg, M. y de Crombrugghe, B.
5	TÍTULO	Nucleotide sequence of the operator-promoter region of the galactose operon of Escherichia coli
	REVISTA	Proc. Natl. Acad. Sci. USA 74 (1), 106-110 (1977)
	PUBMED	319453
	REFERENCIA	142 (bases 1 a 4646332)
	CONSRMT	NCBI Genome Project
10	TÍTULO	Envío directo
	REVISTA	Enviado (10-NOV-2005) National Center for Biotechnology Information, NIH, Bethesda, MD 20894, EE.UU.
	REFERENCIA	143 (bases 1 a 4646332)
	AUTORES	Mori, H., Horiuchi, T. y Hirai, A.
15	TÍTULO	Presentación directa
	REVISTA	Enviado (22-AUG-2005) Hirotada Mori, Graduate School of Biological Sciences, Nara Institute of Science and Technology; 8916-5 Takayama, Ikoma, Nara 630-0101, Japón (Correo electrónico: hmori@gtc.naist.jp, Tel: 81-743-72-5660, Fax: 81-743-72- 5669)
20	COMENTARIO	REFSEQ PROVISIONAL: Este registro aún no ha sido sometido a revisión final del NCBI . La secuencia de referencia se derivó de partir de AP009048. COMPLETO: longitud completa.
	CARACTERÍSTICAS	Ubicación/Calificadores
	fuente	complemento (<1 ..> 5861)
25		/organism = "cepa de Escherichia coli K-12 sustr. W3110"
		/mol_type = "ADN genómico"
		/strain = "K-12 "
		/sub_strain = "W3110"
		/db_xref = "Taxón: 316407"
30	gen	complemento (<1..6)
		/gene = "dcuD"
	CDS	complemento (<1. .6)
		/gene = "dcuD"
		/note = "ECK3216: JW3196: b3227"
35		/codon_start = 1
		/transl_table = 11
		/producto = "transportador predicho"
		/protein_id = "AP 003769.1"
		/db_xref = "GI: 89109989"

ES 2 700 274 T3

```

/translacion="MFGIIISVIVLITMGYLILKNYKPQVVLAAAGIFLMMCGVWLGF
GGVLDPTKSSGYLIVDIYNEILRMLSNRIAGLGLSIMAVGGYARYMERIGASRAMVSL
LSRPLKLIRSPYIILSATYVIGQIMAQFITSASGLGMLLMVTLFPTLVSLGVSRLSAV
AVIATTMSIEWGILETNSIFAAQVAGMKIATYFFHYQLPVASCVIISVAISHFFVQRA
FDKKDKNINHEQAEQKALDNVPPLYYAILPVMPLIILMLGSLFLAHVGLMQSELHLVVV
MLLSLTVTMEVVEFFRKHNLRMDDVQAFFDGMGTQFANVVTLVVAGEIFAKGLTTIG
TVDAVIRGAEHSGLGGIGVMIIMALVIAICAIVMGSGNAPFMSFASLIPNIAAGLHVP
AVVMIMP MHFATTLARAVSPITAVVVVTSGIAGVSPFAVVKRTAIPMAVGFVVNMIAT
ITLFY" (SEQ ID NO: 35)

```

	cebador	330..348	/label="ck nanR3 control primer"
	gene	386..1177	/gene="nanR"
5			/gene="nanR"
	CDS	386..1177	/note="ECK3215:JW3195:b3226"
			/codon_start=1
10			/transl_table=11
			/product="regulador doble transcripcional de enlace de ADN"
			/protein_id="AP_003768.1"
			/db_xref="GI:89109988"

```

/translacion="MGLMNAFDSQTEDSSPAIGRNLRSRPLARKKLSEMVEEELEQMI
RRREFGEGEQLPSERELMAFFNVGRPSVREALAALKRKGLVQINNGERARVSRPSADT
IIGELSGMAKDFLSHPGGIAHFEQLRLFFESSLVRYAAEHATDEQIDLLAKALEINSQ
SLDNNAAFIRSDVDFHRVLAIEIPGNPIFMAIHVALLDWLIAARPTVTDQALHEHNNVS
YQQHIAIVDAIRRHDPDEADRALQSHLNSVSATWHAFGQTTNKKK" (SEQ
ID NO: 36)

```

15	cebador	1005..1025	/label = "cebador de control nanR ck2"
	cebador	1126..1146	/label = "cebador de control nanAFck"
	promotor	1178..1278	
20			/label = "región promotora del operón nan"
	Sitio	1187..1191	

ES 2 700 274 T3

```

/site_type = "sitio de enlace"
/label = "enlace CAP"
Sitio 1198 .. 1202
/site_type = "sitio de enlace"
5 /label = "enlace de CAP"
promotor 1241..1246
/label = -10
enlace de cebador 1252..1301
/label = "cebador H1-dnanA lambda rojo"
10 /note = "para eliminaciones de dnanA:: o dnanATE::marca"
mRNA 1255
/label = + 1
mRNA 1267
/label = + 13
15 mRNA 1279
/label = + 25
gen 1299..2192
/label = "nanA"
CDS 1299..2192
20 /gen = "nanA"
/label = "ECK3214: JW3194: b3225"
/codon_start = 1
/transl_table = 11
/product = "N-acetilneuraminato liasa"
25 /protein_id = "AP_003767.1"
/d_b_xref = "GI: 89109987"
/translation="MATNLRGVMAALLTPFDQQQALDKASLRRLVQFNIQQGIDGLYV
GGSTGEAFVQSLSEREQVLEIVAEEAKGKIKLIAHVGCVSTAESQQLAASAKRYGFDA
VSAVTPFYYPFSFEEHCDHYRAIIDSADGLPMVVYNIPALSGVKLTLDQINTLVTLPG
VGALKQTSGDLYQMEQIRREHPDLVLYNGYDEIFASGLLAGADGGIGSTYNIMGWRYQ
GIVKALKEGDIQTAQKLQTECNKVIDLLIKTGVFRGLKTVLHYMDVVSVP LCRKPFPG
VDEKYLPELKALAAQQMLMQERG" (SEQ ID NO: 37)
Región 1302..4424
/label ="DELECIÓN nanATE"
30 enlace de cebador complemento (2175..2224)

```

ES 2 700 274 T3

```

                                /label = "H2-dnanA lambda rojo cebador "
gen                               2301..3791
                                /gene = "nanT"
CDS                               2301..3791
5                                /gene = "nanT"
                                /note = "ECK3213: JW3193: b3224"
                                /codon_start = 1
                                /transl_table = 11
                                /product = "transportador de ácido siálico"
10                               /protein_id = "AP 003766.1"
                                /db_xref = "GI: 89109986"

/translation="MSTTTQNI PWYRHLNRAQWR AFSAAWLGYLLDGFDFVLIALVLT
EVQGEFGLTTVQAASLISAAFISRWFGLMLGAMGDYGRRLAMVTSIVLFSAGTLAC
GFAPGYITMFIARLVIGMG MAGEY GSSATYVIESWPKHLRNKASGFLISGFSVGAVVA
AQVYSLVVPVWGWRALFFIGILPIIFALWLRKNIPEAEDWKEKHAGKAPVRTMVDILY
RGEHRIANIVMTLAAATALWFCFAGNLQNAAI VAVLGLLCAAIFISFMVQSAGKRWPT
GVMLMVVVLFAFLYSWPIQALLPTYLKTDLAYNPHTVANV LFFSGFGAAVGC CVGGFL
GDWLGTRKAYVCSLLASQLLIIPVFAIGGANVWVLG LLLFFQQMLGQGIAGILPKLIG
GYFDTDQRAAGLGFTYNV GALGGALAPIIGALIAQRLDLGTALASLSFSLTFV VILLI
                                GLDMPSRVQRWLRPEALRTHDAIDGKPFSGAVPFGSAKNDLVKTKS" (SEQ
ID NO: 38)
cebador                          complemento (2329..2350)
                                /label = "cebador de control nanARck"
15  enlace de cebador           3792..3841
                                /label = "cebador H1-dnanE lambda rojo"
gen                               3839..4528
                                /gene = "nanE"
CDS                               3839..4528
20                               /gene = "nanE"
                                /note = "ECK3212: JW3192: b3223"
                                /codon_start = 1
                                /transl_table = 11
                                /product = "N-acetilmannosamina-6-P epimerasa predicha"
25                               /protein_id = "AP 003765.1"
                                /db_xref = "GI: 89109985"

```

ES 2 700 274 T3

/translation="MSLLAQLDQKIAANGGLIVSCQVPVDSPLDKPEIIVAAMALAAEQ
 AGAVAIRIEGVANLQATRAVVSVP IIGIVKRDLEDSVVRITAYIEDVDALAQAGADI I
 AIDGTD RPRPVPVETLLARIHHHGLLAMTDCSTPEDGLACQKLGAEIIGTTLSGYTTP
 ETPEEPDLALVKTLSDAGCRVIAEGRYNTPAQAADAMRHGAWAVTVGSAITRLEHICQ
 WYNTAMKKAVL " (SEQ ID NO: 39)

5 enlace de cebador complemento (4425..4474)
 /note = " para la eliminación de dnanato:: marca"
 /label = "cebador H2-dnanE lambda rojo"
 RBS 4425..4448
 /label = "péptido incomprensible C-terminal fusionado al péptido de
 marca KD13"
 RBS 4449..4451
 10 /label = "péptido incomprensible NEW STOP péptido después de la
 resolución del casete"
 enlace de cebador 4486..4530
 /label = "cebador nanK-H1 lambda rojo"
 RBS 4515..4520
 /label = "nanK RBS"
 15 gen 4525..5400
 /gene = "nanK"
 CDS 4525 .. 5400
 /gene = "nanK"
 /note = "ECK3211: JW5538: b3222"
 20 /codon_start = 1
 /transl_table = 11
 /product = "N-acetilmannosamina quinasa predicha"
 /protein_id = "AP 003764.1"
 /db_xref = "GI: 89109984"

/translation="MTTLAIDIGGTKLAAALIGADGQIRDRRELPTPASQTPEALRDA
 LSALVSPLQAHAQRVAIASTGIIRDGSLALNPHNLGGLLHFPLVKLTLEQLTNLPTIA
 INDAQAAAWAEFQALDGDITDMVFITVSTGVGGGVVSGCKLLTGPGGLAGHIGHTLAD
 PHGPVCGCGRTGCVEAIASGRGIAAAAQGELAGADAKTIFTRAGQGDEQAQQLIHRSA
 RTLARLIADIKATTDCQCVVVGGSVGLAEGYLALVETYLAQEPAAFHVDDLAAHYRHD
 AGLLGAALLAQGEKL " (SEQ ID NO: 40)

25 RBS 4526..4528
 /label = "Native Stop para NanE"

ES 2 700 274 T3

cebador complemento (5065..5083)
/label = "cebador de control nanKckl"

enlace de cebador complemento (5380..5424)
/label = "cebador nanK-H2 lambda rojo"

5 gen 5397..5861
/gene = "yhch"

CDS 5397..5861
/gene = "yhch"
/note = "ECK3210: JW3190: b3221"

10 /codon_start = 1
/transl_table = 11
/product = "proteína hipotética"
/protein_id = "AP_003763.1"
/db_xref = "GI: 89109983"

/translation="MMMGEVQSLPSAGLHPALQDALTLALAARPQEKAPGRYELQGDN
IFMNVMTFNTQSPVEKKAELHEQYIDIQLLLNGEERILFGMAGTARQCEEFHHEDDYQ
LCSTIDNEQAIILKPGMFAVFMPEGPHKPGCVVGEPEIKKVVVKADLMA"
15 (SEQ ID NO: 41)

ORIGEN

ES 2 700 274 T3

1 GAACATTGTT GAACTCCGTG TCAAAAAGAAA ACGGTCAATC CCATAAACGG CAGATTGAAA
 61 ACAACGATGT TATATTTTTT GCAAGGCTAT TTATGGTGCG GATGTCGTGT TTTAATTGT
 121 AGGTGAGGTG ATTTTTCATT AAAAAATATG CGCTTATGAT TATTTTGTA GAACACATTC
 181 ATAATATTCA TAATGCTCGT GAATAGTCTT ATAAATAATT CAAACGGGAT GTTTTTATCT
 241 GCGTTACATT AATTTTTTCG AATAGTTAAT TATTCCGTTA ATTATGGTAA TGATGAGGCA
 301 CAAAGAGAAA ACCCTGCCAT TTTCCCTAC TTTCAATCCT GTGATAGGAT GTCACTGATG
 361 ATGTTAATCA CACTGACCTT ACAGAATGGG CCTTATGAAC GCATTTGATT CGCAAACCGA
 421 AGATTCTTCA CCTGCAATTG GTCGCAACTT GCGTAGCCGC CCGCTGGCGC GTAAAAAACT
 481 CTCCGAAATG GTGGAAGAAG AGCTGGAACA GATGATCCGC CGTCGTGAAT TTGGCGAAGG
 541 TGAACAATTA CCGTCTGAAC GCGAACTGAT GCGTTCCTT AACGTCGGGC GTCCTTCGGT
 601 GCGTGAAGCG CTGGCAGCGT TAAAACGCAA AGGTCTGGTG CAAATAACA ACGGCGAACG
 661 CGCTCGCGTC TCGCGTCTT CTGCGGACAC TATCATCGGT GAGCTTCCG GCATGGCGAA
 721 AGATTTCTT TCTCATCCCG GTGGGATTGC CCATTTGAA CAATTACGTC TGTTCTTTGA
 781 ATCCAGTCTG GTGCGCTATG CCGCTGAACA TGCCACCGAT GAGCAAATCG ATTTGCTGGC
 841 AAAAGCACTG GAAATCAACA GTCAGTCGCT GGATAACAAC GCGGCATTCA TTCGTTTACA
 901 CGTTGATTC CACCGCGTGC TGGCGGAGAT CCCC GGTAAC CCAATCTTCA TGGCGATCCA
 961 CGTTGCCCTG CTCGACTGGC TTATTGCCGC ACGCCCAACG GTTACCGAIC AGGCACTGCA
 1021 CGAACATAAC AACGTTAGTT ATCAACAGCA TATTGCGATC GTTGATGCGA TCCGCCGTCA
 1081 TGATCCTGAC GAAGCCGATC GTGCGTTGCA ATCGCATCTC AACAGCGTCT CTGCTACCTG
 1141 GCACGCTTTC GGTCAGACCA CCAACAAAAA GAAATAATGC CACTTTAGTG AAGCAGATCG
 1201 CATTATAAGC TTTCTGTATG GGGTGTGCT TAATTGATCT GGTATAACAG GTATAAAGGT
 1261 ATATCGTTTA TCAGACAAGC ATCACTTACAG AGGTATTTAT GGCAACGAAT TTACGTGGCG
 1321 TAATGGCTGC ACTCCTGACT CCTTTTGACC AACAAACAAGC ACTGGATAAA GCGAGTCTGC
 1381 GTCGCTGGT TCAGTTCAAT ATTCAGCAGG GCATCGACGG TTTATACGTG GGTGGTTTCA
 1441 CCGGCGAGGC CTTTGTACAA AGCCTTCCG AGCGTGAACA GGTACTGGAA ATCGTCGCCG
 1501 AAGAGGCGAA AGGTAAGATT AAATCATCG CCCACGTCGG TTGCGTCAGC ACCGCCGAAA
 1561 GCCAACAACT TGCGGCATCG GCTAAACGTT ATGGCTTCGA TGCCGTCICG GCCGTCACGC
 1621 CGTTCTACTA TCCTTTCAGC TTTGAAGAAC ACTGCGATCA CTATCGGGCA ATTATTGATT
 1681 CGGCGGATGG TTTGCCGATG GTGGTGTACA ACATTCCAGC CCTGAGTGGG GTAAAACTGA
 1741 CCCTGGATCA GATCAACACA CTTGTTACAT TGCTGGCGT AGGTGCGCTG AAACAGACCT
 1801 CTGGCGATCT CTATCAGATG GAGCAGATCC GTCGTGAACA TCCTGATCTT GTGCTCTATA
 1861 ACGGTTACGA CGAAATCTTC GCCTCTGGTC TGCTGGCGGG CGCTGATGGT GGTATCGGCA
 1921 GTACCTACAA CATCATGGGC TGGCGCTATC AGGGGATCGT TAAGGCGCTG AAAGAAGGCG
 1981 ATATCCAGAC CGCGCAGAAA CTGCAAACTG AATGCAATAA AGTCATTGAT TACTGATCA
 2041 AAACGGGCGT ATCCGCGGC CTGAAAACCTG TCCTCCATTA TATGGATGTC GTTTCTGTGC
 2101 CGCTGTGCCG CAAACCGTTT GGACCGGTAG ATGAAAAATA TCTGCCAGAA CTGAAGGCGC
 2161 TGGCCAGCA GTTGATGCAA GAGCGCGGGT GAGTTGTTT CCCCCTGCTG CCCCTACCGG
 2221 GTGAGGGGAA ATAAACGCAT CTGTACCCTA CAATTTTCAT ACCAAAGCGT GTGGGCATCG
 2281 CCCACCGCGG GAGACTCACA ATGAGTACTA CAACCCAGAA TATCCCGTGG TATCGCCATC
 2341 TCAACCGTGC ACAATGGCGC GCATTTCCG CTGCTGGT TGGATATCTG CTGACGGTT
 2401 TTGATTTCTG TTTAATCGCC CTGGTACTCA CCGAAGTACA AGGTGAATTC GGGCTGACGA
 2461 CGGTGCAGGC GGCAAGTCTG ATCTCTGCAG CCTTTATCTC TCGCTGGTTC GCGGCGCTGA
 2521 TGCTCGGCGC TATGGGTGAC CGCTACGGGC GTCGTCTGGC AATGGTCACC AGCATCGTTC
 2581 TCTTCTCGGC CGGGACGCTG GCCTGCGGCT TTGCGCCAGG CTACATCACC ATGTTTATCG
 2641 CTCGTCTGGT CATCGGCATG GGGATGGCGG GTGAATACGG TTCCAGCGCC ACCTATGTCA
 2701 TTGAAAGCTG GCCAAAACAT CTGCGTAACA AAGCCAGTGG TTTTTTGATT TCAGGCTTCT

ES 2 700 274 T3

2761 CTGTGGGGGC CGTCGTTGCC GCTCAGGTCT ATAGCCTGGT GGTTCCGGTC TGGGGCTGGC
 2821 GTGCGCTGTT CTTTATCGGC ATTTTGCCAA TCATCTTTGC TCTCTGGCTG CGTAAAAACA
 2881 TCCCGGAAGC GGAAGACTGG AAAGAGAAAC ACGCAGGTAA AGCACCAGTA CGCACAATGG
 2941 TGGATATTCT CTACCGTGGT GAACATCGCA TTGCCAATAT CGTAATGACA CTGGCGGCGG
 3001 CTA CTGCGCT GTGGTTCTGC TTCGCCGTA ACCTGCAAAA TGCCGCGATC GTCGCTGTTC
 3061 TTGGGCTGTT ATGCGCCGCA ATCTTTATCA GCTTTATGGT GCAGAGTGCA GGCAAACGCT
 3121 GGCCAACGGG CGTAATGCTG ATGGTGGTCC TGTTGTTTGC TTTCTCTAC TCATGGCCGA
 3181 TTCAGGCGCT GCTGCCAACG TATCTGAAAA CCGATCTGGC TTATAACCCG CATACTGTAG
 3241 CCAATGIGCT GTTCTTTAGT GGCTTTGGCG CGGCGGTGGG ATGCTGCGTA GGTGGCTTCC
 3301 TCGGTGACTG GCTGGGAACC CGCAAAGCGT ACGTTTGTAG CCTGCTGGCC TCGCAGCTGC
 3361 TGATTATTCC GGTATTTGCG ATTGGCGGCG CAAACGTCCT GGTGCTCGGT CTGTTACTGT
 3421 TCTTCCAGCA AATGCTTGA CAAGGGATCG CCGGGATCTT ACCAAAACCTG ATTGGCGGTT
 3481 ATTTTCGATAC CGACCAGCGT GCAGCGGGCC TGGGCTTTAC CTACAACGTT GCGCATTGG
 3541 GCGGTGCACT GGCCCCAATC ATCGGCGCGT TGATCGCTCA ACGTCTGGAT CTGGGTACTG
 3601 CGCTGGCATC GCTCTCGTTC AGTCTGACGT TCGTGGTGAT CCTGCTGATT GGGCTGGATA
 3661 TGCTTCTCG CGTTCAGCGT TGGTTGCGCC CGGAAGCGTT GCGTACTCAT GACGCTATCG
 3721 ACGGTAAACC ATTCAGCGGT GCCGTGCCGT TTGGCAGCGC CAAAAACGAT TTAGTCAAAA
 3781 CCAAAAGTTA ATCCTGTTGC CCGGTCATG TACCGGGCCT TTCGCTAAGG GAAGATGTAT
 3841 GTCGTTACTT GCACAACCTGG ATCAAAAAAT CGCTGCTAAC GGTGGCCTGA TTGTCTCCTG
 3901 CCAGCCGGTT CCGGACAGCC CGCTCGATAA ACCCGAAATC GTCGCCGCCA TGGCATTAGC
 3961 GGCAGAACAG GCGGGCGCGG TTGCCATTCC CATTGAAGGT GTGGCAAATC TGCAAGCCAC
 4021 GCGTGCGGTG GTGAGCGTGC CGATTATTGG AATTGTGAAA CCGCATCTGG AGGATTCTCC
 4081 GGTACGCATC ACGGCCCTATA TTGAAGATGT TGATGCGCTG GCGCAGGCGG GCGCGGACAT
 4141 TATCGCCATT GACGGCACCG ACCGCCCGCG TCCGGTGCCT GTTGAAACGC TGCTGGCACG
 4201 TATTCACCAT CACGGTTTAC TGGCGATGAC CGACTGCTCA ACGCCGGAAG ACGGCCTGGC
 4261 ATGCCAAAAG CTGGGAGCCG AAATTATTGG CACTACGCTT TCTGGCTATA CCACGCCTGA
 4321 AACGCCAGAA GAGCCGGATC TGGCGTGGT GAAAACGTTG AGCGACGCCG GATGTCGGGT
 4381 GATTGCCGAA GGGCGTTACA ACACGCCTGC TCAGGCGGCG GATGCGATGC GCCACGGCGC
 4441 GTGGGCGGTG ACGGTCGGTT CTGCAATCAC GCGTCTTGAG CACATTTGTC AGTGGTACAA
 4501 CACAGCGATG AAAAAGGCGG TGCTATGACC AACTGGCGA TTGATATCGG CGGTACTAAA
 4561 CTTGCCGCCG CGCTGATTGG CGCTGACGGG CAGATCCGCG ATCGTCGTGA ACTTCCTACG
 4621 CCAGCCAGCC AGACACCAGA AGCCTTGCCT GATGCCTTAT CCGCATTAGT CTCTCCGTTG
 4681 CAAGCTCATG CGCAGCGGGT TGCCATCGCT TCGACCGGGA TAAATCCGTA CGGCAGCTTG
 4741 CTGGCGCTTA ATCCGCATAA TCTTGGTGGA TTGCTACACT TTCCGTTAGT CAAAACGCTG
 4801 GAACAACCTA CCAATTTGCC GACCATTGCC ATTAACGACG CGCAGGCCGC AGCATGGGCG
 4861 GAGTTTCAGG CGCTGGATGG CGATATAACC GATATGGTCT TTATCACCGT TTCCACCGGC
 4921 GTTGGCGGCG GTGTAGTGAG CGGCTGCAA CTGCTTACCG GCCCTGGCGG TCTGGCGGGG
 4981 CATATCGGGC ATACGCTTGC CGATCCACAC GGCCAGTCT GCGGCTGTGG ACGCACAGGT
 5041 TGCGTGGAAG CGATTGCTTC TGGTCGCGGC ATTGCAGCGG CAGCGCAGGG GGAGTTGGCT
 5101 GGCGCGGATG CGAAAACCTAT TTTACGCGC GCCGGGCGAG GTGACGAGCA GGCGCAGCAG
 5161 CTGATTCACC GCTCCGCACG TACGCTTGCA AGGCTGATCG CTGATATTAA AGCCACAAC
 5221 GATTGCCAGT GCGTGGTGGT CGGTGGCAGC GTTGGTCTGG CAGAAGGGTA TCTGGCGCTG
 5281 GTGGAAACGT ATCTGGCGCA GGAGCCAGCG GCATTTCATG TTGATTTACT GCGGCGCAT
 5341 TACCGCCATG ATGCAGGTTT ACTTGGGGCT GCGCTGTTGG CCCAGGGAGA AAAATTATGA
 5401 TGATGGGTGA AGTACAGTCA TTACCGTCTG CTGGGTTACA TCCTGCGTTA CAGGACGCGT
 5461 TAACGCTGGC ATTAGCTGCC AGACCGCAAG AAAAAGCGCC GGGTCGTTAC GAATTACAGG
 5521 GCGACAATAT CTTTATGAAT GTCATGACGT TTAACACTCA ATCGCCCGTC GAGAAAAAAG
 5581 CGGAATTGCA CGAGCAATAC ATTGATATCC AGCTGTTATT AAACGGTGAG GAACGGATT
 5641 TGTTTGGCAT GGCAGGCACT GCGCGTCAGT GTGAAGAGTT CCACCATGAG GATGATTATC
 5701 AGCTTTGCAG CACCATTGAT AACGAGCAAG CCATCATCTT AAAACCGGGA ATGTTCCGCC
 5761 TGTTTATGCC AGGTGAACCG CATAAACCAG GATGCGTTGT CGGCGAGCCT GGAGAGATTA
 5821 AAAAGGTTGT GGTGAAGGTT AAGGCTGATT TAATGGCTTA A (SEQ ID NO: 42)

//

REIVINDICACIONES

1. Un método para producir un oligosacárido sialilado en una bacteria que comprende:

proporcionar una bacteria, comprendiendo dicha bacteria una sialil-transferasa exógena, una ruta catabólica de ácido siálico deficiente, una capacidad sintética de ácido siálico, un gen de lactosa permeasa funcional y una mayor capacidad de producción de UDP-GlcNAc que comprende la sobreexpresión de nagC, tal que la bacteria produce al menos 10% más UDP-GlcNAc que una bacteria natural; y

cultivar dicha bacteria en presencia de lactosa.
2. El método de la reivindicación 1, en donde
 - (a) dicha ruta catabólica deficiente en ácido siálico comprende una mutación en cualquiera de los genes seleccionados del gen endógeno de N-acetilneuraminato liasa (nanA), gen endógeno de N-acetilmanosamina quinasa (nanK), gen endógeno de N-acetilmanosamina-6-fosfato epimerasa (nanE), y el gen endógeno del transportador del ácido N-acetilneuramínico (nanT), o cualquier combinación de los mismos; o
 - (b) dicha ruta catabólica deficiente en ácido siálico comprende una mutación en el gen endógeno de la N-acetilneuraminato liasa (nanA), y opcionalmente, una mutación en el gen endógeno del transportador del ácido N-acetilneuramínico (nanT), de preferencia en donde dicho ruta catabólica deficiente en ácido siálico comprende además un gen endógeno de N-acetilmanosamina quinasa (nanK) y un gen endógeno de N-acetilmanosamina-6-fosfato epimerasa (nanE) que no están mutados; o
 - (c) dicha ruta catabólica deficiente en ácido siálico comprende una mutación en el gen endógeno de la N-acetilneuraminato liasa (nanA), una mutación en el gen endógeno de la epimerasa N-acetilmanosamina-6-fosfato (nanE), y opcionalmente, una mutación en el gen endógeno del transportador del ácido N-acetilneuramínico (nanT), preferiblemente en el que dicha ruta catabólica deficiente en ácido siálico comprende además un gen endógeno de quinasa N-acetilmanosamina (nanK) que no está mutado.
3. El método de cualquiera de las reivindicaciones precedentes, en el que la mutación comprende una mutación nula.
4. El método de una cualquiera de las reivindicaciones precedentes, en el que dicha capacidad sintética del ácido siálico comprende un gen exógeno de la sintetasa CMP-Neu5Ac (neuA), un gen exógeno de la sintasa del ácido siálico (neuB) y una UDP-GlcNAc 2-epimerasa exógena (neuC).
5. El método de una cualquiera de las reivindicaciones precedentes, en el que dicho gen exógeno de sialil-transferasa es $\alpha(2,3)$ sialil-transferasa, $\alpha(2,6)$ sialil-transferasa, o $\alpha(2,8)$ sialiltransferasa.
6. El método de cualquiera de las reivindicaciones precedentes, en el que dicho oligosacárido sialilado comprende 3'-sialilactosa (3'-SL) o 6'-sialilactosa (6'-SL).
7. El método de una cualquiera de las reivindicaciones precedentes, en el que dicha bacteria comprende un gen de β -galactosidasa endógeno eliminado o inactivado, por ejemplo, dicho gen de β -galactosidasa eliminado o inactivado comprende el gen de E. coli lac z.
8. El método de la reivindicación 1, en el que dicha bacteria comprende un gen de β -galactosidasa recombinante que proporciona un nivel bajo, pero detectable, de actividad de β -galactosidasa.
9. El método de una cualquiera de las reivindicaciones precedentes, en el que dicha bacteria comprende además un gen lac A delecionado, inactivado o mutado.
10. El método de una cualquiera de las reivindicaciones precedentes, en el que
 - (a) dicha mayor capacidad de producción de UDP-GlcNAc comprende además la sobreexpresión de un gen glmS, un gen glmY, un gen glmZ o cualquier combinación de los mismos; o
 - (b) dicha mayor capacidad de producción de UDP-GlcNAc comprende la sobreexpresión de nagC y glmS; o
 - (c) dicha mayor capacidad de producción de UDP-GlcNAc comprende la sobreexpresión de nagC y glmY; o
 - (d) dicha mayor capacidad de producción de UDP-GlcNAc comprende la sobreexpresión de nagC y glmZ.
11. Un método para producir un oligosacárido que contiene N-acetilglucosamina en una bacteria que comprende:

proporcionar una bacteria, comprendiendo dicha bacteria un gen exógeno UDP-GlcNAc:Gala/pR β 3-N-acetilglucosaminiltransferasa, un gen funcional de la lactosa permeasa y un aumento de la capacidad de producción de UDP-GlcNAc que comprende la sobreexpresión de nagC, de modo que la bacteria produce al menos un 10% más de UDP-GlcNAc que una bacteria nativa; y

cultivar dicha bacteria en presencia de lactosa.

12. El método de la reivindicación 11, en el que dicha bacteria comprende una mayor capacidad de producción de UDP-GlcNAc,

por ejemplo

- 5 (a) dicha mayor capacidad de producción de UDP-GlcNAc comprende además la sobreexpresión de un gen glmS, un gen glmY, un gen glmZ o cualquier combinación de los mismos; o
- (b) dicha mayor capacidad de producción de UDP-GlcNAc comprende la sobreexpresión de nagC y glmS; o
- (c) dicha mayor capacidad de producción de UDP-GlcNAc comprende la sobreexpresión de nagC y glmY; o
- (d) dicha mayor capacidad de producción de UDP-GlcNAc comprende la sobreexpresión de nagC y glmZ.
- 10 13. El método de la reivindicación 11 ó 12, en el que dicho oligosacárido que contiene N-acetilglucosamina comprende cualquiera seleccionado de Lacto-N-triosa 2 (LNT2), Lacto-N-tetraosa (LNT), Lacto-N-neotetraosa (LNnT), Lacto-N-fucopentaosa I (LNF I), Lacto-N-fucopentaosa II (LNF II), Lacto-N-fucopentaosa III (LNF III), Lacto-N-fucopentaosa V (LNF V), Lacto-N-difucohexaosa I (LDFH I), Lacto-N-difucohexaosa II (LDFH II), y Lacto-N-neodifucohexaosa II (LFNnDFH II).
- 15 14. El método de una cualquiera de las reivindicaciones 1-13, en el que dicha bacteria es E. coli.
15. Una bacteria E. coli aislada que comprende
- (a) un gen de β -galactosidasa endógeno eliminado o inactivado, un gen de sialil-transferasa exógeno, una ruta catabólica deficiente en ácido siálico, una capacidad sintética de ácido siálico, un gen funcional de lactosa permeasa y una mayor capacidad de producción UDP-GlcNAc que comprende la sobreexpresión de nagC, de manera que la
- 20 bacteria produce al menos un 10% más de UDP-GlcNAc que una bacteria nativa, por ejemplo, dicha bacteria comprende además un gen de β -galactosidasa recombinante que proporciona un nivel bajo, pero detectable, de actividad de β -galactosidasa; o
- (b) un gen exógeno UDP-GlcNAc:Gala/ β -R β 3-N-acetilglucosaminiltransferasa, un gen funcional de lactosa permeasa y un aumento de la capacidad de producción de UDP-GlcNAc que comprende la sobreexpresión de nagC,
- 25 de modo que la bacteria produce al menos un 10% más de UDP-GlcNAc que una bacteria nativa.
16. El método de cualquiera de las reivindicaciones 1-10, que comprende además recuperar dicho oligosacárido sialilado de dicha bacteria o de un sobrenadante de cultivo de dicha bacteria; o el método de cualquiera de las reivindicaciones 11-13, que comprende además recuperar dicho oligosacárido que contiene N-acetilglucosamina de dicha bacteria o de un sobrenadante de cultivo de dicha bacteria.
- 30 17. El método de la reivindicación 1, en el que dicha bacteria comprende un gen endógeno de N-acetilmanosamina quinasa (nanK) que no está mutado, y un gen endógeno de N-acetilneuraminato liasa (nanA) y un gen endógeno de N-acetilmanosamina-6-fosfato epimerasa (nanE) que están mutados.

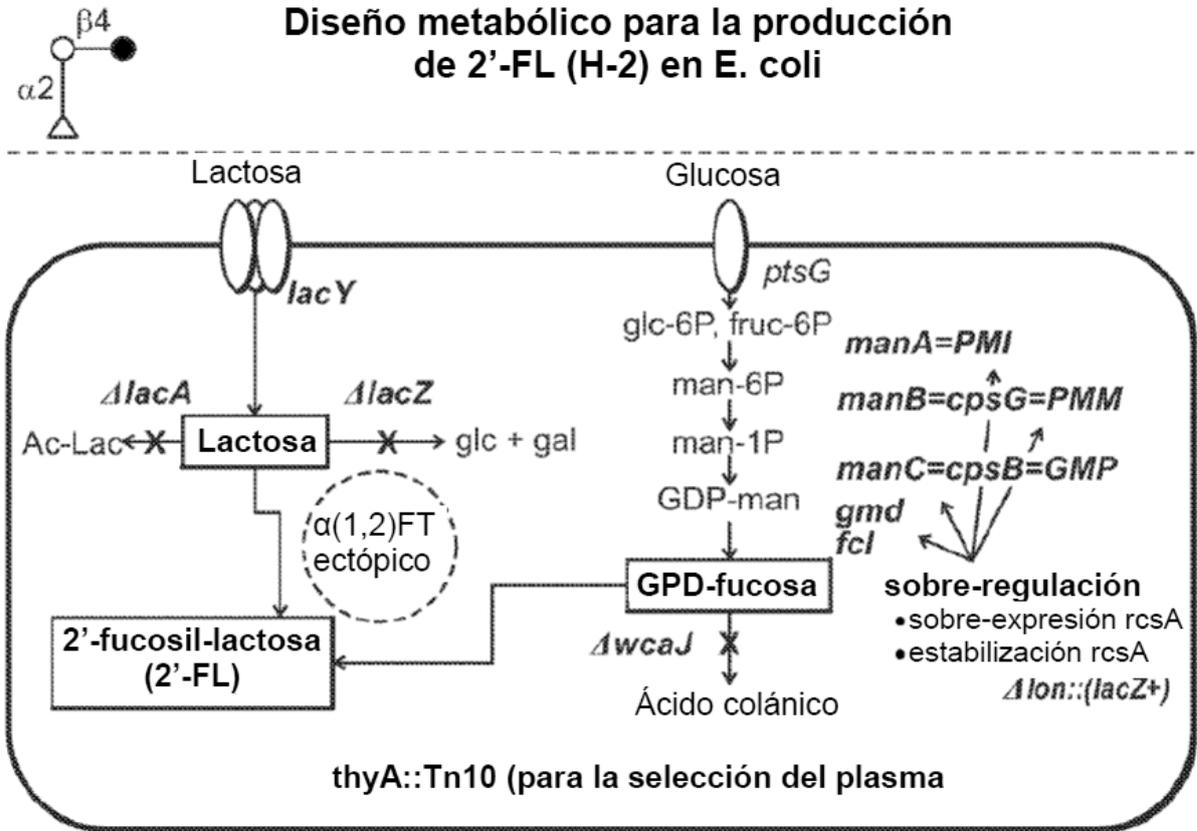


FIG. 1

La acumulación de UDP-GlcNAc en E. coli

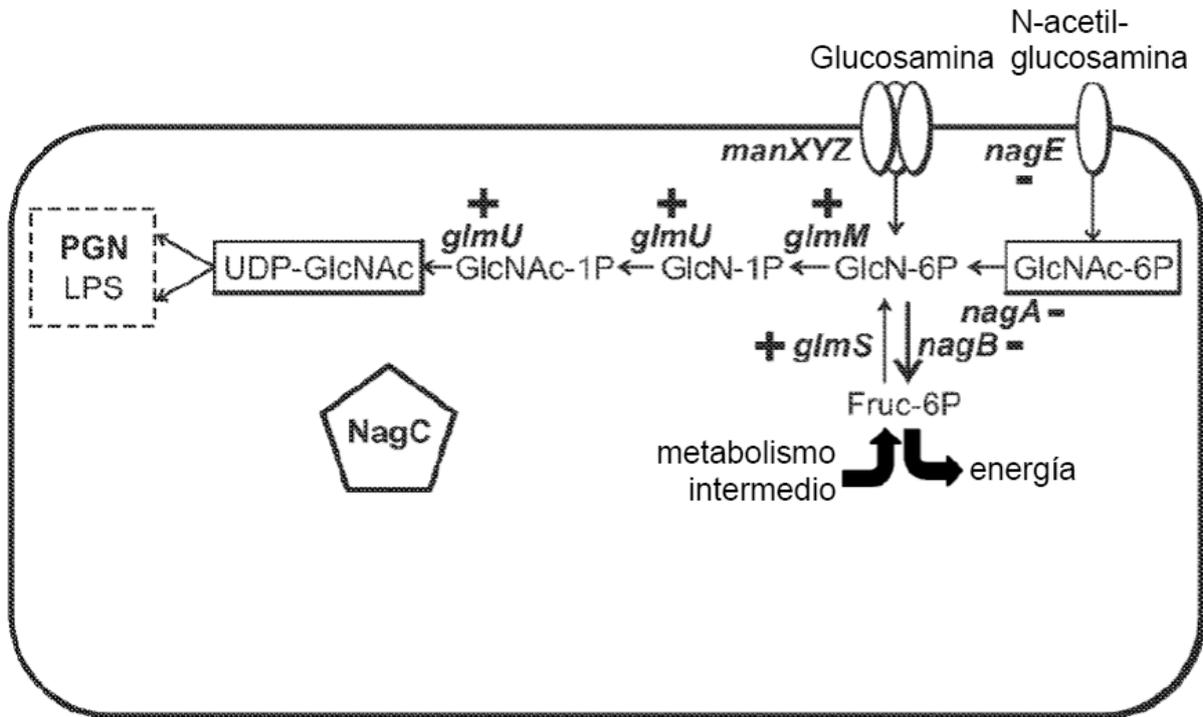


FIG. 2

Diseño de ruta para la producción de 6'-sialil-lactosa en E. coli

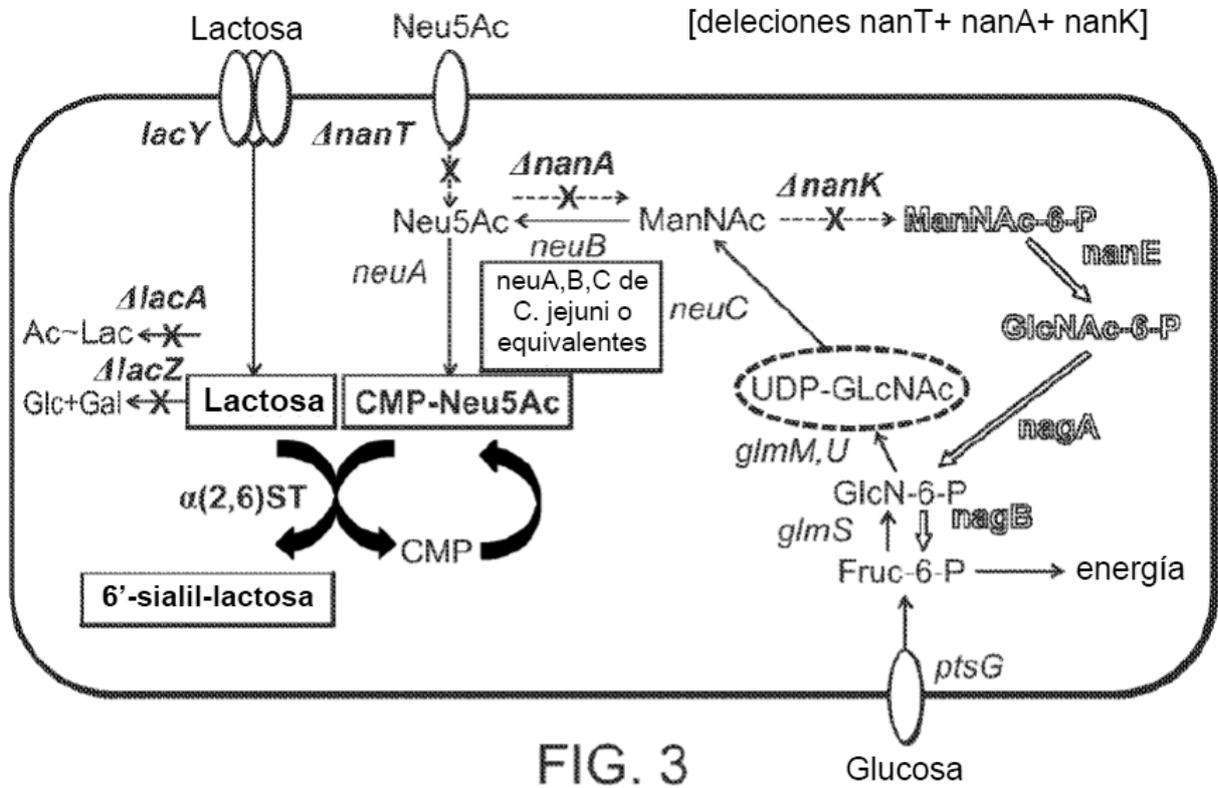


FIG. 3

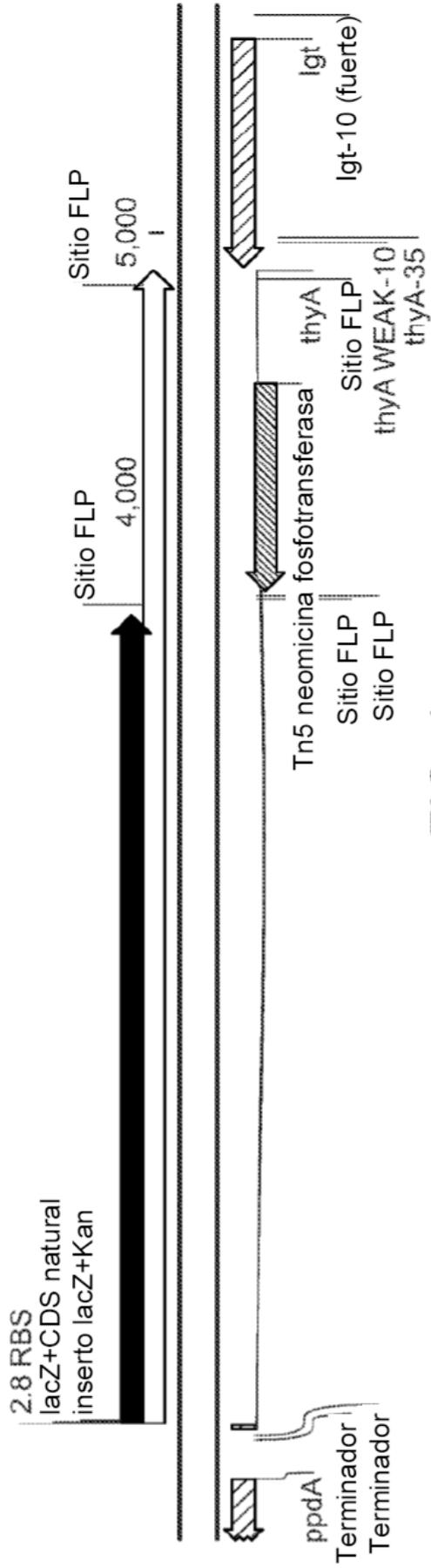


FIG. 4

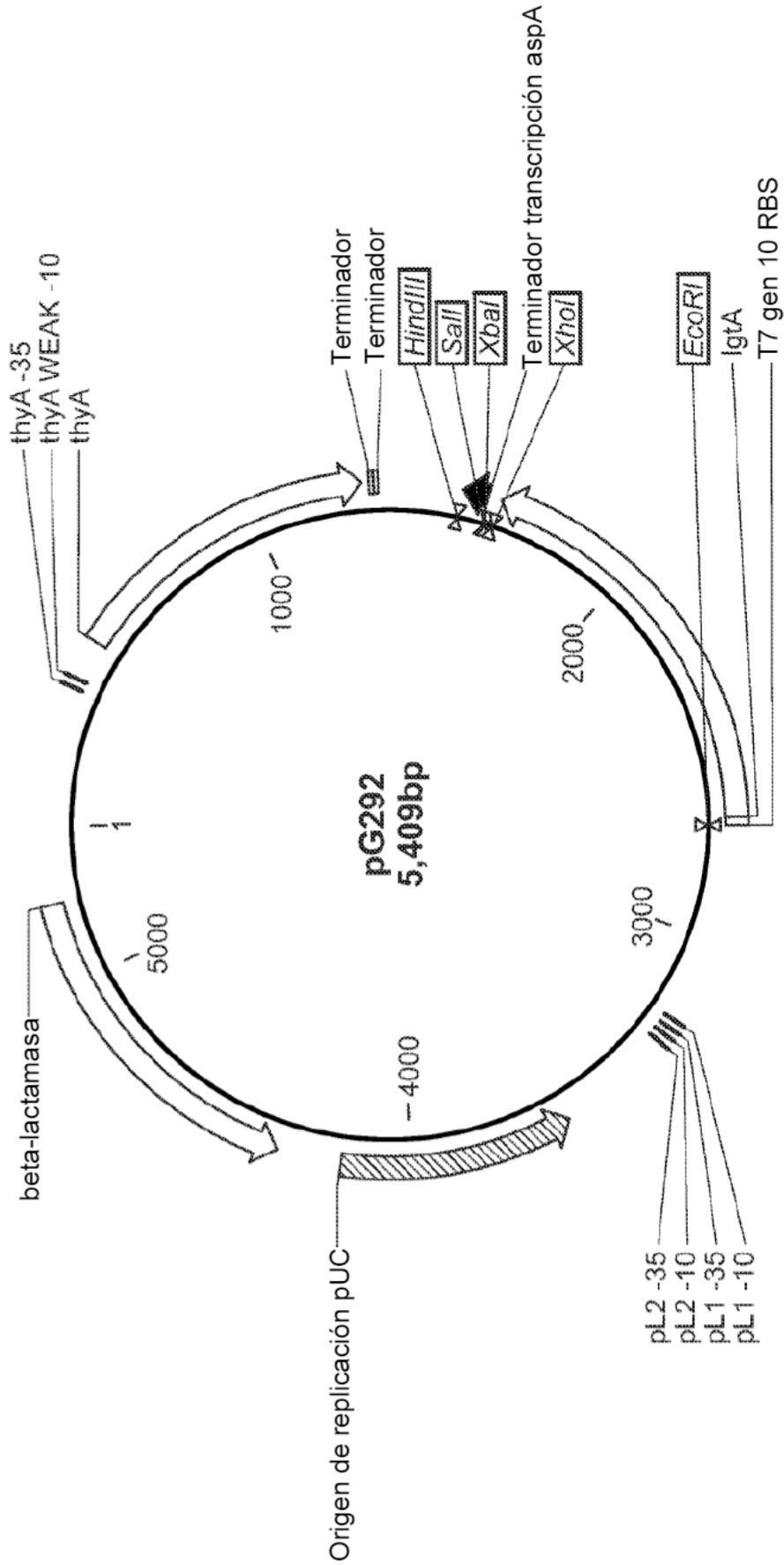


FIG. 5

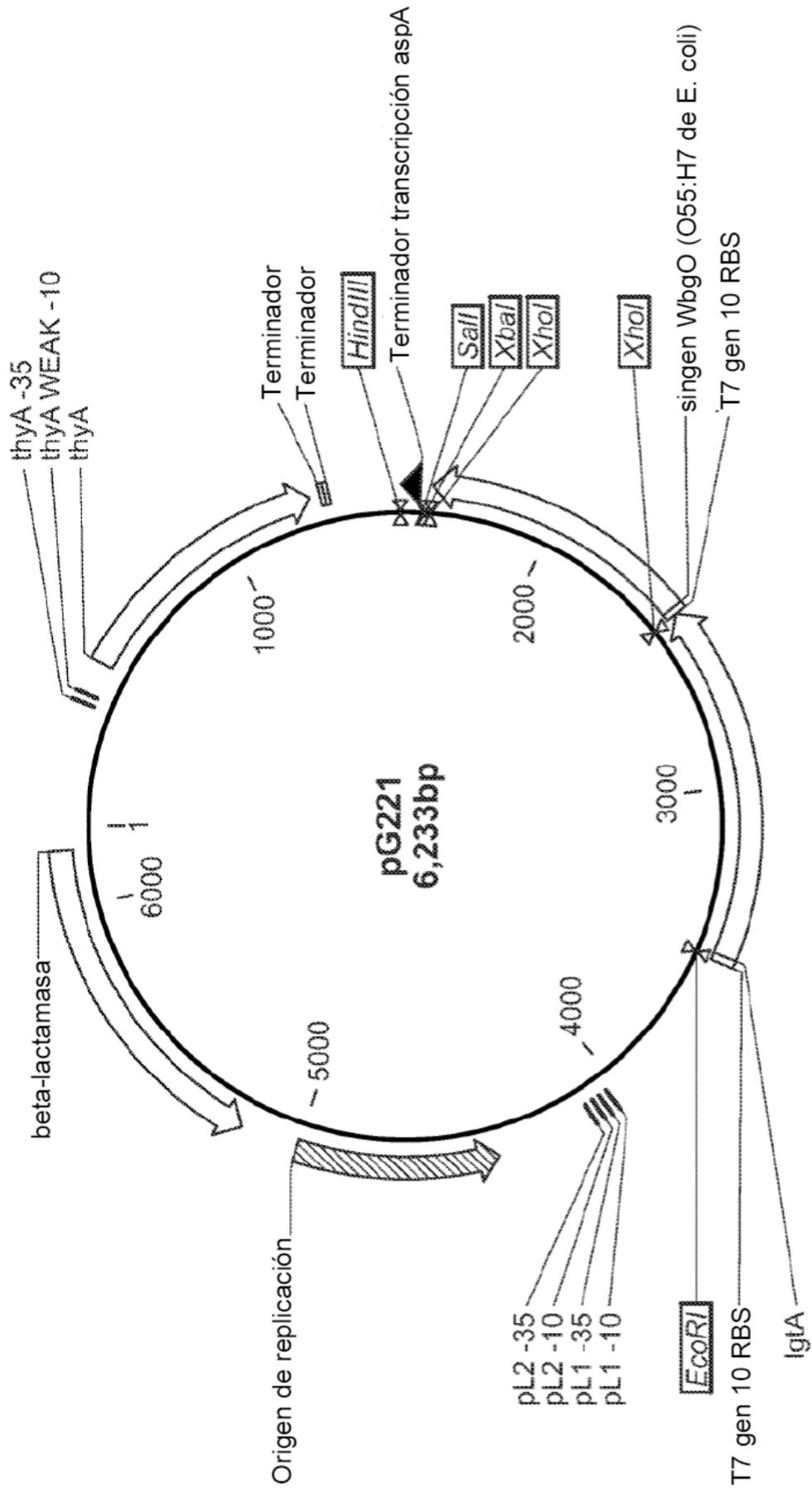


FIG. 6

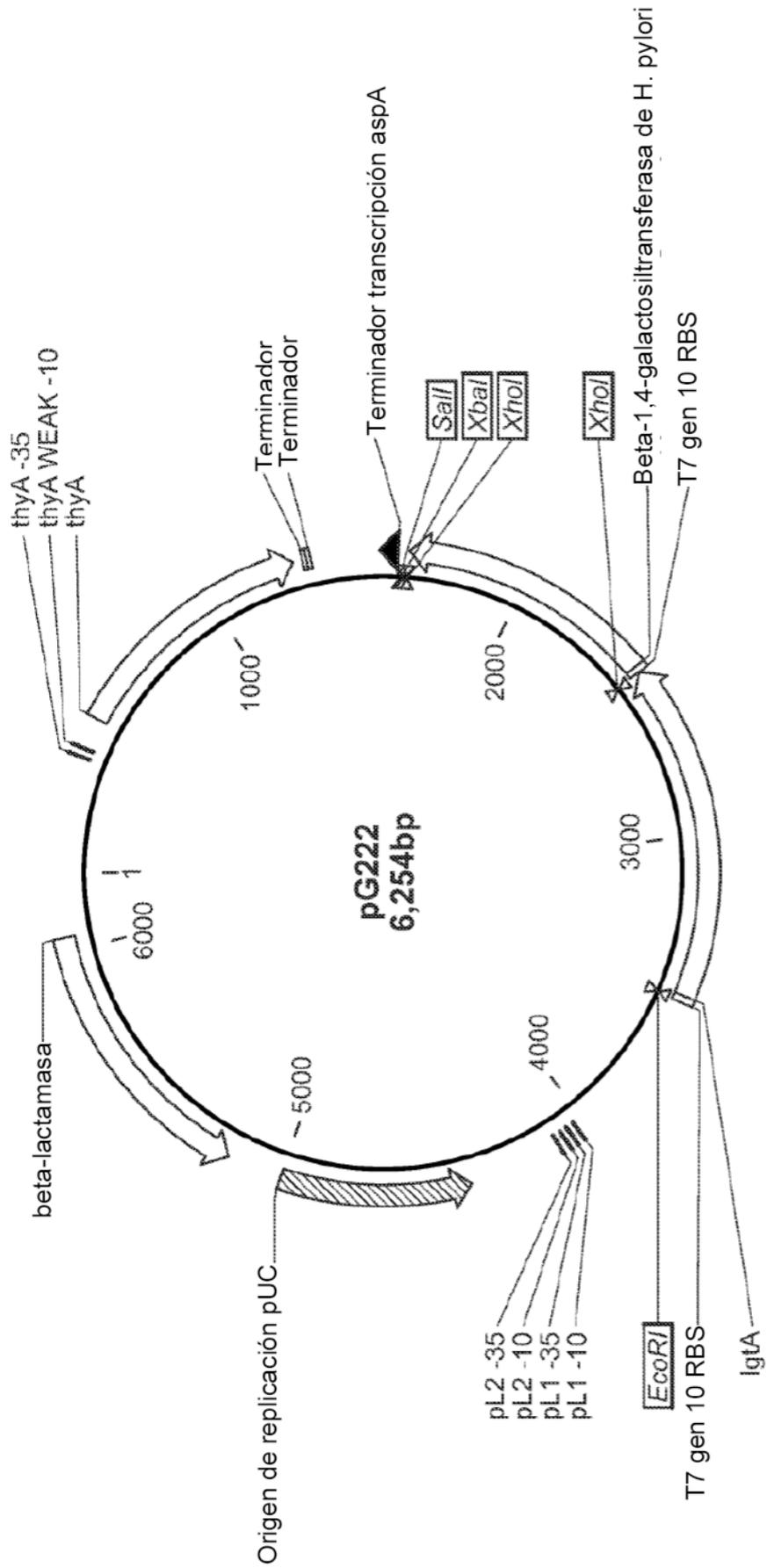
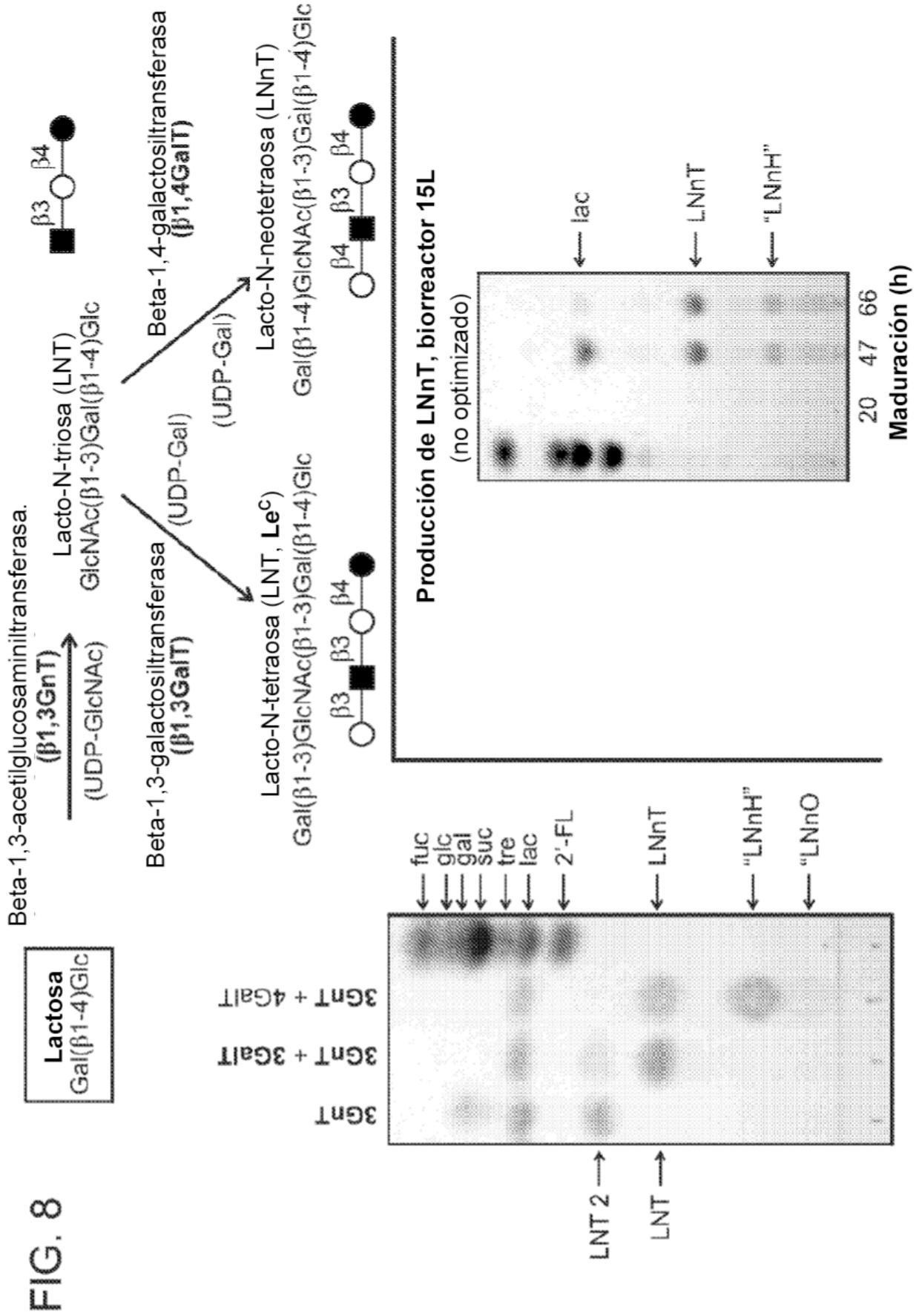


FIG. 7



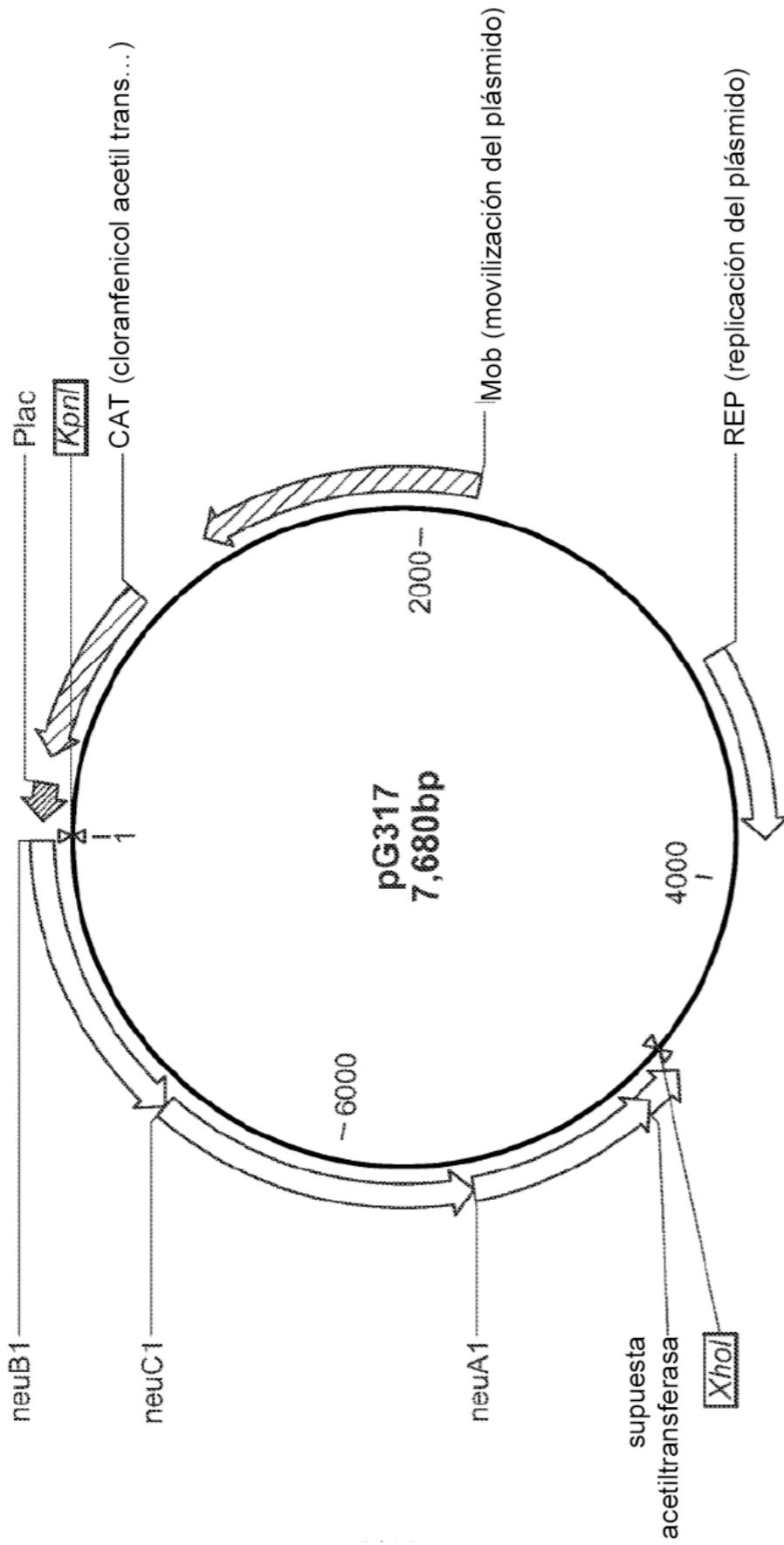


FIG. 9

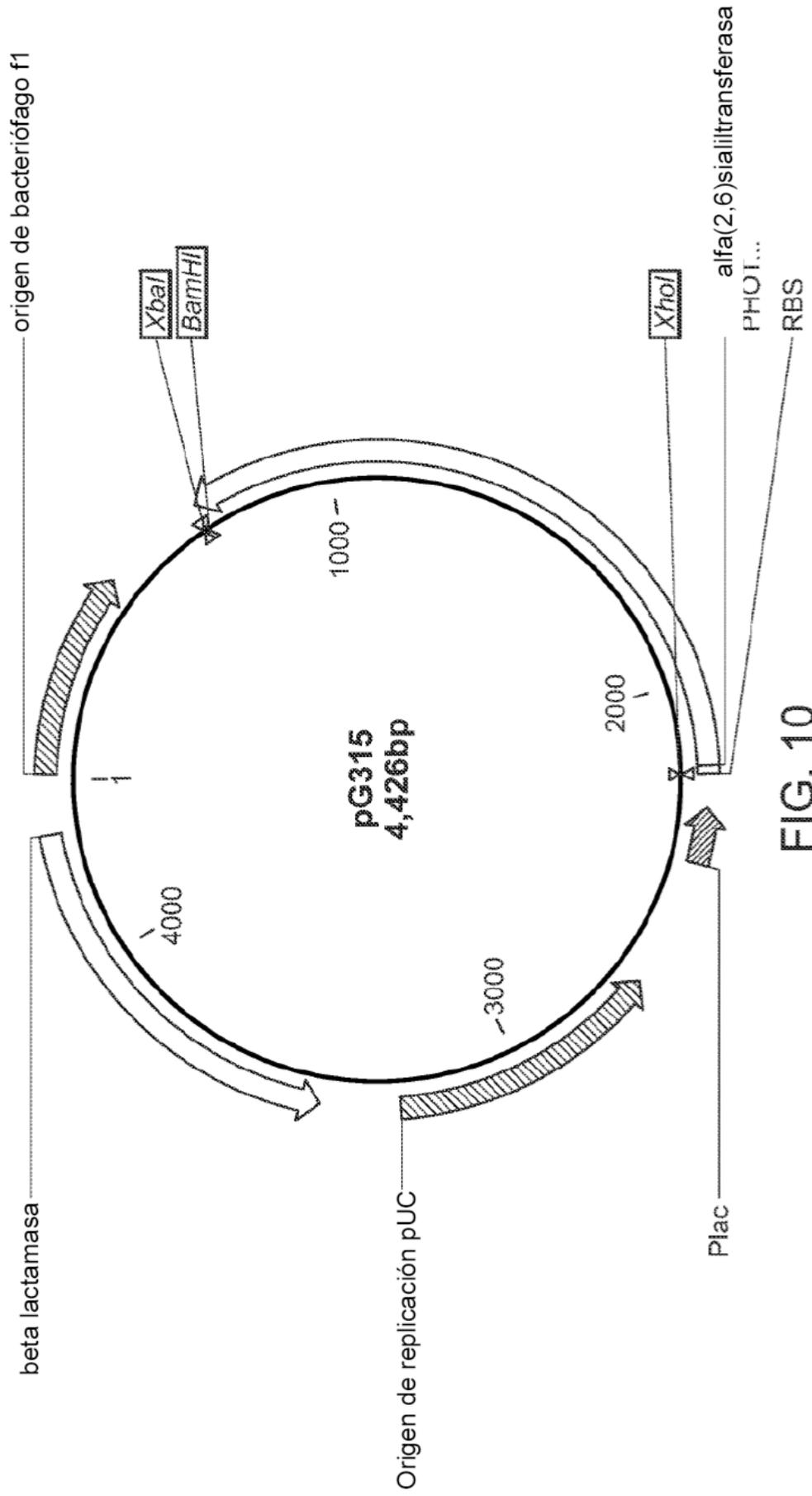
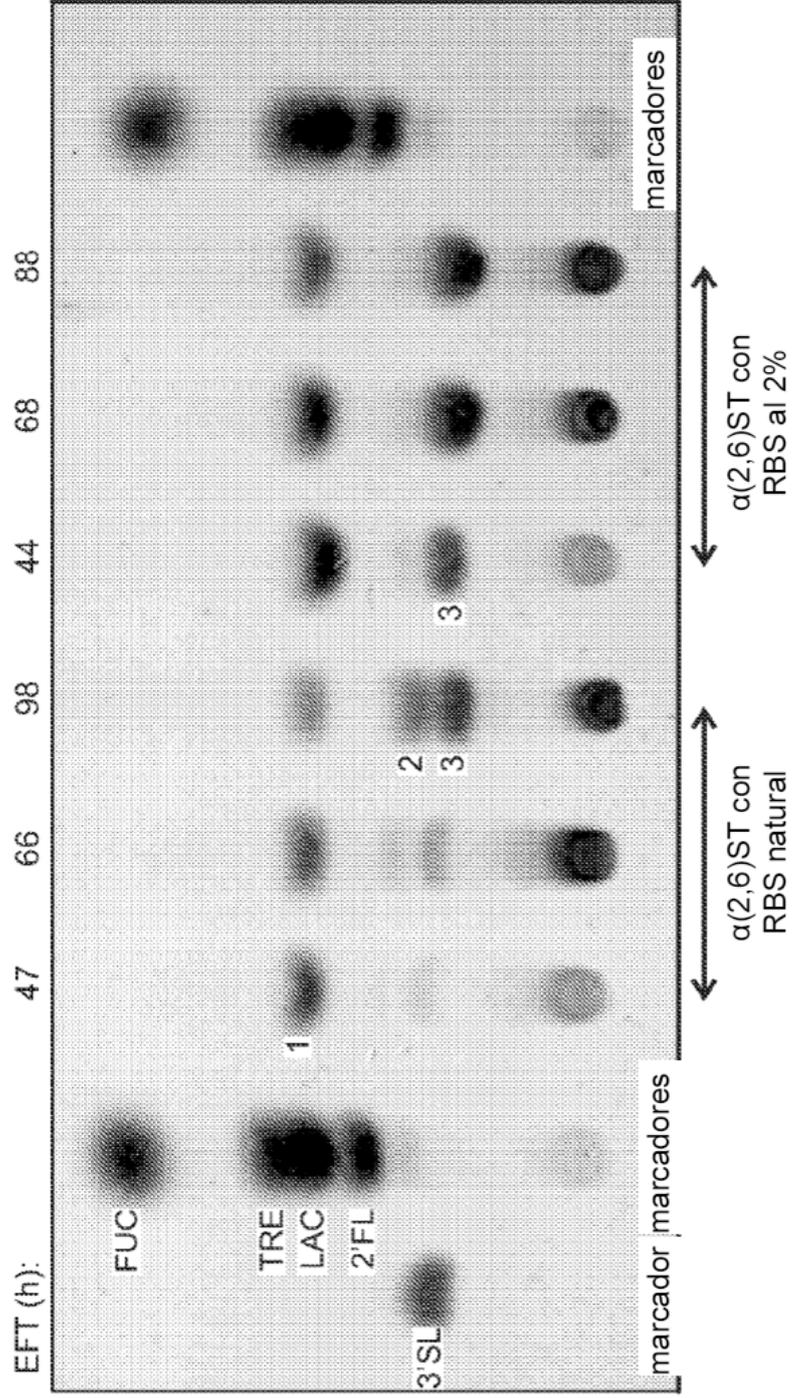


FIG. 10

FIG. 11

Producción de 6'-sialil-lactosa, biorreactor 15 L: análisis TLC de sobrenadantes de



1, lactosa; 2KDO-lactosa; 6'-sialil-lactosa
 3'SL, 3'-sialil-lactosa; 2'FL, 2'-fucosil-lactosa; LAC, lactosa; TRE, trehalosa; FUC, fucosa
 EFT, tiempo de fermentación transcurrido (horas)

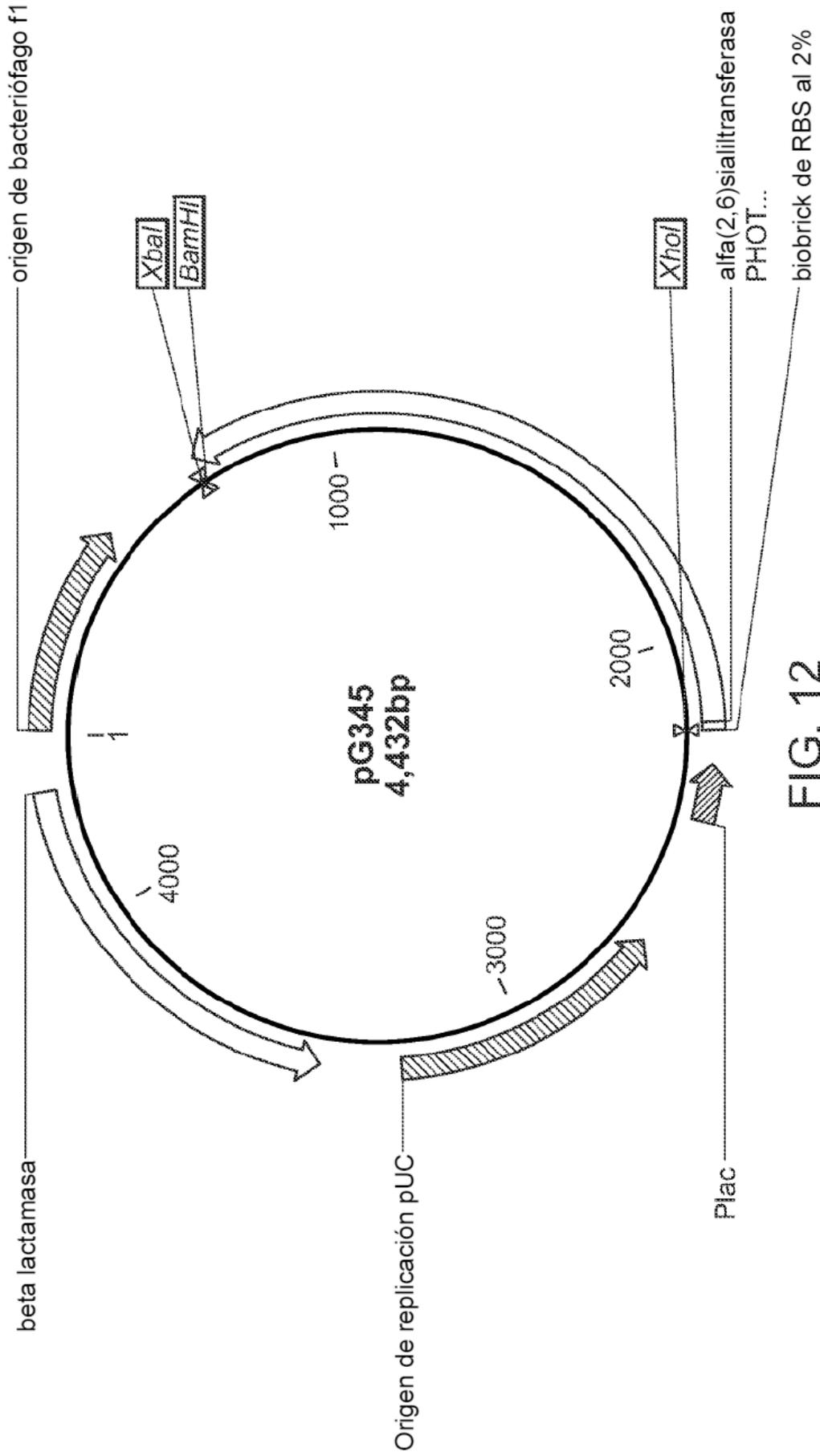


FIG. 12

Diseño de ruta para la producción de 6'-sialil-lactosa en E. coli

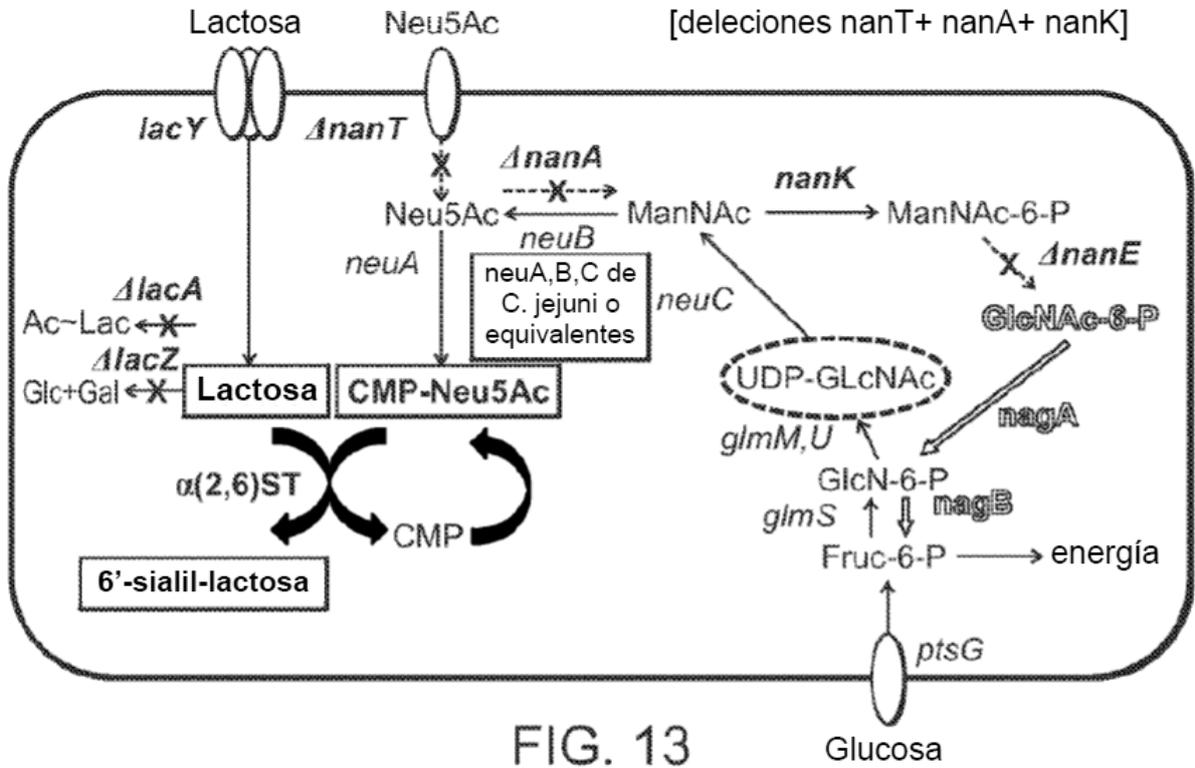
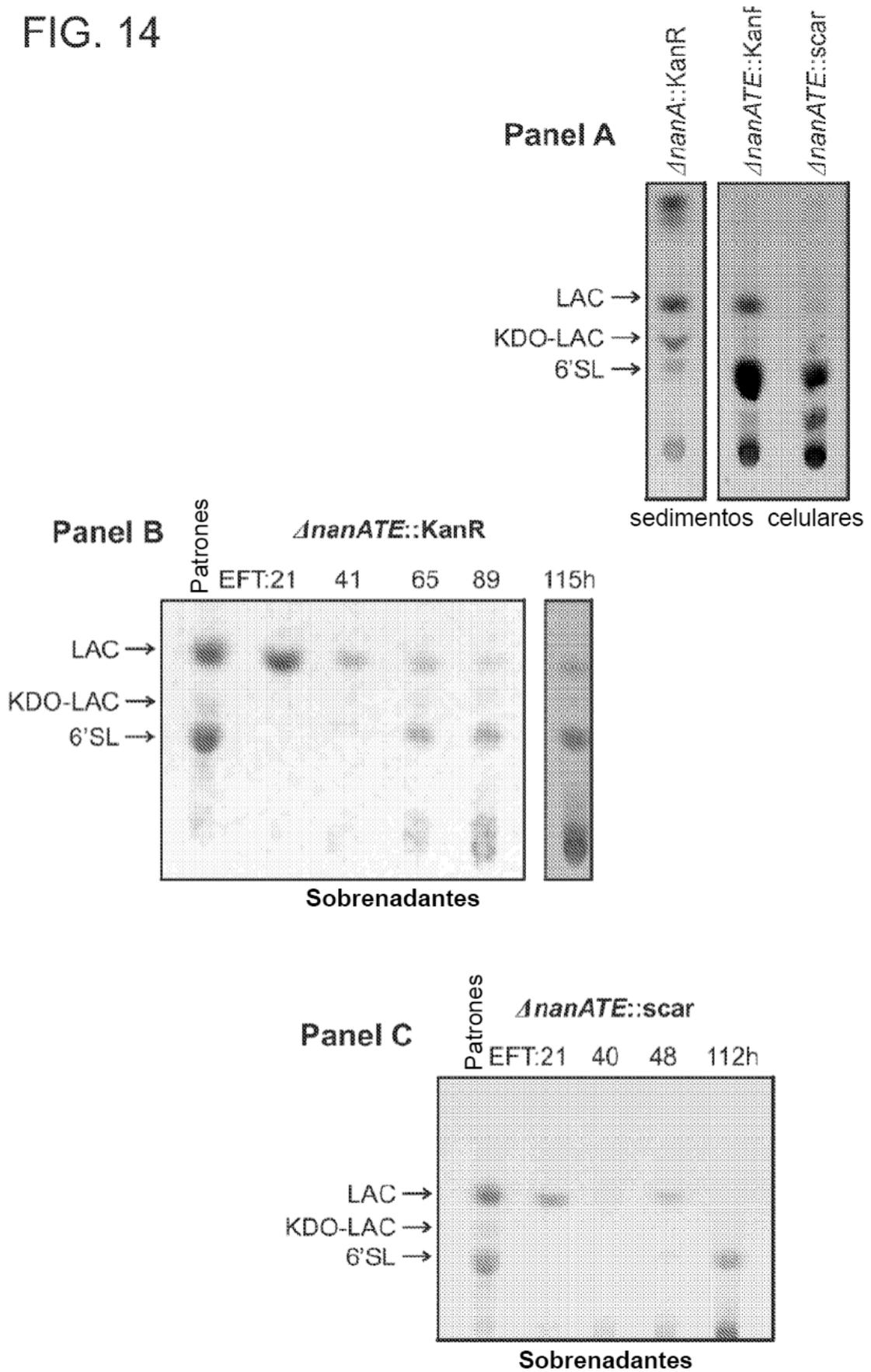


FIG. 13

Glucosa

FIG. 14



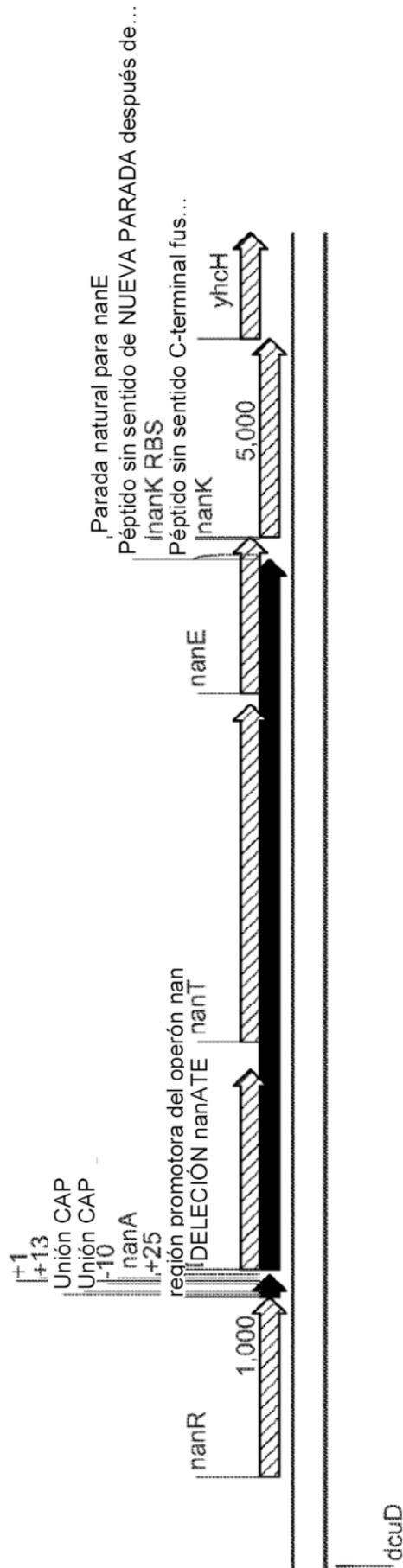
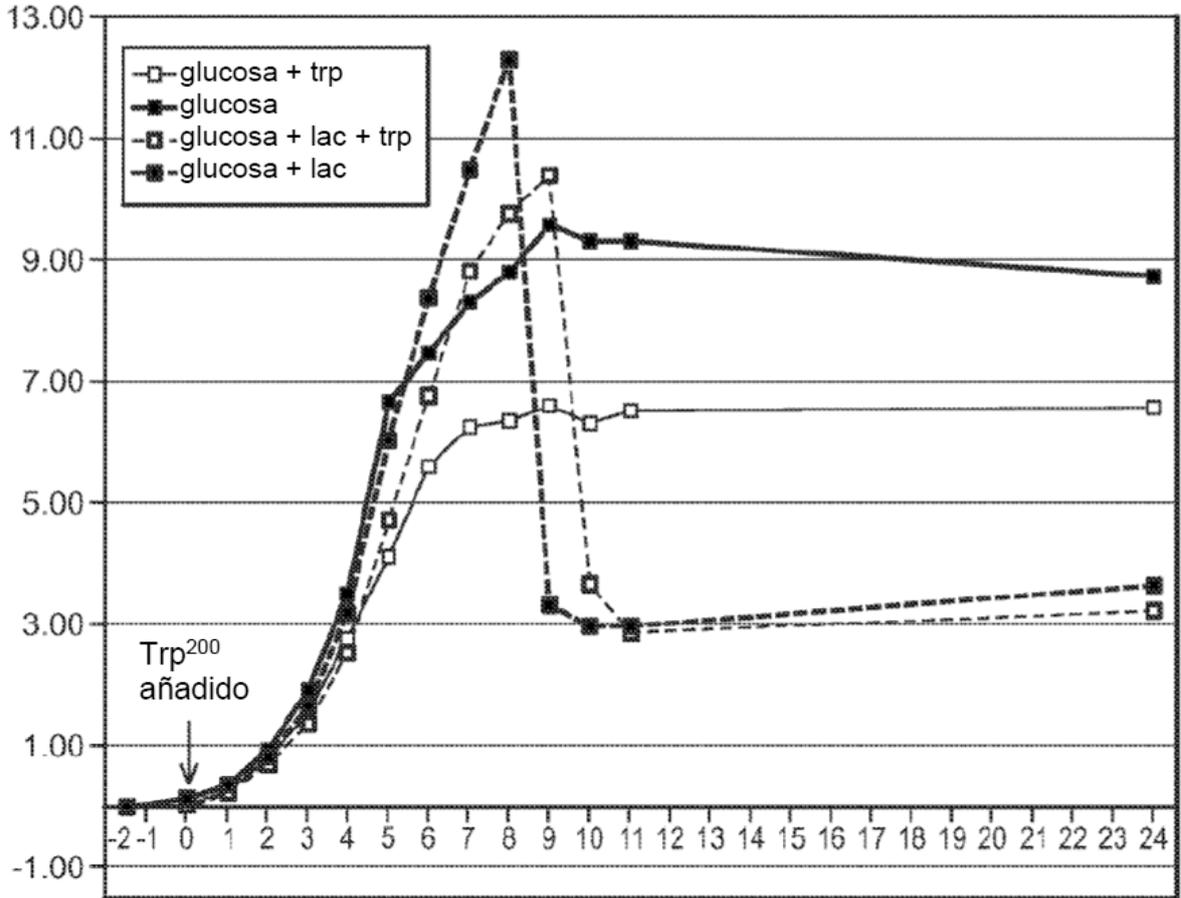


FIG. 15

E680:pG292 (pEC2'-LgtA-thyA, síntesis de LNT2

IMC/ \pm Lactosa al 0,5%/thy⁵⁰/amp¹⁵⁰, \pm inducción de trp²⁰⁰, cultivos de 30 ml, 24 h, 37°C



Toxicidad observada SOLO cuando está presente el aceptor de lactosa

FIG. 16

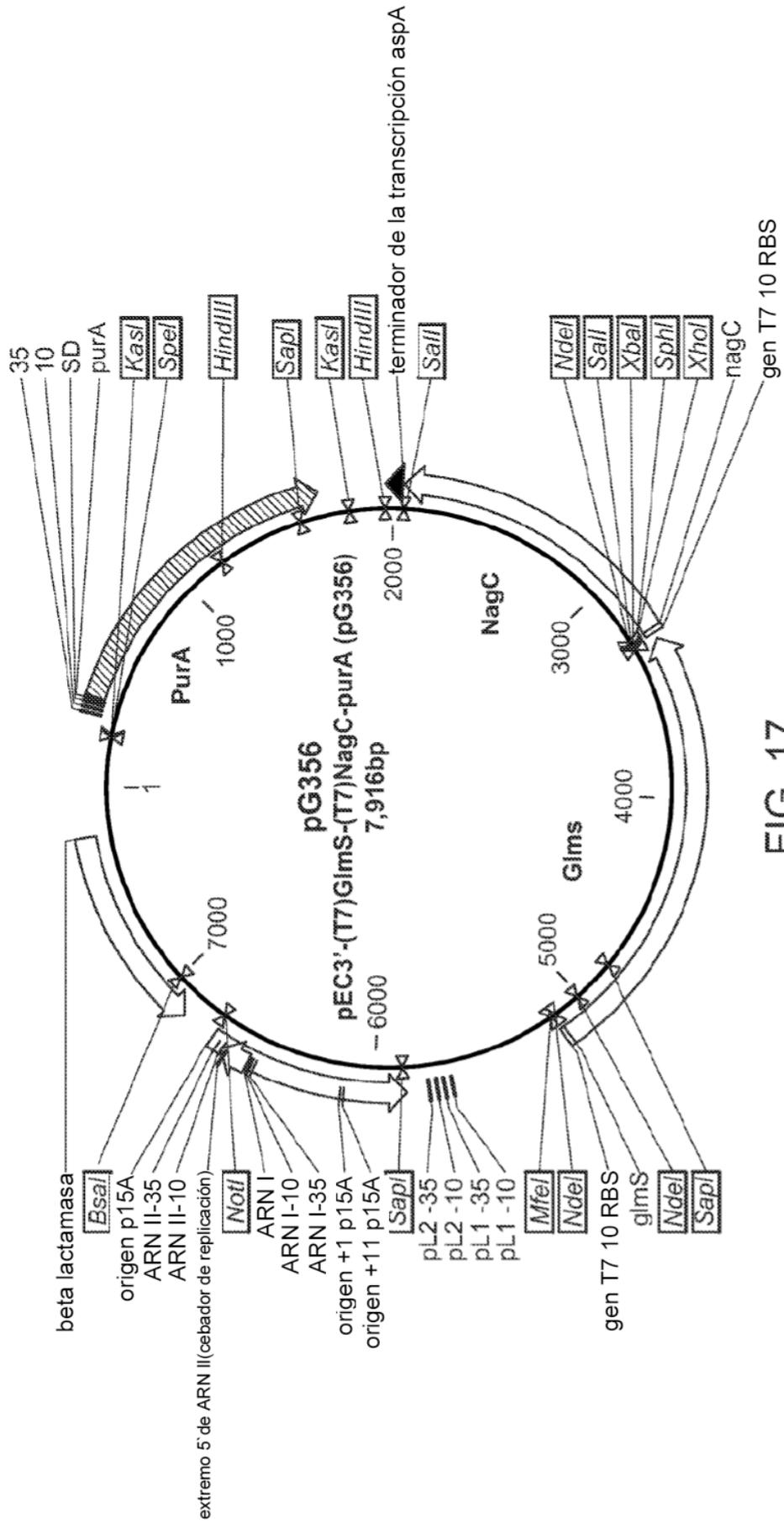


FIG. 17

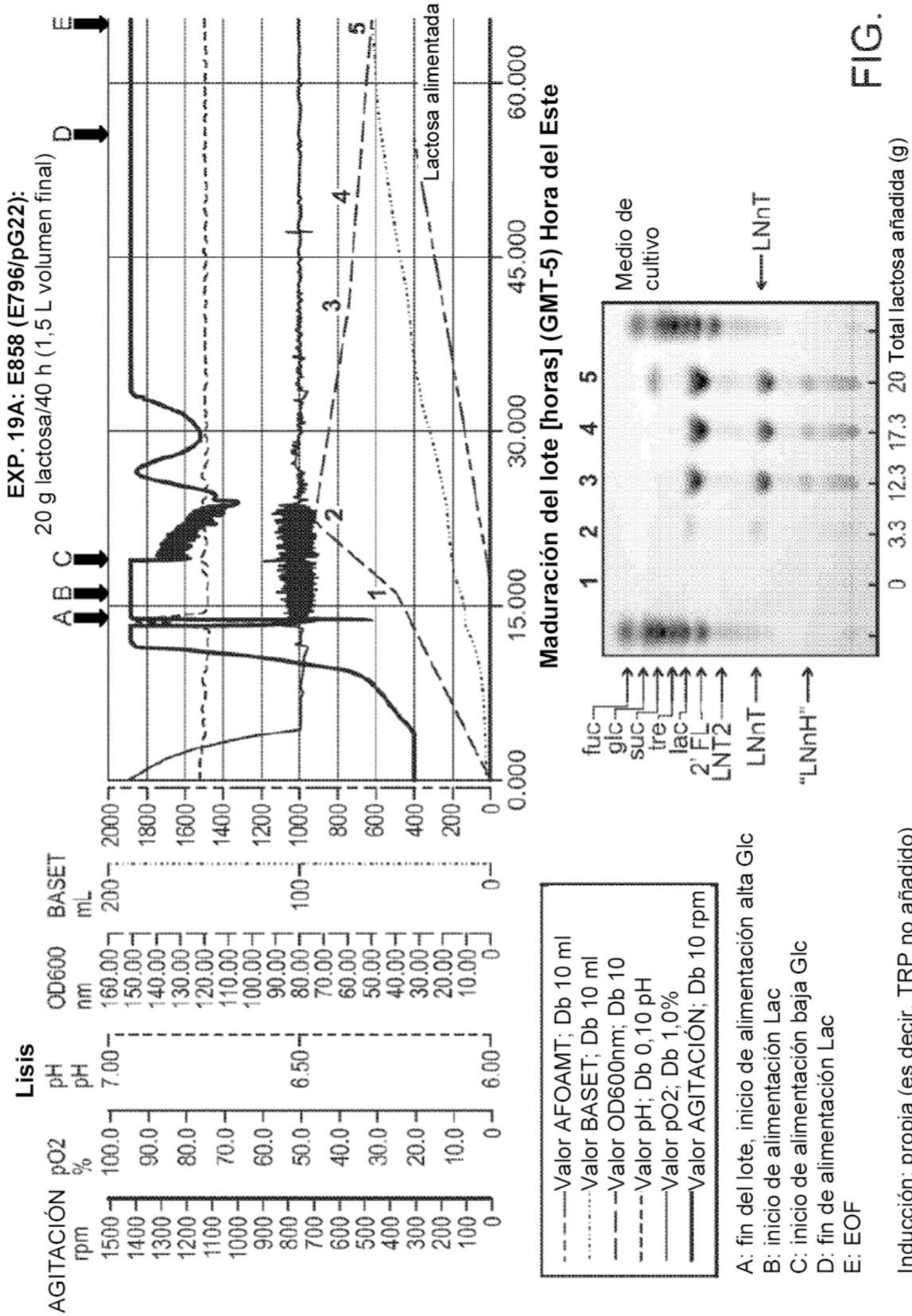


FIG. 18

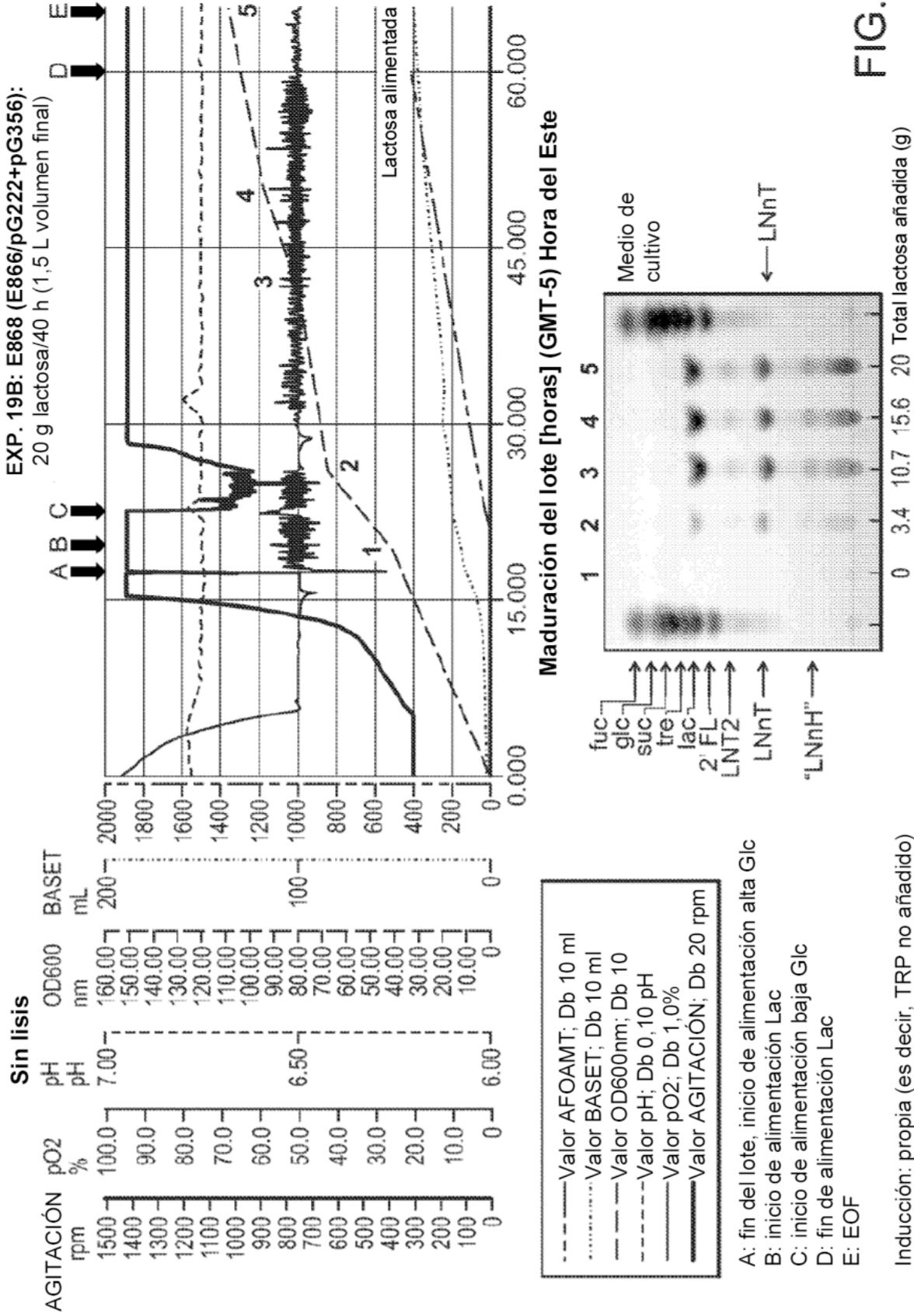


FIG. 19