(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 704 112**

(51) Int. Cl.:

C12N 15/55 (2006.01)
C12N 9/78 (2006.01)
C12N 15/81 (2006.01)
C12N 15/65 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **13.04.2015 PCT/NL2015/050238**

(87) Fecha y número de publicación internacional: **15.10.2015 WO15156675**

(96) Fecha de presentación y número de la solicitud europea: **13.04.2015 E 15726369 (0)**

(97) Fecha y número de publicación de la concesión europea: **24.10.2018 EP 3129486**

(54) Título: **Ureohidrolasas como marcadores dominantes seleccionables en levadura**

(30) Prioridad:

11.04.2014 NL 2012608

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
14.03.2019

(73) Titular/es:

HEINEKEN SUPPLY CHAIN B.V. (100.0%)
Tweede Weteringplantsoen 21
1017 ZD Amsterdam, NL

(72) Inventor/es:

DARAN, JEAN-MARC GEORGES;
PRONK, JACOBUS THOMAS y
ROMAGNOLI, GABRIELE

(74) Agente/Representante:

SÁEZ MAESO, Ana

ES 2 704 112 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

Ureohidrolasas como marcadores dominantes seleccionables en levadura.

Campo

La invención se refiere a los campos de la biología molecular y la ingeniería genética de microorganismos, especialmente de levaduras.

Introducción

El metabolismo de la arginina ha sido objeto de estudios bioquímicos intensivos. Después del descubrimiento del ciclo de la urea para el uso de la arginina como fuente de nitrógeno (Krebs and Henseleit, (1932) Wochenschrift 11: 757-759; Krebs, (1973) Biochemical Education 1: 19-23), la atención se centró en su papel como precursor para la síntesis de poliamina y los compuestos de señalización γ-aminobutirato (GABA) y óxido nítrico (Knowles and Moncada (1994) Biochem J 298: 249-258; Pitkanen et al., (2001) Biochem Biophys Res Commun 287: 1051-1057). La vía más ampliamente distribuida para la degradación de la arginina que ocurre en los tres reinos (Abdelal, (1979) Annu Rev Microbiol 33: 139-168) es iniciada por la arginasa (EC 3.5.3.1), una ureohidrolasa que convierte la arginina en ornitina y urea. Su sitio activo, que contiene varios sitios de unión a Mn^{2+}, también se conserva en otras ureohidrolasas tales como la agmatinasa (EC 3.5.3.11), formiminoglutamase (EC 3.5.3.8) y proclavaminato amidinohidrolasa (EC 3.5.3.22), guanidinobutirasa (EC 3.5.3.7) y guanidinopropionasa (EC 3.5.3.17) (Ouzounis and Kyrpides, (1994) J Mol Evol 39: 101-104). Se supone que los genes que codifican estas enzimas surgieron en una etapa temprana de la evolución (Hartman, (1975) Orig Life 6: 423-427) y se han usado como marcadores en estudios filogenéticos (Ouzounis and Kyrpides (1994) J Mol Evol 39: 101-104; Sekowska et al., (2000) Microbiology 146: 1815-1828).

En los eucariotas, hasta la fecha solo se han descrito dos tipos de ureohidrolasa. Además de la arginasa, los eucariotas superiores expresan agmatinasa (Coleman et al., (2004) Biochem J 379: 849-855), que participa en una ruta alternativa para el catabolismo de la arginina (Figura 1). En esta vía, la arginina se descarboxila primero en agmatina, que se convierte en putrescina y urea por la agmatinasa. La putrescina se puede convertir luego a GABA o a las poliaminas espermina y espermidina (Pegg, (2009) IUBMB Life 61: 880-894). El número cada vez mayor de secuencias genómicas completas ha permitido la identificación putativa de genes de arginasa y agmatinasa en muchos eucariotas. Sin embargo, dado que dicha anotación se basa únicamente en la homología de secuencia, no permite conclusiones definitivas sobre la función catalítica de las proteínas codificadas.

Gran parte del conocimiento sobre el metabolismo de la arginina fúngica se basa en estudios con el organismo modelo *Saccharomyces cerevisiae*. En *S. cerevisiae*, la arginina se transporta a la célula y posteriormente se hidroliza con arginasa (Carl) para producir ornitina y urea (Sumrada and Cooper, (1992) Yeast 8: 311-314; Cooper et al., (1992) J Bacteriol 700 174: 48-55; Shima 64 et al., (2003) Appl Environ Microbiol 69: 715-718). Una amidoliasa dependiente de ATP (Dur1,2) luego convierte la urea en amoníaco y dióxido de carbono. La ornitina se convierte además mediante una transaminasa específica para ornitina (Car2) en glutamato-γ-semialdehído (GSA), que forma espontáneamente 1-pirrolina-5-carboxilato (P5C) (Martin et al., (2003) Appl Environ Microbiol 69: 1623-1628). Debido a la compartimentación subcelular, *S. cerevisiae* no puede convertir P5C citosólico directamente en glutamato (Davis, (1986) Microbiol Rev 50: 280-313). En su lugar, P5C se reduce a prolina usando pirrol-5-carboxilato reductasa (Pro3). Luego, la prolina se transporta a las mitocondrias (Brandriss and Falvey, (1992) Bacteriol 174: 5176), se convierte de nuevo en P5C mediante una oxidasa (Put1) y, finalmente, se convierte en glutamato mediante la deshidrogenasa mitocondrial P5C (Put2) (Davis, (1986) Microbiol Rev 50: 280-313). Dado que solo se han realizado muy pocos estudios fisiológicos sobre el metabolismo de la arginina en levaduras que no son *Saccharomyces*, se desconoce si la vía de la arginasa, que es esencial para el crecimiento de *S. cerevisiae* en la arginina como única fuente de nitrógeno (Bossinger and Cooper, (1977) J Bacteriol 131: 163-173, es la única vía fúngica para el catabolismo de la arginina.

S. cerevisiae y *Kluyveromyces lactis* pertenecen a la familia Saccharomycetaceae. Se considera que estas dos levaduras relacionadas se han separado genéticamente antes del evento de duplicación del genoma completo (WGD) que reformó el genoma de *S. cerevisiae*; además se considera que *K. lactis* se asemeja a un progenitor pre-WGD de *S. cerevisiae* (Dujon, (2010) Nat Rev Genet 11: 512-524). Si bien se han realizado muchos estudios sobre las diferencias en el metabolismo del azúcar entre estas dos especies, las diferencias en el metabolismo de los aminoácidos no se han estudiado en detalle. No obstante, la secuencia completa del genoma de *K. lactis* reveló muchos supuestos ortólogos de genes de *S. cerevisiae* implicados en el metabolismo de la arginina (Dujon et al., (2004) Nature 430: 35-44; Souciet et al., (2009) BMC Genomics 13: 517; Dias et al., (2012) BMC Genomics 13: 517).

La selección de un microorganismo que se transforma con ADN recombinante se facilita en gran medida mediante el uso de un marcador de selección apropiado. El biólogo molecular que trabaja con *Saccharomyces cerevisiae* tiene acceso a una gran cantidad de marcadores seleccionables (Solis-Escalente et al., (2013) FEMS Yeast Research 13: 126-139). Sin embargo, los marcadores auxotróficos y los marcadores de resistencia a antibióticos a veces no son deseados, lo que disminuye vertiginosamente el número de marcadores apropiados. Finalmente, el genetista molecular se queda con un número limitado cuando quiere acceder genéticamente a las cepas de levaduras prototróficas de tipo silvestre, alopoliploides y/o aneuploides.

El estudio del metabolismo de la arginina dio como resultado la identificación de dos nuevos marcadores dominantes de "ganancia de función", que se pueden usar, por ejemplo, para introducir alteraciones genómicas en microorganismos, preferiblemente en cepas de levadura de laboratorio, silvestres e industriales, incluyendo cepas *S. cerevisiae*.

Por lo tanto, la invención proporciona un método de cultivo de un microorganismo seleccionado de los géneros *Saccharomyces sensu stricto*, *Kazachstania*, *Naumovozyma*, *Nakaseomyces* y *Vanderwaltozyma* en presencia de guanidinobutirato como única fuente de nitrógeno, que comprende (a) la introducción de una molécula de ácido nucleico que comprende una secuencia de nucleótido que codifica una guanidinobutirasa de la familia ureohidrolasa en el microorganismo, por lo que la secuencia de nucleótidos está unida operativamente a secuencias del promotor y de terminación; (b) cultivar el microorganismo de modo que la molécula de ácido nucleico que codifica la guanidinobutirasa se exprese en el microorganismo; y (c) cultivar el microorganismo en presencia de guanidinobutirato como única fuente de nitrógeno, en la que la guanidinobutirasa codificada es una hidrolasa ácida de guanidino con el número EC 3.5.3.7. La molécula de ácido nucleico, también denominada casete de marcador dominante proporciona un sistema de marcador seleccionable dominante apropiado para uso en microorganismos, preferiblemente en levadura.

La guanidinobutirosa (hidrolasa ácida de guanidino) e hidrolasa amida de guanidino pertenecen a una familia más grande de proteínas, la ureohidrolasa, que comprende enzimas que comparten una estructura alfa-beta-alfa de 3 capas y juegan un papel importante en el metabolismo de la arginina/agmatina, el ciclo de la urea y degradación de histidina. La guanidinobutirasa cataliza la hidrólisis del 4-guanidinobutanoato en 4-aminobutanoato y urea. La hidrolasa amida de guanidino hidroliza la agmatina a urea y putrescina, el precursor de la biosíntesis de poliaminas, espermidina y espermina. La familia de las ureohidrolasas comprende otros tres subgrupos: i) el hidrolasa aminoácido de guanidino (arginasa, EC3.5.3.1) que cataliza la conversión de arginina en urea y ornitina, ii) el proclavaminato amidinohidrolasa (EC 3.5.3.22), una actividad es decir, cataliza la conversión de proclavaminato de amidino en urea y proclavaminato, un intermedio en la biosíntesis del ácido clavulánico y iii) la formiminoglutamasa (EC 3.5.3.8) que cataliza la cuarta etapa en la degradación de histidina, mediante la hidrólisis de N-formimidoil-L-glutamato a L-glutamato y formamida.

Dicha secuencia de nucleótidos que codifica guanidinobutirasa codifica preferiblemente una hidrolasa ácida de guanidino que tiene uno o ambos dominios conservados como se indica en la figura 2A y se representa en la figura 2B, y/o codifica la secuencia de consenso como se representa en la figura 2A. Dicha hidrolasa ácida de guanidino comprende preferiblemente entre 350 y 450 aminoácidos, preferiblemente aproximadamente 400 aminoácidos. Dicha secuencia de nucleótidos que codifica guanidinobutirasa codifica preferiblemente la proteína hipotética NRRL Y-1140 de *Kluyveromyces lactis*, que tiene la secuencia de GenBank XP_456325.1, como se representa en la figura 3.

Otra secuencia de nucleótidos que codifica hidrolasa amida de guanidino descrita codifica agmatina ureohidrolasa (agmatinasa) (EC.3.5.3.11). La secuencia de nucleótidos codifica preferiblemente una agmatina ureohidrolasa que tiene el dominio conservado como se indica en la figura 4A y se representa en la figura 4B, y/o codifica la secuencia de consenso como se representa en la figura 4A. Dicha agmatina ureohidrolasa comprende preferiblemente entre 290 y 330 aminoácidos, preferiblemente aproximadamente 306 aminoácidos. Dicha agmatina ureohidrolasa codifica preferiblemente la secuencia de GenBank AAC75974.1, como se representa en la figura 5.

Tanto la guanidinobutirasa como la agmatinasa catalizan la formación de urea, una fuente de nitrógeno comúnmente asimilada por microorganismos tal como *S. cerevisiae*. Por lo tanto, estos dos genes de ureohidrolasa presentan las características esenciales de un marcador seleccionable de "ganancia de función" dominante potencial en microorganismos tales como *S. cerevisiae*, cuando se cultivan en guanidinobutirato y/o agmatina como única fuente de nitrógeno.

Una molécula de ácido nucleico que comprende una secuencia de nucleótidos que codifica un marcador de selección de guanidinobutirasa está preferiblemente unida operativamente a secuencias del promotor y de terminación (heterólogas). Dicho promotor dirige la expresión del marcador de selección en el microorganismo. Dicho terminador media en la eficaz formación del extremo 3' del ARNm. Dicho promotor es preferiblemente un promotor de levadura, más preferiblemente un promotor de levadura seleccionado de un gen glicolítico tal como promotor PGI1 (fosfoglucosa isomerasa 1), PFK1 (fosfofructoquinasa-1), PFK2 (fosfofructoquinasa-2), FBA1 (fructosa-bisfosfato aldolasa-1), TPI1 (triosefosfato isomerasa-1), TDH1 (gliceraldehído-3-fosfato deshidrogenasa 1), TDH3 (gliceraldehído-3-fosfato dehidrogenasa3), PGK1 (fosfoglicerato quinasa 1), GPM1 (glicerato fosfomutasa 1), PYK1 (piruvato quinasa 1), ENO1 (alfa-enolasa), y/o ENO2 (enolasa 2), o seleccionado de promotor ACT1 (actina 1), TEF1 (factor de alargamiento de la traducción EF-1 alfa), TEF2 (factor de elongación de la traducción EF-1 alfa 2), AgTEF2 (gen TEF2 de *Ashbya gossypii*), PMA1 (membrana plasmática H+-ATPasa tipo P2). Las secuencias del promotor preferidas se seleccionan de secuencia del promotor de los genes PGI1, FBA1; TPI1; TDH3; PGK1; GPM1; ENO1; ENO2; y PYK1. Los terminadores de un número de genes son conocidos por los expertos y se han empleado, por ejemplo, en vectores de expresión, incluidos los terminadores de los genes CYC1, TRP1, ADH1, MFl, FLP y D (Romanos et al., 1992. Yeast 8: 423- 488).

Las secuencias de promotor preferidas y las secuencias de terminación no comprenden secuencias de *Saccharomyces*. Las secuencias del promotor y secuencia de terminación preferidas son del gen TEF de *Ashbya gossypii* que codifica los factores de elongación de la traducción y/o las secuencias reguladoras de los genes URA3 y LEU2 de *K. lactis* que codifican la orotidina-5'-fosfato (OMP) descarboxilasa y la beta-isopropilmalato deshidrogenasa, respectivamente.

Una molécula de ácido nucleico que codifica una guanidinobutirasa se proporciona preferiblemente como un producto amplificado o como parte de un producto amplificado. Dicho producto amplificado puede comprender además secuencias homólogas a una primera parte del genoma de un microorganismo, preferiblemente una parte corriente arriba de un gen que se va a alterar (denominado gen diana), y secuencias homólogas a la segunda parte del genoma de un microorganismo, preferiblemente una parte corriente abajo del gen diana, adyacente a la molécula de ácido nucleico. El término adyacente se usa para indicar que las secuencias homólogas a la primera parte del genoma están ubicadas en un lado de la molécula de ácido nucleico, mientras que las secuencias homólogas a la segunda parte del genoma están ubicadas en el otro lado de la molécula de ácido nucleico. Las secuencias homólogas a la primera y segunda parte del genoma comprenden preferiblemente entre 20 y 1000 pb, más preferiblemente entre 30 y 500 pb, más preferiblemente entre 40 y 250 pb, más preferiblemente entre 50 y 80 pb de secuencias. Las secuencias homólogas a la parte corriente arriba de un gen incluyen preferiblemente el codón de inicio. Las secuencias homólogas a la parte corriente abajo de un gen incluyen preferiblemente el codón de parada. Los métodos para producir un producto amplificado que comprende una molécula de ácido nucleico que codifica una guanidinobutirasa son conocidos por un experto que incluyen, por ejemplo, reacción en cadena de la polimerasa y amplificación basada en la secuencia de ácido nucleico (NASBA).

Una molécula de ácido nucleico que codifica una guanidinobutirasa está preferiblemente comprendida en un vector. Un vector contiene genes de resistencia bacteriana que, por ejemplo, permiten el crecimiento de bacterias en presencia de un antibiótico. Un vector más preferido es un plásmido, una molécula de ADN de doble cadena que es capaz de replicarse en bacterias independientes del ADN cromosómico.

Dicho vector, preferiblemente plásmido, preferiblemente comprende adicionalmente secuencias homólogas a una primera parte del genoma de un microorganismo, preferiblemente una parte corriente arriba de un gen que se va a alterar (denominado gen diana), y secuencias homólogas a la segunda parte del genoma de un microorganismo, preferiblemente una parte corriente abajo del gen diana, adyacente a la molécula de ácido nucleico. El término adyacente se usa para indicar que las secuencias homólogas a la primera parte del genoma están ubicadas en un lado de la molécula de ácido nucleico, mientras que las secuencias homólogas a la segunda parte del genoma están ubicadas en el otro lado de la molécula de ácido nucleico.

Las secuencias homólogas a la primera parte del genoma comprenden preferiblemente entre 20 y 1000 pb, más preferiblemente entre 30 y 500 pb, más preferidas entre 40 y 250 pb, más preferidas entre 50 a 80 pb de secuencias que son homólogas a una primera parte del genoma, preferiblemente a una parte corriente arriba del gen que se va a alterar o eliminar. Las secuencias homólogas a la parte corriente arriba de un gen incluyen preferiblemente el codón de inicio.

Las secuencias homólogas a la segunda parte del genoma comprenden preferiblemente entre 20 y 1000 pb, más preferiblemente entre 30 y 500 pb, más preferida entre 40 y 250 pb, más preferida entre 50 y 80 pb de secuencias que son homólogas a una segunda parte del genoma, preferiblemente a la parte de corriente abajo del gen que se va a alterar o eliminar. Las secuencias homólogas a la parte de corriente abajo de un gen incluyen preferiblemente el codón de parada.

Dichas secuencias que son homólogas a una primera y/o segunda parte del genoma, preferiblemente un gen diana, pueden comprender secuencias que están alteradas, cuando se comparan con las secuencias del genoma. Los términos que altera, alteración y alterado se refieren a la sustitución de uno o más nucleótidos, la inserción de uno o más nucleótidos y/o la deleción de uno o más nucleótidos en cualquier lugar dentro de las secuencias homólogas.

Una sustitución de uno o más nucleótidos se puede lograr alterando uno o más nucleótidos en la primera parte y/o la segunda parte, preferiblemente en secuencias que son homólogas a una parte corriente arriba y/o corriente abajo de un gen. Cuando la primera parte de la homología y la segunda parte de la homología cubren regiones adyacentes en el genoma, preferiblemente en un gen diana, la integración del vector de direccionamiento dará como resultado una alteración del genoma.

Dicho vector, preferiblemente plásmido, preferiblemente comprende además un sitio de reconocimiento para una endonucleasa en un extremo de la molécula de ácido nucleico que codifica una guanidinobutirasa y una copia de este sitio de reconocimiento en el otro extremo de la molécula de ácido nucleico que codifica una guanidinobutirasa. Dichos sitios de reconocimiento están localizados directamente adyacentes a la molécula de ácido nucleico que codifica una guanidinobutirasa. Dicha endonucleasa preferiblemente es una endonucleasa de corte raro tal como, por ejemplo, PacI (secuencia de reconocimiento diana 5'-TTAATTAA); AscI (secuencia de reconocimiento diana 5'-GGCGCGCC), y AsiSI (secuencia de reconocimiento diana 5'-GCGATCGC). PacI, AscI y AsiSI están disponibles en New England Biolabs. La endonucleasa más preferiblemente es una endonucleasa homing. El término endonucleasa homing se refiere a una endonucleasa que se codifica como genes independientes dentro de los intrones, como una fusión con una proteína huésped o como una inteína de auto corte y empalme. Una lista preferida de endonucleasas homing se proporciona en la tabla 1. Ejemplos adicionales de nucleasas homing son I-DirI, I-NjaI, I-NanI, I-NitI, F-TevI, F-TevII, F-CphI, PI-MgaI, I-CsmI, que son todas conocidas para los expertos. En Benjamin K (Solicitud de Patente US2012/052582) se proporcionan ejemplos adicionales de nucleasas homing.

Dicho vector, preferiblemente plásmido, preferiblemente comprende además una secuencia de ácido nucleico de entre 20 y 200 pb, preferiblemente entre 30 y 100 pb, más preferido aproximadamente 40-50 pb, que está duplicada en cualquier lado de la molécula de ácido nucleico que codifica una guanidinobutirasa, preferiblemente fuera de los sitios de reconocimiento para una endonucleasa. Dicha secuencia de ácido nucleico de entre 20 y 200 pb se localiza preferiblemente entre las secuencias homólogas a la parte corriente arriba de un gen diana y el sitio de reconocimiento de una endonucleasa en un extremo de la molécula de ácido nucleico que codifica una guanidinobutirasa, y entre las secuencias homóloga a la parte corriente abajo del gen diana y el sitio de reconocimiento de una endonucleasa en el segundo extremo de la molécula de ácido nucleico que codifica una guanidinobutirasa. La secuencia de ácido nucleico duplicada es preferiblemente idéntica a una región en el genoma diana, preferiblemente en el gen diana. La secuencia duplicada de ácido nucleico de entre 20 y 200 pb garantiza la eliminación sin problemas del marcador del genoma diana mediante recombinación homóloga.

Se describe adicionalmente un método para producir un microorganismo, preferiblemente una levadura, que comprende la molécula de ácido nucleico que codifica una guanidinobutirosa, el método que comprende proporcionar el microorganismo con la molécula de ácido nucleico que codifica una guanidinobutirasa, seleccionando un microorganismo que tiene dicha molécula de ácido nucleico. produciendo así un microorganismo que comprende la molécula de ácido nucleico que codifica una guanidinobutirasa.

Los métodos para seleccionar un microorganismo, preferiblemente una levadura, que tiene el ácido nucleico que codifica una guanidinobutirasa son conocidos en la técnica e incluyen la transferencia de Southern y la amplificación de un producto de ácido nucleico que comprende al menos una parte de la molécula de ácido nucleico que usa al menos un cebador que es específico para la molécula de ácido nucleico. Alternativamente, un vector que comprende la molécula de ácido nucleico que codifica una guanidinobutirasa preferiblemente comprende además un marcador seleccionable que permite la selección de un microorganismo que comprende la molécula de ácido nucleico que codifica una guanidinobutirasa.

El término específico, como se usa en este documento, se refiere a un cebador o polinucleótido que se hibridará solo a su subsecuencia diana, por lo general en una mezcla compleja de ácidos nucleicos, pero esencialmente a ninguna otra secuencia que aplique condiciones rigurosas como es conocido para el experto. Las condiciones estrictas dependen de la secuencia y serán diferentes en diferentes circunstancias. Una guía extensa para la hibridación de ácidos nucleicos se encuentra en (Tijssen, (1993) Hybridization with Nucleic Acid Probes, vol. 2, Laboratory techniques in biochemistry and molecular biology, Volume 24. Elsevier, Amsterdam). En general, las condiciones rigurosas se seleccionan para que sean aproximadamente 5-10 °C más bajas que el punto de fusión térmico (Tm) para la secuencia específica a un pH de fuerza iónica definido. La Tm es la temperatura (bajo fuerza iónica definida, pH y concentración de ácido nucleico) a la cual el 50% de los cebadores complementarios a la diana se hibridan con la secuencia diana en equilibrio (ya que las secuencias diana están presentes en exceso, a Tm, 50 % de las sondas están ocupadas en el equilibrio). Las condiciones más estrictas serán aquellas en las que la concentración de sal sea inferior a aproximadamente 1.0 M de ion de sodio, por lo general de aproximadamente 0.01 a 1.0 M de concentración de ion de sodio (u otras sales) a un pH de 7.0 a 8.3 y la temperatura es de al menos aproximadamente 30 °C para sondas cortas (por ejemplo, de 10 a 50 nucleótidos) y al menos aproximadamente 60 °C para sondas largas (por ejemplo, más de 50 nucleótidos). Para la hibridación selectiva o específica, una señal positiva es al menos dos veces la referencia, preferiblemente 10 veces la hibridación de referencia.

Un método preferido para alterar un genoma comprende la selección de un microorganismo, preferiblemente una levadura, que expresa funcionalmente la molécula de ácido nucleico que codifica el marcador de selección. Por ejemplo, se puede seleccionar una levadura que exprese guanidinobutirasa, como se determina en un ensayo de actividad enzimática. Los métodos para determinar la actividad de guanidinobutirasa son conocidos en la técnica. La cantidad de urea producida se puede cuantificar usando, por ejemplo, el método de Archibald (Archibald, (1945) J Biol Chem 157: 507-518). Se puede generar una línea de calibración que varía desde 0 a 1 mM de urea para la cuantificación de la cantidad de urea.

El término "expresar funcionalmente" en esta especificación se usa para indicar que un gen de interés expresa la proteína que está codificada por el gen, en una forma activa.

Dicho microorganismo se selecciona preferiblemente ya que es capaz de crecer sobre un sustrato, preferiblemente un medio sintético, que comprende, como única fuente de nitrógeno, guanidinobutirato.

La invención proporciona además un microorganismo seleccionado de los géneros *Saccharomyces sensu stricto*, *Kazachstania*, *Naumovozyma*, *Nakaseomyces* y *Vanderwaltozyma* que comprenden una secuencia de nucleótidos transformada que codifica una guanidinobutirasa codificada, en la cual la guanidinobutirasa codificada de la familia de la ureohidrolasas comprende uno o ambos dominios conservados como se indica en la figura 2A y se representa en la figura 2B. Dicha secuencia de nucleótidos transformada preferiblemente se integra en el genoma de dicho microorganismo.

Dicho microorganismo comprende preferiblemente una secuencia de nucleótidos que codifica una secuencia de nucleótidos que codifica guanidinobutirasa que codifica la proteína hipotética NRRL Y-1140 de *Kluyveromyces lactis*.

El método como se describe en este documento se puede aplicar a todos los microorganismos. Si un microorganismo tiene actividad de guanidinobutirasa endógena, se pueden proporcionar mutantes de estos organismos en los que esta actividad está inactivada. Los métodos para inactivar un gen que codifica guanidinobutirasa en un microorganismo son conocidos en la técnica. Los métodos apropiados se han descrito, por ejemplo, en Akada et al., (2002). Yeast 19: 393-402; McNabb et al., (1997). Biotechniques, 22: 1134-1139; Storici et al., (1999). Yeast 15: 271-283; Gueldener et al., (2002). Nucleic Acids Res 30: e23; e Iwaki and Takegawa, (2004). Biosci Biotechnol Biochem 68: 545-550. Cualquiera de estos métodos se puede aplicar para generar un microorganismo en los que un gen endógeno que codifica guanidinobutirasa se puede inactivar funcionalmente.

El término funcionalmente inactivado se usa en este documento para indicar una presencia funcional reducida de un producto proteico de un gen en un microorganismo, que se debe ya sea a un nivel de expresión reducido o a un nivel de actividad reducido de la proteína. Dicha presencia funcional reducida produce preferiblemente una reducción de más del 90% de la cantidad y/o actividad de la proteína, más preferiblemente una reducción de más del 95% de la cantidad y/o actividad de la proteína, la más preferida es una reducción de más del 99% de la cantidad y/o actividad de la proteína, en comparación con la actividad de la proteína correspondiente en un microorganismo relacionado que no comprende el gen funcionalmente inactivado. En una realización más preferida, un gen funcionalmente inactivado no tiene actividad residual y es equivalente a un gen inactivado. El término gen inactivado se refiere al gen que se ha hecho funcionalmente inactivo por deleción parcial o completa de la región de codificación del ADN genómico que codifica dicho gen.

Un microorganismo de los géneros *Saccharomyces sensu stricto*, *Kazachstania*, *Naumovozyma*, *Nakaseomyces* y *Vanderwaltozyma* surgió después del evento de duplicación del genoma completo (post WGD) que jugó un papel importante en la evolución de este subfilo.

Una comparación del metabolismo del nitrógeno entre *Saccharomyces cerevisiae* y *Kluyveromyces lactis* reveló diferencias sustanciales en la asimilación de arginina entre estas dos especies de levaduras. Mientras que la deleción del gen CAR1 de *S. cerevisiae* arginasa fue suficiente para anular el crecimiento de la arginina como única fuente de nitrógeno, el correspondiente mutante Klcar1Δ en *K. lactis* todavía podía crecer en la arginina. Esta diferencia fenotípica fue causada por la presencia en *K. lactis* de un gen (KLLA0F27995g/KlGBU1) que codifica una guanidinobutirasa (EC.3.5.3.7), una enzima no demostrada previamente en hongos. La presencia de esta enzima proporciona la capacidad de crecer en guanidinobutirato como única fuente de nitrógeno.

Se encontró que los ortólogos de guanidinobutirasa no están presentes en Saccharomycetaceae después de WGD. Sorprendentemente, una de las 34 secuencias anotadas del genoma de *S. cerevisiae*, disponible en la base de datos SGD (www.yeastgenome.org/) perteneciente a la cepa EC1118, albergaba un gen cuya secuencia proteica predicha compartía un 62% de identidad de secuencia con una secuencia de guanidinobutirasa de *Kluyveromyces lactis.* Una inspección adicional de la estructura genómica de la cepa EC1118 reveló que este ortólogo se encontró en una región genómica adquirida por transferencia horizontal, más probablemente desde *Torulaspora delbrueckii*. El origen de este gen se ajusta a la distribución filogenética de los ortólogos de guanidinobutirasa en los géneros pre WGD Saccharomycetaceaee (*Zygosaccharomyces, Lachancea, Torulaspora, Kluyveromyces y Eremothecium*) y en los géneros que pertenecen al grupo CTG y Dipodascacae (Kurtzman, (2003) FEMS Yeast Res 4: 233-245; Dujon, (2010) Nat Rev Genet 11: 512-524).

Una levadura preferida es una levadura de cerveza Lager de laboratorio, silvestre e industrial, *Saccharomyces pastorianus*, o una cepa de *S. cerevisiae* de laboratorio, silvestre e industrial. Una levadura preferida de *Saccharomyces sensu stricto* (*Saccharomyces paradoxus, S. mikatae, S. bayanus, S. eubayanus, S. kudriavzevii, S. paradoxus, S. arboricolus*), o una cepa de *Kazachstania*, *Naumovozyma*, *Nakaseomyces* o *Vanderwaltozyma*, es una levadura de cerveza de laboratorio, silvestre e industrial, *Saccharomyces pastorianus* o una cepa de *S. cerevisiae*.

La invención proporciona adicionalmente un método para alterar un genoma, preferiblemente un genoma diana, en un microorganismo seleccionado de los géneros *Saccharomyces sensu stricto*, *Kazachstania*, *Naumovozyma*, *Nakaseomyces* y *Vanderwaltozyma*, que comprende proporcionar un conjunto de construcciones de la invención que codifican guanidinobutirasa, más preferiblemente que codifican la proteína hipotética NRRL Y-1140 de *Kluyveromyces lactis*, para dicho microorganismo, y seleccionando un microorganismo en el cual el genoma ha sido alterado por la selección de un microorganismo que exprese funcionalmente dicha guanidinobutirasa. Un método preferido comprende seleccionar un microorganismo cultivando en presencia de guanidinobutirato como única fuente de nitrógeno.

La presente invención proporciona además un conjunto de construcciones, que comprende una primera construcción que comprende una primera parte de la secuencia de nucleótidos que codifica una guanidinobutirasa, y una segunda construcción que comprende una segunda parte de la secuencia de nucleótidos que codifica la guanidinobutirasa, en la que un fragmento de la primera parte de la secuencia de nucleótidos se superpone con un fragmento que está presente en la segunda parte de la secuencia de nucleótidos, lo que permite la recombinación entre la primera y la segunda parte de la secuencia de nucleótidos para expresar funcionalmente el marcador de selección de guanidinobutirasa, en el que la guanidinobutirasa codificada de la familia ureohidrolasa comprende uno o ambos dominios conservados como se indica en la figura 2A y se representa en la figura 2B.

Este conjunto de construcciones supera una eficiencia de direccionamiento baja al proporcionar un conjunto de construcciones de direccionamiento, en las que la expresión correcta de un marcador de selección depende de un evento de recombinación entre las construcciones de direccionamiento. Se encontró que la ocurrencia de un evento de recombinación entre las construcciones de direccionamiento se mejora notablemente después de la integración de las construcciones de direccionamiento en el locus de direccionamiento correcto. Por lo tanto, el sistema diana de la presente invención, que comprende un conjunto de construcciones de direccionamiento, aumenta considerablemente el porcentaje de construcciones correctamente integradas en microorganismos que expresan el marcador de selección, en comparación con un sistema de selección de un solo vector. La división del marcador en dos construcciones separadas limita la ocurrencia de falsos positivos debido a eventos de cruce único. El enfoque del marcador dividido mejora la proporción de positivos verdaderos sobre falsos positivos (Nielsen et al., 2006. Fungal Gen Biol 43: 54-64).

El término construcción o construcción de direccionamiento, como se usa en este documento, se refiere a un segmento de ácido nucleico construido artificialmente. Una construcción preferida es un vector, preferiblemente un vector que contiene genes de resistencia bacteriana para el crecimiento en bacterias. Una construcción más preferida es un plásmido, un ADN de doble cadena lineal o circular que es capaz de replicarse en bacterias independientemente del ADN cromosómico.

El término superposición, como se usa en este documento, se refiere a una región duplicada de la secuencia de nucleótidos que codifica un marcador de selección que está presente en ambos conjuntos de construcciones. La región duplicada es sustancialmente idéntica y preferiblemente está entre 40 y 400 pb, preferiblemente de aproximadamente 200 pb. El término sustancialmente, como se usa en este documento, se usa para indicar que la región es al menos un 90% idéntica para asegurar una recombinación eficiente entre las construcciones de direccionamiento, más preferida al menos 95% idéntica, más preferida al menos 99% idéntica, más preferida 100 % idéntica.

La primera construcción en el conjunto de construcciones de direccionamiento preferiblemente comprende además un sitio de reconocimiento para una endonucleasa y una primera región de homología con un genoma diana de un microorganismo, y la segunda construcción comprende además una segunda región de homología con el genoma diana del microorganismo, y una copia del sitio de reconocimiento de endonucleasas, por lo que una secuencia codificante que codifica la endonucleasa y que está acoplada a un promotor inducible está presente en la primera o segunda construcción; y una parte de la primera región de homología con el genoma diana en la primera construcción se duplica entre la copia del sitio de reconocimiento de endonucleasas y la segunda región de homología con el genoma diana en la segunda construcción; o una parte de la segunda región de homología con el genoma diana en la segunda construcción se duplica entre la primera región de homología con el genoma diana y el sitio de reconocimiento de endonucleasa en la primera construcción.

Dicha región duplicada de homología con el genoma diana en la primera y segunda construcción de direccionamiento está preferiblemente entre 20 y 200 pb, preferiblemente entre 40 y 100 pb, preferiblemente aproximadamente 80 pb. Dicha región duplicada de homología con el genoma diana en la primera y la segunda construcción de direccionamiento permite la eliminación sin cicatrización del marcador del genoma diana mediante recombinación homóloga.

La primera construcción comprende preferiblemente, en este orden, una primera región de homología con un genoma diana de un microorganismo, un sitio de reconocimiento para una endonucleasa y una primera parte de un marcador de selección. La segunda construcción comprende preferiblemente, en este orden, una región de solapamiento con la primera parte de la secuencia de nucleótidos que codifica un marcador de selección seguido de una segunda parte de la secuencia de nucleótidos que codifica el marcador de selección, una secuencia codificante que codifica la endonucleasa y que se acopla a un promotor inducible, una copia del sitio de reconocimiento de endonucleasas, una copia de una parte de la primera región de homología con el genoma diana que está presente en la primera construcción, y una segunda región de homología con el genoma diana del microorganismo. Esta configuración se representa en la figura 6

El genoma diana puede ser cualquier ubicación, preferiblemente un gen, en el genoma de un microorganismo, preferiblemente de una levadura, preferiblemente de una levadura de los géneros *Saccharomyces sensu stricto*, *Kazachstania*, *Naumovozyma*, *Nakaseomyces* y *Vanderwaltozyma*, de los cuales la secuencia genómica debe ser alterada. El término gen, como se usa en este documento, se refiere a una parte del genoma del microorganismo que comprende partes intrónicas y exónicas de un gen, la región promotora de dicho gen y secuencias genómicas que median la expresión de dicho gen, tal como por ejemplo secuencias potenciadoras.

El experto entenderá que las construcciones de direccionamiento se usan preferiblemente para alterar un gen de un microorganismo. Por lo tanto, la invención proporciona además un conjunto de construcciones de direccionamiento, que comprenden una primera construcción que comprende una primera región de homología con un gen diana de un microorganismo, un sitio de reconocimiento para una endonucleasa y una primera parte de una secuencia de nucleótidos que codifica un marcador de selección, y una segunda construcción que comprende una región de solapamiento con la primera parte de la secuencia de nucleótidos que codifica el marcador de selección seguido de una secuencia de nucleótidos que codifica una segunda parte del marcador de selección, una copia del sitio de reconocimiento de endonucleasas y una segunda región de homología con la gen diana del microorganismo, mediante el cual los fragmentos superpuestos permiten la recombinación entre la primera y la segunda parte de la secuencia de nucleótidos que codifica el marcador de selección; por lo que una secuencia codificante que codifica la endonucleasa y que está acoplada a un promotor inducible está presente en la primera o segunda construcción; y mediante el cual una

parte de la primera región de homología con el gen diana en la primera construcción se duplica entre la copia del sitio de reconocimiento de endonucleasa y la segunda región de homología con el gen diana en la segunda construcción; o una parte de la segunda región de homología con el gen diana en la segunda construcción se duplica entre la primera región de homología con el gen diana y el sitio de reconocimiento de endonucleasa en la primera construcción.

Dicha región duplicada de homología con el gen diana en la primera y segunda construcción de direccionamiento está preferiblemente entre 20 y 200 pb, preferiblemente entre 40 y 100 pb, preferiblemente aproximadamente 80 pb.

El término alteración de la secuencia genómica incluye la sustitución de uno o más nucleótidos, la inserción de uno o más nucleótidos y/o la deleción de uno o más nucleótidos en cualquier lugar dentro de un genoma, preferiblemente dentro de un gen.

Por ejemplo, si la primera y la segunda región de homología con un gen diana comprenden secuencias genómicas adyacentes del gen, una sustitución de uno o más nucleótidos en la primera región de homología y/o en la segunda región de homología, dará como resultado una alteración del gen después del direccionamiento homólogo con el conjunto de construcciones de direccionamiento según la invención. Dicha sustitución de uno o más nucleótidos está preferiblemente en la región de homología con el gen diana que está presente en la primera y en la segunda construcción.

Dicha alteración de la secuencia genómica es preferiblemente una deleción de uno o más nucleótidos, preferiblemente en cualquier lugar dentro del gen. Por ejemplo, si la primera y la segunda región de homología con un gen diana comprenden secuencias genómicas del gen que están separadas en el genoma del organismo, una alteración del gen después del direccionamiento homólogo con el conjunto de construcciones de direccionamiento según la invención el resultado es una deleción de la región que estaba ubicada entre la primera y la segunda región de homología en el cromosoma parental.

Dicha primera construcción comprende preferiblemente una primera parte, preferiblemente la primera mitad o los primeros dos tercios, de una región que codifica el marcador de selección. Por ejemplo, la proteína guanidinobutirasa de *K. lactis* tiene 410 aminoácidos, que está codificada por una secuencia de ácido nucleico de 1230 pb. Dicha primera construcción comprende preferiblemente entre 400 y 800 pb de la región codificante de esta proteína, más preferiblemente entre 500 y 700 pb. La segunda construcción comprende preferiblemente entre 400 y 800 pb de la región codificante de esta proteína, más preferiblemente entre 500 y 700 pb.

La región de solapamiento entre la primera y la segunda parte del marcador de selección está preferiblemente entre aproximadamente 50 pb y aproximadamente 600 pb, preferiblemente aproximadamente 200 pb.

La primera o segunda construcción de direccionamiento comprende una secuencia codificante que codifica una endonucleasa y que está acoplada a un promotor inducible. La endonucleasa es preferiblemente una endonucleasa de corte raro tal como, por ejemplo, PacI (secuencia de reconocimiento diana 5'-TTAATTAA); AscI (secuencia de reconocimiento diana 5'-GGCGCGCC), y AsiSI (secuencia de reconocimiento diana 5'-GCGATCGC). PacI, AscI y AsiSI están disponibles en New England Biolabs. La endonucleasa más preferiblemente es una endonucleasa homing. El término endonucleasa homing se refiere a una endonucleasa que se codifica ya sea como genes independientes dentro de los intrones, como una fusión con una proteína huésped o como una inteína de auto corte y empalme. Una lista preferida de endonucleasas homing se proporciona en la tabla 1. Ejemplos adicionales de nucleasas homing son I-DirI, I-NjaI, I-NanI, I-NitI, F-TevI, F-TevII, F-CphI, PI-MgaI, I-CsmI, que son todas conocidas para los expertos. Otros ejemplos de nucleasas homing se proporcionan en la Solicitud de Patente US 2012/052582.

Una nucleasa homing preferida es PI-PspI (New England Biolabs; secuencia de reconocimiento 5'-TGGCAAACAGCTATTATGGGTATTATGGGT)) o PI-SceI (New England Biolabs; secuencia de reconocimiento 5'-ATCTATGTCGGGTGCGGAGAAAGAGGTAAT). Las secuencias codificantes de la mayoría de las endonucleasas homing son conocidas. Por ejemplo, la secuencia codificante de PI-SceI y de PI-PspI está disponible en las bases de datos públicas (número de acceso de GenBank Z74233.1 y número de acceso de Genbank U00707.1, respectivamente). El experto entenderá que una secuencia que difiere de la secuencia disponible públicamente para una nucleasa, todavía puede codificar la nucleasa. Por ejemplo, el término región codificante de PI-PspI puede incluir una secuencia que se desvía de la secuencia disponible públicamente, por ejemplo, por optimización de codones, pero que todavía expresa una endonucleasa activa que reconoce y digiere la secuencia de reconocimiento diana indicada.

Dicha endonucleasa está preferiblemente bajo el control de un promotor inducible. El término promotor inducible, como se usa en este documento, se refiere a un promotor cuya expresión puede ser regulada. Los promotores inducibles son conocidos para los expertos. Ejemplos de promotores inducibles que se han empleado en levaduras son el promotor GAL1 y el promotor GAL10, ambos inducibles por galactosa, el promotor SUC2, que es inducible por sacarosa, el promotor MAL12, que es inducible por maltosa; el promotor CUP1, que es inducible por cobre, y los promotores tetO7 y tetO2, que son inducibles por tetraciclina [Gari et al., (1997) Yeast 13: 837-48; Yen et al., (2003) Yeast 20 1255-62]. Un promotor inducible preferido es el promotor GAL1.

Un sitio de reconocimiento que comprende la secuencia de reconocimiento diana para la endonucleasa, está ubicado adyacente a (detrás) la primera región de homología con un gen diana de un microorganismo en la primera

construcción. Una copia de este sitio de reconocimiento se encuentra adyacente a (frente a) la segunda región de homología con el gen diana del microorganismo en la segunda construcción. El experto entenderá que cuando una parte de la primera región de homología con el gen diana en la primera construcción se duplica entre la copia del sitio de reconocimiento de endonucleasa y la segunda región de homología con el gen diana en la segunda construcción, dicha copia del sitio de reconocimiento se encuentra adyacente a (frente a) la duplicación de la primera región de homología con el gen diana en la segunda construcción. Alternativamente, el sitio de reconocimiento está ubicado adyacente a (detrás) la parte duplicada de la segunda región de homología con el gen diana en la primera construcción cuando una parte de la segunda región de homología con el gen diana en la segunda construcción se duplica en la primera construcción. El marcador de selección, incluidas las secuencias del promotor y de terminación, y la región codificante de la endonucleasa, incluido el promotor inducible, se encuentran entre el sitio de reconocimiento en la primera construcción y la copia de este sitio de reconocimiento en la segunda construcción.

La invención proporciona además un método para alterar un genoma, preferiblemente un gen diana, en un microorganismo seleccionado de los géneros *Saccharomyces sensu stricto*, *Kazachstania*, *Naumovozyma*, *Nakaseomyces* y *Vanderwaltozyma*, que comprende proporcionar el conjunto de construcciones de direccionamiento según la invención a dicho microorganismo, y selección de un microorganismo en el cual el genoma ha sido alterado por selección de un microorganismo que exprese funcionalmente dicha guanidinobutirasa.

Como se indica en este documento anteriormente, la aparición de un evento de recombinación entre las construcciones de direccionamiento se mejora notablemente después de la integración de las construcciones de direccionamiento en el locus de direccionamiento correcto. Por lo tanto, la presencia de un marcador de selección funcionalmente recombinante es altamente indicativa de la presencia de construcciones de direccionamiento correctamente integradas en el genoma diana y, por lo tanto, de un genoma alterado en el microorganismo.

Como se indica en este documento anteriormente, los términos que altera, alteración y alterado se refieren a la sustitución de uno o más nucleótidos, la inserción de uno o más nucleótidos y/o la deleción de uno o más nucleótidos en cualquier parte del genoma, preferiblemente dentro de un gen diana.

Se puede realizar una sustitución de uno o más nucleótidos alterando uno o más nucleótidos en la primera región de homología y/o en la segunda región de homología. Cuando la primera región de homología y la segunda región de homología con el genoma diana cubren regiones adyacentes del genoma, preferiblemente el gen diana, la integración de los vectores de direccionamiento dará como resultado una alteración del genoma. Cuando está presente, dicha sustitución de uno o más nucleótidos se realiza preferiblemente mediante la alteración de uno o más nucleótidos en la región de superposición de la homología con el genoma que está presente en la primera y en la segunda construcción.

Dicha alteración de una secuencia genómica es preferiblemente una deleción de uno o más nucleótidos en cualquier lugar dentro de un genoma, preferiblemente dentro de un gen. Por ejemplo, si la primera y segunda región de homología con un genoma diana comprenden secuencias genómicas que se separan en el genoma del organismo, una alteración del genoma después del direccionamiento homólogo con el conjunto de construcciones de direccionamiento según la invención resultará en una deleción de la región que estaba ubicada entre la primera y la segunda región de homología en el cromosoma parental.

Se describe adicionalmente un método para producir un microorganismo que comprende un genoma alterado, preferiblemente un gen alterado, el método que comprende proporcionar el conjunto de construcciones de direccionamiento según la invención a dicho microorganismo, y seleccionar un microorganismo en el que el genoma ha sido alterado y que expresa funcionalmente un marcador de selección recombinada.

El método para producir un microorganismo que comprende un genoma alterado comprende preferiblemente inducir el promotor inducible para la expresión de la endonucleasa, eliminando así el marcador de selección y la región codificante de la endonucleasa, incluido el promotor inducible, del genoma diana.

Se describe adicionalmente un microorganismo, que comprende una alteración genómica que se produce mediante los métodos de la invención. Cuando están presentes, las regiones duplicadas de homología con el genoma diana en la primera y segunda construcción de direccionamiento aseguran la eliminación sin problemas del marcador del genoma diana mediante recombinación homóloga. El microorganismo resultante comprende solo la alteración o las alteraciones que estaban presentes en la primera y/o segunda construcción de direccionamiento, o que fueron inducidas por recombinación de las construcciones de direccionamiento en el genoma de direccionamiento, tal como una inserción en el genoma de direccionamiento o una deleción del genoma de direccionamiento.

Se describe adicionalmente un microorganismo que comprende una alteración genómica, preferiblemente una alteración de un gen diana, la alteración que comprende una inserción de un marcador de selección recombinante funcionalmente y una secuencia codificante para una endonucleasa que está acoplada a un promotor inducible, por lo que el genoma diana comprende una copia de una secuencia de reconocimiento para la endonucleasa en ambos sitios de la inserción.

La invención proporciona además un kit que comprende el conjunto de construcciones de la invención. Dicho kit puede comprender además métodos y medios para el crecimiento de un microorganismo, preferiblemente una levadura,

preferiblemente de la Saccharomycetaceae, en medio sintético que comprende guanidinobutirato, y/o para la identificación de la enzima guanidinobutirasa tal como, por ejemplo, ácido guanidinobutírico.

Se describe adicionalmente un método de cultivo de un microorganismo, preferiblemente una levadura, preferiblemente de Saccharomycetaceae, en presencia de guanidinobutirato o agmatina como única fuente de nitrógeno, que comprende: (a) introducir la molécula de ácido nucleico de la invención en el microorganismo y (b) cultivar el microorganismo de modo que la molécula de nucleótido se exprese en el microorganismo.

Se describe adicionalmente un método de cultivo de un microorganismo, preferiblemente una levadura, preferiblemente de Saccharomycetaceae, en presencia de guanidinobutirato o agmatina como única fuente de nitrógeno, que comprende: (a) introducir el conjunto de construcciones de la invención en el microorganismo y (b) cultivar el microorganismo de manera que después de la recombinación de los construcciones de direccionamiento, la molécula de nucleótido que codifica el marcador de selección se exprese en el microorganismo.

Se describe adicionalmente un método para producir un microorganismo que comprende un genoma alterado, el método que comprende proporcionar un microorganismo que comprende una alteración del genoma, preferiblemente de un gen diana, la alteración que comprende una inserción de una secuencia de nucleótidos funcionalmente recombinada que codifica un marcador de selección y una secuencia codificante para una endonucleasa que está acoplada a un promotor inducible, por lo que el genoma diana comprende una copia de una secuencia de reconocimiento para la endonucleasa en ambos sitios de la inserción, e inducir al promotor inducible a eliminar las secuencias de ácido nucleico entre las secuencias de reconocimiento de la endonucleasa. Nuevamente, cuando están presentes, las regiones duplicadas de homología con el gen diana en la primera y la segunda construcción de direccionamiento aseguran la eliminación perfecta del marcador del genoma diana mediante recombinación homóloga al proporcionar al ADN genómico con una pequeña pieza homóloga para volver a conectar las cadenas de ADN rotas eficientemente. El microorganismo resultante comprende solo la alteración o las alteraciones que estaban presentes en la primera y/o segunda construcción de direccionamiento, o que fueron inducidas por recombinación de las construcciones de direccionamiento en el genoma de direccionamiento, tal como una inserción en el genoma de direccionamiento o una deleción del genoma, preferiblemente una inserción en un gen dirigido o una deleción del gen dirigido o una deleción desde dentro del gen dirigido

Tabla 1

Enzima	Secuencia de reconocimiento	Corte	SF	Fuente	D	SCL
I-AniI	5' TTGAGGAGGTTTCTCTGTAAATAA 3' 3' AACTCCTCCAAAGAGACATTTATT 5'	5' ---TTGAGGAGGTTTC TCTGTAAATAA--- 3' 3' ---AACTCCTCC AAAGAGACATTTATT--- 5'	HI	Aspergillus nidulans	E	mito
I-CeuI	5' TAACTATAACGGTCCTAAGGTAGCGA 3' 3' ATTGATATTGCCAGGATTCCATCGCT 5'	5' ---TAACTATAACGGTCCTAA GGTAGCGA--- 3 3' ---ATTGATATTGCCAG GATTCCATCGCT--- 5'	HI	Chlamydomonas eugametos	E	cloro
I-ChuI	5' GAAGGTTTGGCACCTCGATGTCGGCTCATC 3' 3' CTTCCAAACCGTGGAGCTACAGCCGAGTAG 5'	5' ---GAAGGTTTGGCACCTCG ATGTCGGCTCATC--- 3' 3' ---CTTCCAAACCGTG GAGCTACAGCCGAGTAG--- 5'	HI	Chlamydomonas humicola	E	cloro
I-CpaI	5' CGATCCTAAGGTAGCGAAATTCA 3' 3' GCTAGGATTCCATCGCTTTAAGT 5'	5' ---CGATCCTAAGGTAGCGAA ATTCA--- 3 3' ---GCTAGGATTCCATC GCTTTAAGT--- 5'	HI	Chlamydomonas pallidostigmata	E	cloro
I-CpaII	5' CCCGGCTAACTCTGTGCCAG 3' 3' GGGCCGATTGAGACACGGTC 5'	5' ---CCCGGCTAACTC TGTGCCAG--- 3' 5' ---GGGCCGAT TGAGACACGGTC--- 3'	HI	Chlamydomonas pallidostigmata	E	cloro
I-CreI	5' CTGGGTTCAAAACGTCGTGAGACAGTTTGG 3' 3' GACCCAAGTTTTGCAGCACTCTGTCAAACC 5'	5' ---CTGGGTTCAAAACGTCGTGA GACAGTTTGG--- 3 3' ---GACCCAAGTTTTGCAG CACTCTGTCAAACC--- 5'	HI	*Chlamydomonas reinhardtii*	E	cloro
DmoI	5' ATGCCTTGCCGGGTAAGTTCCGGCGCGCAT 3' 3' TACGGAACGGCCCATTCAAGGCCGCGCGTA 5'	5' ---ATGCCTTGCCGGGTAA GTTCCGGCGCGCAT--- 3' 3' ---TACGGAACGGCC CATTCAAGGCCGCGCGTA--- 5'		*Desulfurococcus mobilis*	A	Chrm
H-DreI[	5' CAAAACGTCGTAAGTTCCGGCGCG 3' 3' GTTTTGCAGCATTCAAGGCCGCGC 5'	5' ---CAAAACGTCGTAA GTTCCGGCGCG--- 3' 3' ---GTTTTGCAG CATTCAAGGCCGCGC--- 5'	HI	Escherichia coli	B	
I-HmuI	5' AGTAATGAGCCTAACGCTCAGCAA 3' 3' TCATTACTCGGATTGCGAGTCGTT 5'	• * 3' ---TCATTACTCGGATTGC GAGTCGTT--- 5'	HIII	*Bacillus subtilis*SPO1	B	fago
I-HmuII	5' AGTAATGAGCCTAACGCTCAACAA 3' 3'TCATTACTCGGATTGCGAGTTGTT 5'	3' ---TCATTACTCGGATTGCGAGTTGTTN$_{35}$ NNNN-5'	HIII	*Bacillus subtilis* phage SP82	B	fago
I-LlaI	5' CACATCCATAACCATATCATTTTT 3' 3' GTGTAGGTATTGGTATAGTAAAAA 5'	5' ---CACATCCATAA CCATATCATTTTT--- 3' 3' ---GTGTAGGTATTGGTATAGTAA AAA--- 5'	HIII	*Lactococcus lactis*	B	chrm
I-MsoI	5' CTGGGTTCAAAACGTCGTGAGACAGTTTGG 3' 3' GACCCAAGTTTTGCAGCACTCTGTCAAACC 5'	5' ---CTGGGTTCAAAACGTCGTGA GACAGTTTGG--- 3' 3' ---GACCCAAGTTTTGCAG CACTCTGTCAAACC--- 5'		*Monomastix* sp	E	
PI-PfuI	5' GAAGATGGGAGGAGGGACCGGACTCAACTT 3' 3' CTTCTACCCTCCTCCCTGGCCTGAGTTGAA 5'	5' ---GAAGATGGGAGGAGGG ACCGGACTCAACTT--- 3' 3' ---CTTCTACCCTCC TCCCTGGCCTGAGTTGAA--- 5'		Pyrococcus furiosus Vc1	A	
PI-PkoII	5' CAGTACTACGGTTAC3' 3' GTCATGATGCCAATG5'	5' ---CAGTACTACG GTTAC--- 3' 3' ---GTCACG ATGCCAATG--- 5'		*Pyrococcus kodakaraensis* KOD1	A	

(continuación)

Enzima	Secuencia de reconocimiento	Corte	SF	Fuente	D	SCL
I-PorI	5' GCGAGCCCGTAAGGGTGTGTACGGG 3' CGCTCGGGCATTCCCACACATGCCC	5' ---GCGAGCCCGTAAGGGT GTGTACGGG--- 3' 3' ---CGCTCGGGCATT CCCACACATGCCC--- 5'	HIII	Pyrobaculum organotrophum	A	chrm
I-PpoI	5' TAACTATGACTCTCTTAAGGTAGCCAAAT 3' ATTGATACTGAGAGAATTCCATCGGTTTA	5' ---TAACTATGACTCTCTTAA GGTAGCCAAAT--- 3' 3' ---ATTGATACTGAGAG AATTCCATCGGTTTA--- 5'	HIV	*Physarum polycephalum*	E	nuclear
PI-PspI	5' TGGCAAACAGCTATTATGGGTATTATGGGT 3' ACCGTTTGTCGATAATACCCATAATACCCA	5' ---TGGCAAACAGCTATTAT GGGTATTATGGGT--- 3' 3' ---ACCGTTTGTCGAT AATACCCATAATACCCA--- 5'	HI	*Pyrococcus* sp	A	chrm
I-SeaI	5' TGTCACATTGAGGTGCACTAGTTATTAC 3' ACAGTGTAACTCCACGTGATCAATAATG	5' ---TGTCACATTGAGGTGCACT AGTTATTAC--- 3' 3' ---ACAGTGTAACTCCAC GTGATCAATAATG--- 5'	HI	Saccharomyces capensis	E	mito
I-SceI	5' AGTTACGCTAGGGATAACAGGGTAATATAG 3' TCAATGCGATCCCTATTGTCCCATTATATC	5' ---AGTTACGCTAGGGATAA CAGGGTAATATAG--- 3' 3' ---TCAATGCGATCCC TATTGTCCCATTATATC--- 5'	HI	*Saccharomyces cerevisiae*	E	mito
PI-SceI	5' ATCTATGTCGGGTGCGGAGAAAGAGGTAATGAAATGGCA 3' TAGATACAGCCCACGCCTCTTTCTCCATTACTTTACCGT	5'-ATCTATGTCGGGTGC GGAGAAAGAGGTAAT-GAAATGGCA-3' 3'-TAGATACAGCC CACGCCTCTTTCTCCATTACTT-TACCGT-5'	HI	*Saccharomyces cerevisiae*	E	
I-SceII	5' TTTTGATTCTTTGGTCACCCTGAAGTATA 3' AAAACTAAGAAACCAGTGGGACTTCATAT	5' ---TTTTGATTCTTTGGTCACCC TGAAGTATA--- 3' 3' ---AAAACTAAGAAACCAG TGGGACTTCATAT--- 5'	HI	*Saccharomyces cerevisiae*	E	mito
I-SecIII	5' ATTGGAGGTTTTGGTAACTATTTATTACC 3' TAACCTCCAAAACCATTGATAAATAATGG	5' ---ATTGGAGGTTTTGGTAAC TATTTATTACC--- 3' 3' ---TAACCTCCAAAACC ATTGATAAATAATGG--- 5'	HI	*Saccharomyces cerevisiae*	E	mito
I-SceIV	5' TCTTTTCTCTTGATTAGCCCTAATCTACG 3' AGAAAAGAGAACTAATCGGGATTAGATGC	5' ---TCTTTTCTCTTGATTA GCCCTAATCTACG--- 3' 3' ---AGAAAAGAGAAC TAATCGGGATTAGATGC--- 5'	HI	*Saccharomyces cerevisiae*	E	mito
I-SceV	5' AATAATTTTCTTCTTAGTAATGCC 3' TTATTAAAAGAAGAATCATTACGG	5' ---AATAATTTTCT TCTTAGTAATGCC--- 3' 3' ---TTATTAAAAGAAGAATCATTA CGG--- 5'	HIII	*Saccharomyces cerevisiae*	E	mito
I-SeeVI	5' GTTATTTAATGTTTTAGTAGTTGG 3' CAATAAATTACAAAATCATCAACC	5' ---GTTATTTAATG TTTTAGTAGTTGG--- 3' 3' ---CAATAAATTACAAAATCATCA ACC--- 5'	HIII	*Saccharomyces cerevisiae*	E	mito
I-SceVII	5' TGTCACATTGAGGTGCACTAGTTATTAC	Desconocido	HI	*Saccharomyces*	E	mito
	3' ACAGTGTAACTCCACGTGATCAATAATG			*cerevisiae*		
I-Ssp6803I	5' GTCGGGCTCATAACCCGAA 3' CAGCCCGAGTATTGGGCTT	5' ---GTCGGGCT CATAACCCGAA--- 3' 3' ---CAGCCCGAGTA TTGGGCTT--- 5'		*Synechocystis* sp. PCC 6803	B	

(continuación)

Enzima	Secuencia de reconocimiento	Corte	SF	Fuente	D	SCL
I-TevI	5' AGTGGTATCAACGCTCAGTAGATG 3' TCACCATAGT TGCGAGTCATCTAC	5' ---AGTGGTATCAAC GCTCAGTAGATG--- 3' 3' ---TCACCATAGT TGCGAGTCATCTAC--- 5'	HII	*Escherichia coli* phage T4	B	fago
I-TevII	5' GCTTATGAGTATGAAGTGAACACGTTATTC 3' CGAATACTCATACTTCACTTGTGCAATAAG	5' ---GCTTATGAGTATGAAGTGAACACGT TATTC--- 3' 3' ---CGAATACTCATACTTCACTTGTG CAATAAG--- 5'	HII	*Escherichia coli* phase T4	B	fago
I-TevIII	5' TATGTATCTTTTGCGTGTACCTTTAACTTC 3' ATACATAGAAAACGCACATGGAAATTGAAG	5' ---T ATGTATCTTTTGCGTGTACCTTTAACTTC--- 3' 3' ---AT ACATAGAAAACGCACATGGAAATTGAAG--- 5'	HIII	*Escherichia coli* phase RB3	B	fago
PI-TliI	5' TAYGCNGAYACNGACGGYTTYT 3' ATRCGNCTRTGNCTGCCTAARA	5' ---TAYGCNGAYACNGACGG YTTYT--- 3' 3' ---ATRCGNCTRTGNC TGCCTAARA--- 5'	HI	*Thermococcus litoralis*	A	chrm
PI-TliII	5' AAATTGCTTGCAAACAGCTATTACGGCTAT 3' TTTAACGAACGTTTGTCGATAATGCCGATA	Desconocido **	HI	*Thermococcus litoralis*	A	chrm
I-Tsp061I	5' CTTCAGTATGCCCCGAAAC 3' GAAGTCATACGGGGCTTTG	5' ---CTTCAGTAT GCCCCGAAAC--- 3' 3' ---GAAGT CATACGGGGCTTTG--- 5'		*Thermoproteus* sp. IC-061	A	
I-Vdi141I	5' CCTGACTCTCTTAAGGTAGCCAAA 3' GGACTGAGAGAATTCCATCGGTTT	5' ---CCTGACTCTCTTAA GGTAGCCAAA--- 3' 3' ---GGACTGAG AGAATTCCATCGGTTT--- 5'		Vulcanisaeta distributa IC-141	A	

Tabla 1: descripción general de las endonucleasas homing y sus secuencias diana.

Abreviaturas: SF Familia estructural: HI: Familia LAGLIDADG; HII: familia GIY-YIG; HIII: familia H-N-H; VIH: familia caja de his-cys.

D: Dominio biológico de la fuente: A: archaea; B: bacterias; E: eukarya.

SCL: Ubicación subcelular: cloro: cloroplasto; chrm: cromosomica; mito: mitocondrial; nuclear: extracromosómico nuclear; fago: bacteriófago.

Tabla 2: cepas de *Saccharomyces cerevisiae* usadas en este estudio.

Cepa	Genotipo	Referencia
CEN.PK113-7D	Cepa de referencia prototrófica *MATa*	(Nijkamp et al., 2012)
CEN.PK113-5D	*MATa ura3-52*	(Entian & Kötter, 2007)
IMZ312	*MATa ura3-52* pAG426GPD-ccdB (*TDH3*$_{pr}$*-CYC1*$_{ter}$ *URA3* 2μ)	Este estudio
IME215	*MATa ura3-52* pUDE264 (*TDH3*$_{pr}$*-KlGBU1-CYC1*$_{ter}$ *URA3* 2μ)	Este estudio
IMX598	*MAT a ade2Δ::TDH3*$_{pr}$*- KlGBU1-CYC1*$_{ter}$	Este estudio

Referencias: Nijkamp et al., (2012) Microb Cell Fact 11: 36

Entian & Kötter, (2007) Meth Microbiol 36: 629-666

Tabla 3: cebadores

Cebador	Secuencia 5' a 3'
pUG-RV	GGGAGATCTCCGCCAAGCGAATTGAAGGACCGTGCGTAGAATGAAGAACA TTAAGGGTTGTCGACCTGC
pUg-FW	GGGTCTAGAATGACAAGAGGGTCGAACTCGCCTAAGTCGTAATTGAGTCC AGATCCACTAGTGGCCTATG
pDS-RV	GGGTCTAGACCGCCAAGCGAATTGAAGGACCGTGCGTAGAATGAAGAACCTATATTACCC TGTTATCCCTAGCGTAACTTTAAGGGTTCTCGAGAGCTC
pDS-FW	GGGAGATCTATGACAAGAGGGTCGAACTCGCCTAAGTCGTAATTGAGTCAGTTACGCTAG GGATAACAGGGTAATATAGCTGTTTAGCTTGCCTCGTCC
FK140 pUDI065 fw	GTATCACGAGGCCCTTTC
SLT1_control_rv	CAATTCAACGCGTCTGTGAG
KANMX4 fw	TCTTTCCTGCGTTATCCC
FK105-MP1	CTCGGTGAGTTTTCTCCTTCAT
Cebador hacia adelante GBU1	CATCCGAACATAAACAACC**ATG**AAGGTTGCAGGATTTATATTG
Cebador reverso GBU1	CAAGAATCTTTTTATTGTCAGTACTGATCAGGCTTGCAAAACAAATTGTTC
Hacia adelante esqueleto	CAATATAAATCCTGCAACCTTCATGGTTGTTTATGTTCGGATG
Reverso esqueleto	GAACAATTTGTTTTGCAAGCCTGATCAGTACTGACAATAAAAAGATTCTTG
GBU1-ADE2ko-CENPK-fw	GTAAAATCGTTGGATCTCTCTTCTAAGTACATCCTACTATAACAATCAAGAAAAACAAGA AAATCGGACAAAACAATCAAGTATGCGCTGCAGGTCGACAACCCTTAATG
GBU1-ADE2ko-CENPK-rv	GATGTAATCATAACAAAGCCTAAAAAATAGGTATATCATTTTATAATTATTTGCTGTACA AGTATATCAATAAACTTATATATTAGCCGCATAGGCCACTAGTGGATCTG

Tabla 4: Plásmidos usados en este estudio. [a] DR Repetición directa.

Plásmido	Característica	Referencia
pAG426GPDccdB	2μ ori *URA3 TDH3*$_{pr}$-*ccdB-CYC1*$_{ter}$	(Alberti et al., 2007)
pUDE264	2μ ori *URA3 TDH3*$_{pr}$-*KIGBU1-CYC1*$_{ter}$	Este estudio
pDS1	ori *AmpR* DR[a]-sitio I-SceI-*AgTEF2*$_{pr}$-*KanMX-AgTEF2*$_{ter}$-sitio I-SceI-DR	
pDS8	ori *AmpR* DR[a]-sitio I-SceI-*AgTEF2*$_{pr}$-*KIGBU1-AgTEF2*$_{ter}$-sitio I-SceI-DR	Este estudio

Referencias: Alberti et al., (2007) Yeast 24: 913-919

Tabla 5: Actividades de guanidinobutirasa medidas en extractos celulares de las cepas IME215 y CEN.PK113-7D de *S. cerevisiae* cultivadas en cultivos en lotes con arginina como única fuente de nitrógeno. Las cepas de *S. cerevisiae* e IME215 se cultivaron previamente en medio sintético con glucosa y amonio como única fuente de nitrógeno. B.D. indica por debajo del límite de detección, que se estimó en 0.005 μmol min^{-1} mg^{-1} de proteína. Los datos representan el promedio de ± desviación media de los cultivos por duplicado biológicos independientes.

Cepa	Descripción	Actividad (μmol min^{-1} mg^{-1} de proteína)
S. cerevisiae IME215	*MATa ura3-52* pUDE264 (*TDH3*$_{pr}$-*KIGBU1-CYC1*$_{ter}$ *URA3).*	0.17 ± 0. 006
S.cerevisiae CEN.PK113-7D	Referencia prototrófica *MATa*	B.D.

Figuras legendarias

Figura 1. Visión general de las reacciones clave en el metabolismo de la arginina eucariota. Las líneas gruesas indican reacciones de ureohidrolasa. EC 3.5.3.1: arginasa, EC 4.1.1.17, ornitina descarboxilasa, EC 2.6.1.13: ornitina aminotransferasa, EC 1.5.1.2: pirrolina-5-carboxilato reductasa, EC 1.5.99.8: prolina deshidrogenasa, EC 1.5.1.12: 1-pirrolina-5-carboxilato deshidrogenasa, EC 2.6.1.-aminotransferasa, EC 4.1.1.75 2-oxo ácido descarboxilasa, EC 1.2.1.54 gammaguanidinobutiraldehído deshidrogenasa, EC 3.5.3.7 guanidinobutirato, EC 2.6.1.19 GABA transaminasa, EC 1.2.1.16 succinato-semialdehído deshidrogenasa.

Figura 2. A- Alineación de secuencias de aminoácidos hidrolasa ácida de guanidino (EC.3.5.3.7) de levaduras Saccharomycotina. Las secuencias de aminoácidos se alinearon usando Clustal W (V1.7). Los residuos de aminoácidos conservados en todas las secuencias se representan con el símbolo *. XP_456325: referencia GBU1 de *Kluyveromyces lactis* NRRL Y-1140; XP_002498240: *Zygosaccharomyces rouxii* CBS 732; XP_716668: *Candida albicans* SC5314; XP_461566: *Debaryomyces hansenii* CBS767; EIF45280: *Dekkera bruxellensis* AWRI1499; BAO40383: *Kluyveromyces marxianus* DMKU3-1042; AADM01000201.1: *Lachancea waltii* NCYC 2644; XP_503530: *Yarrowia lipolytica* CLIB122; EFW95653.1: *Ogataea parapolymorpha* DL-1 (*Hansenula polymorpha*)*;* XP_002552049: *Lachancea thermotolerans* CBS 6340 (*Kluyveromyces thermotolerans* CBS6340); XP_001523956: *Lodderomyces elongisporus* NRRL YB-4239; XP_001482640: *Meyerozyma guilliermondii* ATCC 6260 (*Pichia guilliermondii* ATCC6260); XP_004196483: *Millerozyma farinosa* CBS 7064 (*Pichia sorbitophila*)*;* XP_001385334: *Scheffersomyces stipitis* CBS 6054 (*Pichia stipitis* CBS 6054); XP_003679661: *Torulaspora delbrueckii* CBS 1146; AACE03000003.1: *Lachancea kluyveri* NRRL Y-12651. La secuencia de consenso se presenta bajo la alineación de la secuencia y el aminoácido conservado en al menos el 50% de las secuencias. Los aminoácidos en negrita y subrayados indican un dominio conservado encontrado en hidrolasa ácida de guanidino.

B- Weblogo (Schneider and Stephens (1990). Nucleic Acids Research 18, 6097-6100) que representan los motivos de hidrolasa ácida de guanidino conservados de la secuencia de alineación.

Figura 3. Entrada de GenBank de la proteína *K. lactis* XP_456325.1

Figura 4. Alineación de secuencias de aminoácidos de hidrolasa amida de guanidino bacteriana (EC.3.5.3.11). A. Las secuencias de aminoácidos se alinearon usando Clustal W (V1.7). Los residuos de aminoácidos conservados en todas las secuencias se representan con el símbolo *. NP_289508: *Escherichia coli;* WP_000105576: *Shigella flexneri;* YP_005016506: *Klebsiella oxytoca* KCTC 1686; YP_001337000: *Klebsiella pneumoniae subsp. pneumoniae* MGH 78578; YP_003614749: *Enterobacter cloacae subsp. cloacae* ATCC 13047; YP_001455807: *Citrobacter koseri* ATCC BAA-895; XP_004532666: *Ceratitis capitata;* WP_006734551: *Salmonella enterica;* AHE29794: *Burkholderia pseudomallei* NCTC 1317. B- Secuencia de consenso generada a partir de la alineación de la secuencia. La secuencia

de consenso se presenta bajo la alineación de la secuencia y el aminoácido conservado en al menos el 50% de las secuencias. Los aminoácidos en negrita y subrayados indican dominios conservados encontrados en hidrolasa amida de guanidino.

B- Weblogo que representa los motivos de hidrolasa guanidino amida conservados de la secuencia de alineación.

Figura 5. Entrada en GenBank de la proteína de *E. coli* (agmatinasa) AAC75974.1.

Figura 6. Vector 1 y 2 con todas las partes esenciales para el casete de deleción estándar. La superposición de 400 bases en el marcador de selección KlGBU1 (indicado por una cruz) está diseñada para recombinarse debido a la homología.

Figura 7. Mapa del plásmido pDS8 que comprende el módulo marcador KlGBU1YM.

Figura 8. Perfil de crecimiento típico de las cepas de *S. cerevisiae* en guanidinobutirato: las cepas de *S. cerevisiae* CEN.PK113-7D (círculo cerrado) e IME215 ($TDH3_{pr.}$::*KlGBU1*::$CYC1_{ter}$) (cuadrado cerrado) se cultivaron en matraces de agitación en medio sintético de glucosa que contiene guanidinobutirato como única fuente de nitrógeno.

Figura 9. Transformación de *S. cerevisiae* y S. *pastorianus* con pUDE264. Las cepas *S. cerevisiae* CEN.PK113-7D y S. *pastorianus* CBS1483 se transformaron con el plásmido pUDE264 que porta el gen de guanidinobutirasa de *K. lactis*. Las células transformadas (50 μl) se sembraron en placas en SM con diferentes fuentes de nitrógeno ($(NH_4)_2SO_4$, guanidinobutirato y agmatina). Las cepas no transformadas (-ADN) también se sembraron en medios similares a los del control negativo.

Figura 10. Ejemplos de casetes de deleción. Los casetes de deleción contienen varias regiones: A- (1) una secuencia de 50 a 80 pb homóloga a la parte corriente arriba del gen que se eliminará, incluido el codón de inicio, y una secuencia de 50 a 80 pb homóloga a la parte corriente abajo del gen que se eliminará, incluido el codón de parada y (4) el casete seleccionable que incluye el promotor TEF2 de *Ashbya gossypii*, el gen GBU1 de *K. lactis* y el terminador TEF2 de *A. gossypii*. B- Además de la región (1) y (4), el casete de deleción incluye (2) una secuencia de 40 pb flanqueada por (3) un sitio de restricción I-SceI ubicado corriente arriba y corriente abajo del módulo marcador.

Ejemplos

Ejemplo 1

Materiales y métodos

Cepas y mantenimiento

Las cepas de *Saccharomyces cerevisiae* usadas en este estudio se enumeran en la tabla 2. Las cepas de *S. cerevisiae* se construyeron en el fondo de CEN.PK (Nijkamp et al., (2012) Microb Cell Fact 11: 36; Entian & Kötter, (2007) Meth Microbiol 36: 629-666). Las cepas de levadura que no tenían un plásmido se mantuvieron en medio YPD (agua desmineralizada; 10 g/L de extracto de levadura; 20 g/L de peptona; 20 g/L de glucosa). Las cepas de levadura que llevan el plásmido se mantuvieron en un medio sintético que contenía sales, oligoelementos y vitaminas, preparadas y esterilizadas como se describió anteriormente (Verduyn et al., (1992) Yeast 8: 501-517) en las que se usó urea en lugar de sulfato de amonio cuando fue aplicable. Las reservas de cultivo se prepararon a partir de matraz agitado. Los cultivos se incubaron a 30 °C y se agitaron a 200 rpm, mediante la adición de glicerol al 20% (v/v) y se almacenaron a -80 °C.

Condiciones de los medios y el cultivo.

Los experimentos de crecimiento se llevaron a cabo en un medio sintético que contenía sales, oligoelementos y vitaminas, preparado y esterilizado como se describe anteriormente (Verduyn et al., (1992) Yeast 8: 501-517). Se añadió glucosa a una concentración final de 20.0 g/L. Cuando el sulfato de amonio no era la fuente de nitrógeno en el medio sintético, se reemplazó por guanidinobutirato, que se esterilizó por filtración y se añadió a un medio estéril a concentraciones de 2.9 g/L. Además, se añadieron 3.3 g/L de sulfato de potasio para compensar la eliminación del sulfato de amonio.

Si fuera necesario, se añadieron 0.15 g/L de uracilo y/o 200 mg/L de G418 (Geneticin) para completar el medio. Las placas de agar de selección se hicieron añadiendo 20.0 g/L de agar a estos medios sintéticos.

Los cultivos en matraz de agitación se realizaron en matraces de agitación de 500 ml o 250 ml que contenían 100 ml o 20 ml de medio líquido respectivamente y se incubaron en un agitador orbital (New Brunswick Scientific, Edison, NJ) a 200 rpm a 30 ºC.

Clonación y sobreexpresión de *K. lactis* KLLA0F27995g en *S. cerevisiae*. El ADN genómico de la cepa de referencia prototrófica *S. cerevisiae* CEN.PK113-7D se preparó como se describió anteriormente (Burke et al., 2000. Cold Spring Harbor Laboratory. Methods in yeast genetics: a Cold Spring Harbor Laboratory course manual). ORF KLLA0F27995g (*KlGBU1*) se clonó a partir de ADN genómico usando Phusion Hot-Start polimerasa (Finnzymes) y los cebadores cebador hacia adelante GBU1/cebador reverso GBU1, tabla 3.

Los productos de la PCR se digirieron con SpeI y XhoI clonados en pAG426GPDccdB (Alberti et al., (2007) Yeast 24: 913-919; Tabla 4) digeridos preliminarmente con las mismas enzimas, usando la ligasa T4 (Life Technologies, Breda, Países bajos) resultando en el plásmido pUDE264.

El plásmido pUDE264 se transformó en *S. cerevisiae* CEN.PK113-5D (ura3-52) usando el método LiAc como se describió anteriormente en (Gietz & Woods, (2002) Methods Enzymol 350: 87-96) dando como resultado la cepa IME215. (MATa ura3-52 pUDE264 ($TDH3_{pr}$-KlGBU1-$CYC1_{ter}$ URA3 2μ). El plásmido del esqueleto pAG426GPD-ccdB se transformó en *S. cerevisiae* CEN.PK113-5D (ura3-52) usando el método LiAc como se describió anteriormente en (Gietz & Woods, (2002) Methods Enzymol 350: 87-96) resultando en la cepa IMZ312 (MATa *ura3-52* pAG426GPD-ccdB (*$TDH3_{pr}$-$CYC1_{ter}$ URA3* 2μ*).*

Construcción del plásmido pDS8

El ORF de KLLA0F27995g (KlGBU1) se transfirió de pUDE264 a pDS1 en lugar del gen kanR. Para la construcción del plásmido pDS1, el vector pUG6 ((Guldener et al. (1996) Nucleic Acids Res 24:2519-24) se usó como plantilla para la PCR usando la polimerasa de ADN PhusionTM Hot Start II (Thermo Scientific, Waltham, MA) según a las instrucciones del fabricante y los pares de cebadores pUGfw/pUGrv para amplificar el esqueleto del vector y pDSfw/pDSrv para amplificar el marcador de selección. El cebador pUGfw contenía el sitio de restricción XbaI y una secuencia sintética de 40 pb (repetición B) que se usó para construir una repetición directa de 80 pb. El cebador pUGrv contenía el sitio de restricción BglII y otra secuencia sintética de 40 pb (repetición A) para construir la segunda repetición directa de 80 pb. El cebador pDSfw contenía el sitio de restricción BglII seguido de una repetición B de 40 pb. El cebador pDSrv contenía el sitio de restricción XbaI y la repetición A de 40 pb. Todas las PCR se visualizaron con electroforesis en gel. Los fragmentos se aislaron del gel de agarosa usando el kit de recuperación de ADN en gel ZymocleanTM. (Zymo Research, Irvine, CA). Los fragmentos de PCR se restringieron con XbaI y BglII y se ligaron con ADN ligasa T4 (Thermo Scientific) según las instrucciones del fabricante. La sustancia química competente Escherichia coli, cepa DH5a, se transformó con la mezcla de unión y se seleccionó una colonia correcta mediante PCR usando los siguientes cebadores FK140 pUDI065 fw, SLT1_control_rv, KANMX4 fw y FK105-MP1.

La transferencia de KlGBU1 en lugar de la secuencia kanR se logró mediante el ensamblaje de (Gibson et al., (2009) Nat Methods 6: 343-345). El fragmento portador del gen KlGBU1 se amplificó por PCR usando la polimerasa Phusion Hot-Start (Finnzymes) y los cebadores GBU1-fw y GBU1-rv (Tabla 3). El plásmido pUDE264 se usó como plantilla. El plásmido principal se amplificó por PCR usando la polimerasa de inicio en caliente de Phusion (Finnzymes) y los cebadores Esqueleto-fw y Esqueleto-rv. El plásmido pDS1 se usó como plantilla. Los dos fragmentos compartieron al menos 43 nucleótidos de identidad en sus flancos y se ensamblaron in vitro usando el kit de clonación de ensamblaje Gibson de New England Biolabs (Ipswich, MA). La mezcla ensamblada se transformó en *E. coli* DH5α y se sembró en una placa LB que contenía ampicilina (100 mg/L). El plásmido ensamblado que contiene el módulo marcador KlGBU1YM se verificó mediante análisis de restricción y secuenciación y un clon correcto se denominó pDS8.

Deleción de ADE2 en *S. cerevisiae*.

Las deleciones de genes en *S. cerevisiae* se realizaron mediante la integración de los casetes KlGBU1YM (sitio SceI::*$AgTEF2_{pr}$-KlGBU1-$AgTEF2_{ter}$*::sitio SceI) a través del método de PCR de homología de flanco corto (Wach et al., (1994) Yeast 10: 1793-1808). Las secuencias de los cebadores oligonucleotídicos se muestran en la tabla 3. El casete de deleción para ADE2 se amplificó usando la polimerasa Phusion Hot-Start (Finnzymes, Landsmeer, Países Bajos) y el plásmido de plantilla pDS8 usando los cebadores GBU1-ADE2ko-CENPK-fw/ GBU1-ADE2ko-CENPK-rv. La transformación de *S. cerevisiae* CEN.PK113-7D con los casetes de deleción ADE2 se realizó usando el método LiAc como se describió anteriormente en (Gietz & Woods, (2002) Methods Enzymol 350: 87-96) dando como resultado la cepa IMX598.

La integración correcta del casete KlGBU1SY y la sustitución del gen de interés se verificó mediante PCR diagnóstica usando un cebador hacia adelante específico para la UTR 5 '(región no traducida) de ADE2 y el cebador reverso para el casete de deleción (Tabla 3).

Preparación de extractos celulares.

Para la preparación de extractos celulares, se recogieron muestras de cultivo mediante centrifugación, se lavaron dos veces con solución reguladora de fosfato de potasio 10 mM (pH 7.5) que contenía EDTA 2 mM y se almacenaron a -20 °C. Antes de la rotura celular, las muestras se descongelaron a temperatura ambiente, se lavaron y se resuspendieron en solución reguladora de fosfato de potasio 100 mM (pH 7.5) que contenía MgCl2 2 mM y ditiotreitol 2 mM. Los extractos se prepararon por sonicación con perlas de vidrio de 0.7 mm a 0 °C, durante 2 minutos a intervalos de 0.5 minutos con un sonicador MSE (Wolf Laboratories Limited, Pocklington, Reino Unido) (150 W de salida; 8 μm de amplitud pico a pico) (Luttik et al., (2008) Metab Eng 10: 141-153). Las células y los residuos no rotos se eliminaron mediante centrifugación a 4 °C (20 min; 36000 x g). El extracto celular resultante se usó para ensayos enzimáticos.

Ensayos de actividad enzimática

Para el ensayo enzimático de la arginasa, se activaron 50 μL de extracto celular en 950 μL de solución reguladora de maleato de manganeso (sulfato de manganeso 50 mM, ácido maleico 50 mM, pH 7) durante 1 hora a 37 °C (Messenguy et al., (1971) Eur J Biochem 22: 277-286). La mezcla de reacción para los ensayos de arginasa, preparada en tubos eppendorf oscuros, contenía 60 μL de extracto celular activado, 400 μL de solución de arginina 713 mM (pH 9.5) y agua desmineralizada hasta 1 mL. La mezcla de reacción se incubó durante 30 minutos a 37 ºC. Para detener la reacción, se añadieron a la mezcla de reacción 0.7 mL de mezcla de ácido sulfúrico y fosfórico (20% v/v de ácido sulfúrico concentrado y 60% v/v de ácido fosfórico almibarado en agua desmineralizada). La cantidad de urea producida se cuantificó usando el método de Archibald (Archibald, (1945) J Biol Chem 157: 507-518) con una línea de calibración que varía desde 0 hasta 0.6 mM de urea. Se añadieron 0.06 mL de una solución de 4% v/v de α-isonitroso-propiofenona en etanol y las muestras se mezclaron completamente antes de hervir durante 1 h en un baño de agua a 100 °C para desarrollar el color. Las muestras se enfriaron a temperatura ambiente durante 15 minutos y la absorbancia a 540 nm se midió en un espectrofotómetro Libra S11 (Biochrom, Cambridge, Reino Unido).

La mezcla de reacción para los ensayos enzimáticos de guanidinobutirosa (GBU) se preparó en tubos eppendorf oscuros, que contenían en un volumen final de 1 ml: solución reguladora de glicina 50 mM (pH 9), $MnSO_4$ 5 mM y 50 μL a 100 μL de extracto celular. La reacción se inició mediante la adición de ácido guanidinobutírico 50 mM. Después de 30 minutos de incubación a 37 °C, la reacción se detuvo mediante la adición de 700 μL de mezcla de ácido sulfúrico-fosfórico (20% v/v de ácido sulfúrico concentrado y 60% v/v de ácido fosfórico almibarado en agua desmineralizada). La cantidad de urea producida se cuantificó usando el método de Archibald (Archibald, 1945) con una línea de calibración que va desde 0 hasta 0.6 mM de urea. Se añadieron 0.06 mL de una solución de 4% v/v de α isonitrosopropiofenona en etanol y las muestras se mezclaron completamente antes de hervir durante 1 h en un baño de agua a 100 °C para desarrollar el color. Las muestras se enfriaron a temperatura ambiente durante 15 minutos y la absorbancia a 540 nm se midió con un espectrofotómetro Libra S11 (Biochrom).

Resultados

La expresión de KlGBU1 en *S. cerevisiae* confiere la capacidad de crecer en guanidinobutirato como única fuente de nitrógeno.

Para caracterizar la función del gen ureohidrolasa putativo de *K. lactis*, KLLA0F27995g se clonó bajo el control del promotor TDH3 constitutivo fuerte ($TDH3_{pr}$) en un vector de expresión y se transformó en *S. cerevisiae* (Figura 8).

El plásmido de expresión pUDE264 se transformó en la cepa de laboratorio CEN.PK113.7D (protótrofo) y la cepa CBS1483 de *Saccharomyces pastorianus* de elaboración de cerveza silvestre (www.cbs.knaw.nl/Collections/). Las células transformadas se siembran en placas en medio sintético que contenía guanidinobutirato o agmatina o sulfato de amonio. Como era de esperar, para ambas cepas, todos los transformantes podrían crecer en placas no selectivas que contienen sulfato de amonio. En línea con las mediciones de enzimas, no se rescataron transformantes en placas de agmatina, mientras que se detectaron más de mil y cien transformantes en placas de guanidinobutirato para CEN.PK113-7D y CBS1483, respectivamente (Figura 9). Esto demostró que KlGBU1 podría usarse eficientemente como un marcador seleccionable dominante en la cepa prototrófica de *S. pastorianus*.

Construcción de plásmidos y casetes de deleción.

La secuencia codificante del gen GBU1 de *K. lactis*, flanqueada por el promotor y terminador TEF2 de *Ashbya. gossypii* se clonó en el vector pDS1 reemplazando el gen KanMX, dando como resultado el plásmido pDS8 (Figura 7). El módulo KlGBU1YM resultante solo contenía secuencias heterólogas, lo que reduce la probabilidad de integración mal dirigida (Wach et al., (1994) Yeast 10: 1793-1808). El plásmido pDS8 se puede usar fácilmente como plantilla para los casetes de deleción que contienen el nuevo módulo marcador KlGBU1YM y se usó para la construcción de todos los casetes de deleción usados en este estudio.

Los casetes de deleción contenían tres regiones principales (Fig. 10): (1) una secuencia de 50 a 55 pb homóloga a la parte corriente arriba del gen que se eliminará, incluido el codón de inicio, y una secuencia de 50 a 55 bp homóloga a la parte de corriente abajo del gen que se va a eliminar, incluido el codón de parada. Estas regiones se usaron para la recombinación homóloga dirigida (Baudin et al., (1993) Nucleic Acids Res 21: 3329-3330), (2) una secuencia de 40 pb flanqueada por (3) un sitio de restricción I-SceI ubicado corriente arriba y corriente abajo del módulo marcador y (4) el marcador KlGBU1YM. Tras la restricción por la endonucleasa, la recombinación homóloga de la repetición directa sería suficiente para extraer y reciclar el módulo marcador.

Eliminación de genes en *S. cerevisiae* usando KlGBU1YM

Para evaluar si el nuevo marcador KlGBU1YM era apropiado para la desactivación del gen en *S. cerevisiae*, se intentó eliminar un gen en la cepa de laboratorio CEN.PK113-7D. Se seleccionó ADE2 para este experimento de prueba de principio porque el fenotipo causado por la deleción de ADE2 se puede detectar visualmente, lo que brinda una evaluación preliminar rápida de la integración dirigida. ADE2 codifica para la enzima fosforibosilaminoimidazol carboxilasa, que participa en la biosíntesis de los nucleótidos de purina. Los mutantes ade2 requieren una fuente externa de adenina y acumulan precursores de nucleótidos de purina en la vacuola que dan a las colonias un color rojo (Zonneveld & van der Zanden, (1995) Yeast 11: 823-827).

El potencial de KlGBU1YM como marcador dominante se probó transformando un casete de deleción para interrumpir ADE2 en CEN.PK113-7D. Después de la transformación, las células se cultivaron en placas de agar de medio sintético (SM) que contenían guanidinobutirato como única fuente de nitrógeno. La deleción de genes dirigida se confirmó por la incapacidad de las colonias individuales para crecer en SM (en ausencia de adenina) y por PCR. La eficiencia de transformación promedio fue de 5 transformantes por microg de ADN, con el 100% de las colonias albergando la integración correcta y pudiendo crecer sobre guanidinobutirato como única fuente de nitrógeno.

REIVINDICACIONES

1. Un método de cultivo de un microorganismo seleccionado de los géneros *Saccharomyces sensu stricto*, *Kazachstania*, *Naumovozyma*, *Nakaseomyces* y *Vanderwaltozyma* en presencia de guanidinobutirato como única fuente de nitrógeno, que comprende:

(a) introducir una molécula de ácido nucleico que comprende una secuencia de nucleótidos que codifica una guanidinobutirasa de la familia ureohidrolasa en el microorganismo, por lo que la secuencia de nucleótidos está unida operativamente a las secuencias promotora y terminadora;

(b) cultivar el microorganismo de modo que la molécula de ácido nucleico que codifica la guanidinobutirasa se exprese en el microorganismo; y

(c) cultivar el microorganismo en presencia de guanidinobutirato como única fuente de nitrógeno,

en el que la guanidinobutirasa codificada es una hidrolasa ácida de guanidino con el número EC 3.5.3.7.

2. El método según la reivindicación 1, en el que la guanidinobutirasa codificada comprende uno o ambos dominios conservados como se indica en la figura 2A y se representa en la figura 2B.

3. El método según la reivindicación 1 o la reivindicación 2, en el que dicha secuencia de nucleótidos que codifica guanidinobutirasa codifica guanidinobutirasa de *Kluyveromyces lactis* que tiene el número de acceso de GenBank XP_456325.1, como se representa en la figura 3.

4. El método según una cualquiera de las reivindicaciones 1-3, en el que las secuencias promotora y/o terminadora se seleccionan de un gen glicolítico seleccionado de PGI1, PFK1, PFK2, FBA1, TPI1, TDH1, TDH3, PGK1, GPM1, PYK1, ENO1, y/o ENO2.

5. Un microorganismo seleccionado de los géneros *Saccharomyces sensu stricto*, *Kazachstania*, *Naumovozyma*, *Nakaseomyces* y *Vanderwaltozyma* que comprende una secuencia de nucleótidos transformada que codifica una guanidinobutirasa, en la que la guanidinobutirasa codificada de la familia ureohidrolasa comprende uno o ambos dominios conservados como se indica en la figura 2A y se representa en la figura 2B.

6. El microorganismo según la reivindicación 5, en el que dicha secuencia de nucleótidos que codifica guanidinobutirasa codifica guanidinobutirasa de *Kluyveromyces lactis* que tiene el número de acceso de GenBank XP_456325.1, como se representa en la figura 3.

7. Un conjunto de construcciones, que comprende una primera construcción que comprende una primera parte de una secuencia de nucleótidos que codifica una guanidinobutirasa, y una segunda construcción que comprende una segunda parte de la secuencia de nucleótidos que codifica la guanidinobutirasa, por lo que un fragmento de la primera parte de la secuencia de nucleótidos se superpone con un fragmento que está presente en la segunda parte de la secuencia de nucleótidos, lo que permite la recombinación entre la primera y la segunda parte de la secuencia de nucleótidos para expresar funcionalmente el marcador de selección de guanidinobutirasa, en el que la guanidinobutirasa codificada de la familia ureohidrolasa comprende uno o ambos dominios conservados como se indica en la figura 2A y se muestra en la figura 2B.

8. El conjunto de construcciones según la reivindicación 7, en el que la primera construcción comprende además un sitio de reconocimiento para una endonucleasa y una primera región de homología con un genoma diana de un microorganismo, y la segunda construcción comprende además una segunda región de homología con el genoma diana del microorganismo, y una copia del sitio de reconocimiento de endonucleasa, mediante el cual

una secuencia codificante que codifica la endonucleasa y que está acoplada a un promotor inducible está presente en la primera o segunda construcción; y

una parte de la primera región de homología con el genoma diana en la primera construcción se duplica entre la copia del sitio de reconocimiento de endonucleasas y la segunda región de homología con el genoma diana en la segunda construcción; o una parte de la segunda región de homología con el genoma diana en la segunda construcción se duplica entre la primera región de homología con el genoma diana y el sitio de reconocimiento de endonucleasa en la primera construcción.

9. El conjunto de construcciones según la reivindicación 7 o la reivindicación 8, en el que el fragmento superpuesto de la secuencia de nucleótidos que codifica la guanidinobutirasa es de aproximadamente 200 pares de bases.

10. El conjunto de construcciones según una cualquiera de las reivindicaciones 7-9, en el que la región duplicada de homología con el genoma diana en la primera y la segunda construcciones está preferiblemente entre 20 pb y 200 pb.

11. Un kit que comprende el conjunto de construcciones de una cualquiera de las reivindicaciones 7-10.

12. Un método para alterar un genoma, preferiblemente un gen diana, en un microorganismo seleccionado de los géneros *Saccharomyces sensu stricto*, *Kazachstania*, *Naumovozyma*, *Nakaseomyces* y *Vanderwaltozyma*, que comprende proporcionar el conjunto de construcciones según una cualquiera de las reivindicaciones 7-10 a dicho microorganismo, y seleccionando un microorganismo en el cual el genoma ha sido alterado por la selección de un microorganismo que exprese funcionalmente dicha guanidinobutirasa.

13. El método según la reivindicación 12, en el que un microorganismo se selecciona cultivando un microorganismo en presencia de guanidinobutirato como única fuente de nitrógeno.

14. El método de la reivindicación 12 o 13, en el que se selecciona un microorganismo en el que el genoma se ha alterado mediante la inserción de guanidinobutirasa funcional recombinada.

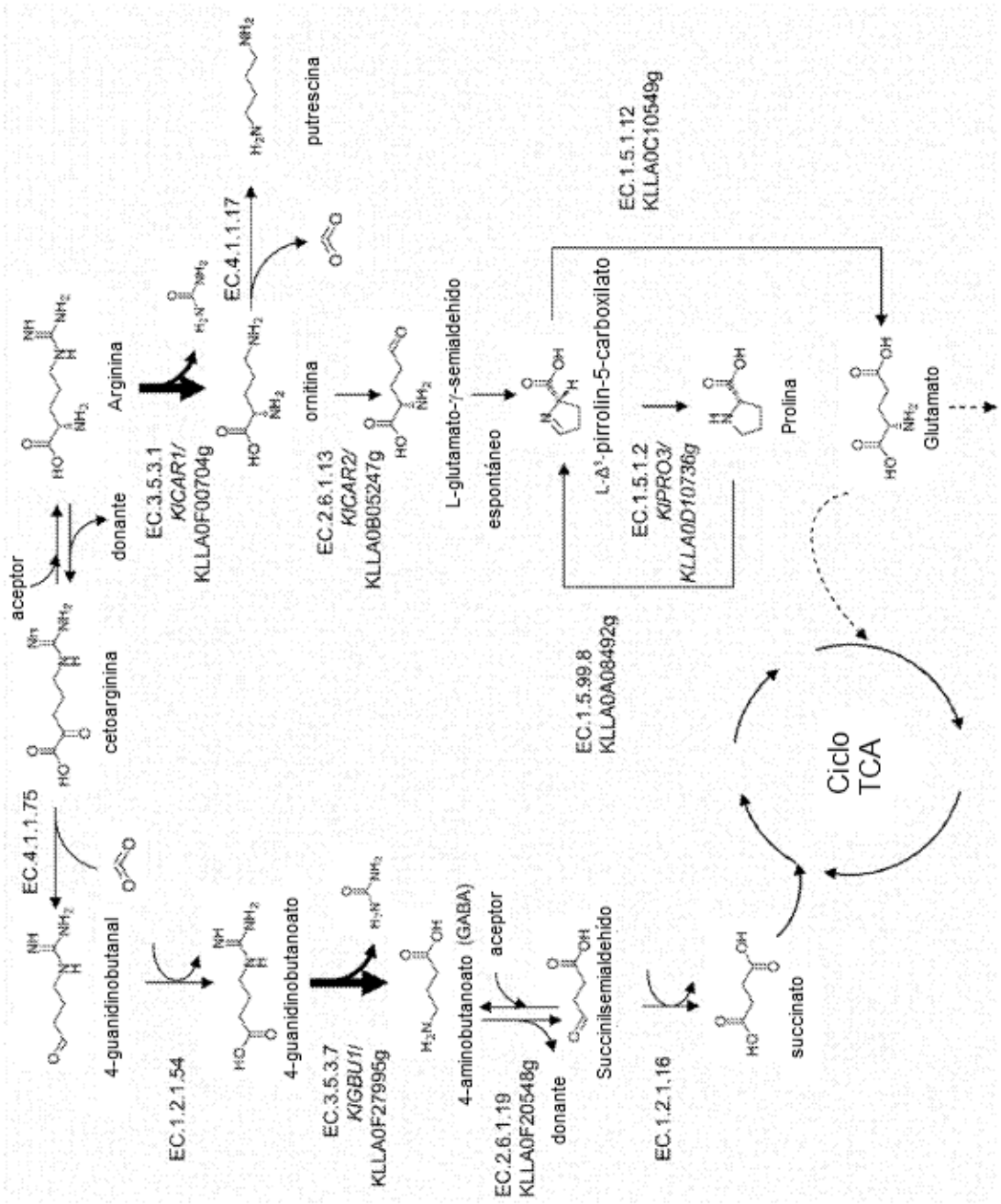
aceptor
donante
Arginina
cetoarginina
EC.4.1.1.75
4-guanidinobutanal
EC.1.2.1.54
4-guanidinobutanoato
EC.3.5.3.7
KlGBU1/
KLLA0F27995g
4-aminobutanoato (GABA)
aceptor
EC.2.6.1.19
KLLA0F20548g
donante
Succinilsemialdehido
EC.1.2.1.16
succinato
Ciclo
TCA
EC.3.5.3.1
KlCAR1/
KLLA0F00704g
ornitina
EC.4.1.1.17
putrescina
EC.2.6.1.13
KlCAR2/
KLLA0B05247g
L-glutamato-γ-semialdehido
espontáneo
L-Δ¹-pirrolin-5-carboxilato
EC.1.5.1.2
KlPRO3/
KLLA0D10736g
EC.1.5.99.8
KLLA0A08492g
Prolina
EC.1.5.1.12
KLLA0C10549g
Glutamato

FIG. 1

FIG. 2A

```
XP_456325          ----MKVAGFI-LG-ALI---------QFSLTE----------------GHVE-------
XP_002498240       ----MRFSTLASLA-ALF---------EFT-TA----------------GR---------
XP_716668          ----MKLLALL----TLL---------PLVI----------------------------
XP_461566          ----MKVLNIS-TI-LFI---------ATTTTA----------------NQLH-------
EIF45280           ----------------------------MPFPK----------------SQSE-------
BAO40383           MINWSYYALFC-L--AIV---------PFAKCE----------------YYNE-------
AADM01000201.1     ------------------------------------------------------------
XP_503530          -----MDSEYG-DL-SVE---------QKPIKDHSHSHHHHDHGHGHGNGHVE-------
EFW95653           ----MAVAK-A-LQ-SPI---------QFA------------------------------
XP_002552049       ML--MRSVLFS-L--ACA---------NAVLAS----------------GTTV-------
XP_001523956       ----MKLLLTL-LG--IL---------QVCIA----------------------------
XP_001482640       ----MKLLPLL-FATALA---------QLTLKN----------------DNVV-------
XP_004196483       MLKSISYWAFI-LS-IYLGVVCADEAEQIASAE----------------GFAK-------
XP_001385334       ----MK-CSHI-VS-API---------FLSVAA----------------AVLQPVAWKSD
XP_003679661       -------MLYL-FS-LCV---------LLSLAW----------------CY-E-------
AACE03000003.1     ------------------------------------------------------------

Consenso                   M
```

```
XP_456325          -------Q-N---------------ENA-NLT-EMWGEDWP-------FSGIQTFAHLPH
XP_002498240       -----------------------------------WTEDWP-------FAGFPSFAKLPF
XP_716668          --------------------------ST-NLE-EKWGGLWP-------FQGIATFAHLEH
XP_461566          -------S-DPFSDNVVYFDTHSAEREP-NLK-DMWDDLWP-------FQGINTFAHLEH
EIF45280           -------S-R---------------KSP-TLE-QLWGQXWP-------FTGIPTFAHLNT
BAO40383           -------DIG---------------TNS-TLN-DMWGEEWP-------FSGIQTFAHLPH
AADM01000201.1     ------------------------------------------------------------
XP_503530          -------V-----------------EYL-NLPGNNFYEDLDAELNGPLYAGIQTFAHLGH
EFW95653           -------S-N---------------SEP-TLD-QLWGEDWP-------FNGIQTFAHLNY
XP_002552049       -------L-E---------------ED---LE-QMWGQDWP-------FSGIQTFAHLPH
XP_001523956       ---------------------------T-DLE-EKWGSLWQ-------FQGIATFAHLEH
XP_001482640       -------Y-D---------------SNPPSLS-EMWDGLWP-------FQGINTFAHLDH
XP_004196483       -------M-Q---------------RNA-TLE-EMWGELWP-------FQGINTFAHLEH
XP_001385334       ELINDSID-G---------------SSP-SLK-EMWDDLWP-------FQGINTFAHLDH
XP_003679661       -------D-R---------------ATD-DLD-RLWGQDWP-------FSGINTYAHLPH
AACE03000003.1     ------------------------------------------------------------

Consenso           ------------------------------L----WG-WP        F-GI-TFAHL-H
```

FIG. 2A, Continuación

```
XP_456325          HKCLIDME---KKFDIGVIGVPFDTAVSFRGGARFGPQAIRKASQRQTSMRGFNFRADIN
XP_002498240       HTCLVD-N---PEFDIALIGVPFDTAVSFRPGARFGPQAIRRASQRQNGLRGFNARAGIN
XP_716668          FQCLIESE---KQFDIGIIGVPFDTAVSYRPGARFGPRAIRDASQRQNNLRGFNPKALFD
XP_461566          NKCLLDPD---QEYDIGIIGVPFDTATSYRPGARFGPRAIRSASQRQTSLRGYNQRADFN
EIF45280           TKCLLSPE---SXYDIGIIGVPFDTATSYRPGARFGPQAIRLASQRQNSMRGFNTRAGIN
BAO40383           QKCLLDMG---TKFDIGVIGMPFDTAVSYRPGARFGPQGIRKASQRQNSMRGFNFRAGIN
AADM01000201.1     --------------------MPFDTAVSYRPGARFGPQAIRRSSQRQNSMRGFNSRAGIN
XP_503530          VSCFDPTNFASEQFDIALVGAPFDTAVTFRSGARFGPAGIRKGSRRMSPGQVSPYREGFM
EFW95653           TKCLVDPE---TSFDIGVIGVPFDTATTYRSGARFGPRAIRTGSQRQTSKRAFNTRAGIN
XP_002552049       EKCLMNRS---LDFDIGVIGIPFDTAVTYRPGARFGPQAIRKSSQRQTSMRGFNFRAGIN
XP_001523956       TQCLVNPK---ESFDVAVIGVPFDTAVSYRPGARFGPRAIRDASQRQYSLRGFNHRALFN
XP_001482640       KVCLTQPD---EIYDFAIIGVPFDTAVLYRPGARFGPRAIRAAAQRQTSLRGYNQRANFN
XP_004196483       FKCLVEQD---EKYDIGIIGVPYDTSVSYRPGARFGPRAIRTASQRQTSLRGFNQRAFFN
XP_001385334       HKCLLEPG---NTYDVALIGVPFDTAVSYRPGARFGPRAIRAASQRQTSLRGYNQRANFN
XP_003679661       QKCLLDKN---FTFDIGIIGVPFDSAVTYRPGARFGPQAIRAASQRQIPIRSFNFRAGIN
AACE03000003.1     ----------------------------------------------MRGFNFRAGIN

Consenso           -KCL----   --FDIG-IGVPFDTAVS-R-GARFGP-AIR-ASQRQ-S-RGFN-RA-IN

XP_456325          PYQDWASVVDCGDVPVTPMDNCLALKMMTAAYENLLSHESQTSDNNLP--PRFVTLGGDH
XP_002498240       PYDNWAHIIDCGDIPVTPMDNQLALEQMNAAYDELV-NGTTTSGNTAGALPRLVSLGGDH
XP_716668          PYQSWARIIDCGDIPVTPMDNSAAYKQMSEAFKDLLNRKS-SNNTEIP--PRYIALGGDH
XP_461566          PYTSWAKVIDCGDMPVTPMDNSLAFKQMNKGFEELIARRNSKNSTVTP--PRYIALGGDH
EIF45280           PYQNWASLVDCGDIPVTPMDNKVALDQMTAAFEELLLRRNSSLGDAHP--PRYVALGGDH
BAO40383           PYNNWASVIDCGDVPVTPMDNNLALQMMTAAYDNLLSHESKAESNELP--PRLVTLGGDH
AADM01000201.1     PYQNWAKIMDCGDVPVTPMDNQLALKMMTSAYETLLNHSSTTKDSKLP--PRLVTLGGDH
XP_503530          LYDDWAKFVDCGDVAMHPLDNRYALNQLYRGMRAIHNHTTSTLNATHI--PRAILMGGDH
EFW95653           PYQDWAKVIDCGDIPVTPMDNELALDQMTKAFEELLLKRKNAVDGSGP--PKLVALGGDH
XP_002552049       PYQDWAKVLDCGDVPVTPMDNNLALQMMGAAYHNLLNRNSTLKQAELP--PRFATLGGDH
XP_001523956       PYKSWAKIIDCGDIPVTPMDNHLAFKQMDIAFDELL-QRSSANDSRVP--PRYVVLGGDH
XP_001482640       PYTSWAKVLDCGDIPVTPMDNHLAFKQMDLAFEELILR--RNSSSKAP--PRYIALGGDH
XP_004196483       PYTSWARIVDCGDVPVTPVDNELAFKQMTAAFEELLLRRSAKNDSSMP--PRHIILGGDH
XP_001385334       PYASWAKIVDCGDLPITPMDNSIAFTQMTKGFEELLLRRSSNSSSELP--PRYVALGGDH
XP_003679661       PYQKWAKVVDCGDIPVTPMDSSLALEMMTAAYENLLDRDSEYSKSSMP--PRLLSLGGDH
AACE03000003.1     PYQSWAKVMDCGDIPVTPMDNQLALKMMDAAYENLLDRNSTAAESPLP--PRFASLGGDH
                    *  **   ****    * *   *                          *     ****
Consenso           PY--WA---DCGD-PVTPMDN-LAL- M--A-E-LL---S-------P  PR---LGGDH
```

FIG. 2A, Continuación

```
XP_456325          SIILPALRALRKTYGRLAVIHFDSHLDTWAPSKYPSFWHSDTSEFTHGSMLWIAHNEGLL
XP_002498240       SVILPALRALHKHYGPISVIHLDSHLDTWSPDSYPSYWHSNTSEFTHGSMLWLAAQEGLI
XP_716668          SVLLPHIRALHKIYGPVNIIHFDAHLDTWKPNKYPT---SEKNDINHGSMLWKAYEEGLT
XP_461566          SVLLPHLRALHEVYGKINILHFDAHLDTWGPDKYPSFWHSKQAELNHGSMLWKANKE-CL
EIF45280           SIILPHLRALHEVYGKIAVIHFDAHLDTWTPXKYPSFWSSEQSKFTHGSMLWMAKKEGIL
BAO40383           SIILPALRSLHKLYGRLAVIHFDSHLDTWSPSKYPSFWHSDTSEFTHGSMLWIAHNEGLI
AADM01000201.1     SILLPVLRSLKEVYGPIAVIHFDSHLDTWAPAKYPSFWHSDTSEFTHGSMLWLASQEGLL
XP_503530          TTTLSALQAIYEKIGPVSVIHFDSHIDTWDPMVLGGN-VSSYMQVNHGTFLHYAAERGYL
EFW95653           SILLPHLRALNKVYGKVAVIHFDAHLDTWSPSKYPSFWSSDQSKFTHGSMLWMANEEDLL
XP_002552049       SIILPILRQLHKIYGPISVIHFDSHLDTWAPSKYPSYWHSNNSDFTHGSMLWIAKQEGLL
XP_001523956       SVLLPHLRALKKHYGRLNVLHFDAHLDTWSPLKYPSFWRTDQNDLNHGSMLWQAHEEG-L
XP_001482640       SVLLPHLRALKKAYGPLNVIHFDAHLDTWSPDKYPSFWHSDQSEITHGSMLWTAFEEG-L
XP_004196483       SVILPHLRALSKVHGPINVIHLDAHLDTWAPDKYPSFWHSAQSEVNHGSMLWKAHQEGLL
XP_001385334       SVLLPHLRALHEVYGRINVIHFDAHLDTWAPDKYPSFWHSDQSEINHGSMLWKAHHEG-L
XP_003679661       SIILPVIRNLYKLYGPITVLHFDSHLDTWSPSKYPSYWHSKSSKFTHGSMLWMAKQEGLL
AACE03000003.1     SVILPILRQLHKIYGPISVIHFDSHLDTWAPSKYPSYWHSDTSEFTHGSMLWIAKQEGLL
                      *          *     * * * *** *               **  *  *
Consenso           S--LP-LRAL-K-YG---VIHFD HLDTW-P-KYPSFWHS--SEFTHGSMLW-A--EGLL
```

```
XP_456325          TENNNIHAGLRT---RLSGSSFEDYDDDDKVGFHRIEADEIMD---GGIKS---IVEKIK
XP_002498240       NKGHCVHGGLRT---RLSGDDWSDYEEDDRVGFHRIHADEMME---IGPRG---IAERIK
XP_716668          TK-HNIHVGVRT---RL--SELDDLQDDDEQNFVRIEADDIWL---KGPQW---VVDKIL
XP_461566          TSEHNVHAGVRT---KLSG--IEDYVDDDSQNFTRITADDIWI---KGVQY---VVDKIL
EIF45280           SDDYNVHVGIRT---RISGVSWEDFDEDDDQGWLRFSADDVWV---GGKQSLDQIVASIK
BAO40383           TENSNVHAGLRT---RLSGTSYEDYDEDDQVGFYRIEADEIMD---GGPSA---IVEKIK
AADM01000201.1     SGGHNVHAGLRT---RLSGTSWEDYDEDDEVGFYRIQADEIMD---IGVHG---VAKKII
XP_503530          NHGHNLHVGSRAPYVRKHG----DIEHDKHCGFAIVNAREIDE---VGIAG---VVQKIK
EFW95653           SDDYNVHIGLRT---RISGKDWEDYEDDDDQGWARFSADDIWINGLGGLKE---IVRSIN
XP_002552049       SENSNVHAGLRT---RLSGVGWDDYEEDKETGFHRIECDEILD---IGVRG---IARKIL
XP_001523956       TTNRNVHAGVRT---KLSG--IEDYQDDDAQNWVRIEADDIWL---KGPQY---VVDKIL
XP_001482640       TTNTNIHAGLRT---KLSG--LEDYEDDDKQNFVRIYADDIWI---DGVQS---VIAKIN
XP_004196483       S-HNNVHAGLRT---KLSG--IADYEDDDAQHFTRITADDIWI---KGPDY---VLDTIL
XP_001385334       TSHHNVHAGLRT---KLSG--LEDYEDDDSQHFIRIDADDIWL---KGPQW---VVQKIL
XP_003679661       SE-HNVHAGLRT---RLSGVDWEDYEDDDDVGFHRIESDDIIR---LGVQG---LAEKIK
AACE03000003.1     AENSNVHAGLRT---RLSGVGWDDYEEDSETGFHRIECDEILK---VGVNG---IAERIL
                         * * *            *   *                    *          *
Consenso           ----N-HAGLRT   RLSG---EDY-DDD--GF-RI-AD-I--   -G---   -V-KI-
```

FIG. 2A, Continuación

```
XP_456325          SKIPSDVPVYISVDIDVLDPSAAPGTGTMEVGGWMTRELIRIIRELEDLNLVGADIVEVS
XP_002498240       QIVPKNVPVYLSVDIDVLDPSAAPGTGTVEPGGWLTRELISLIRQLQDLPLVGADVVEVS
XP_716668          ATIPKDTATYISVDVDVLDPGFTSGTGTQEPGGFLPRELIYLLRSIDGLTVVGADVVEVS
XP_461566          ETIPPDTPTYLSVDIDVLDPAFGSGTGTQEPGGWLPRELIHVLRSIENLTIVGADIVEVS
EIF45280           KRIPAHYPVYISVDVDCMDPGFTPGTGTIEPGGMMPREVIYLLRHL-DLDLVGADVVEVA
BAO40383           SKIPDNVPVYISVDIDVLDPSAAPGTGTMEAGGWLTRELIRIIRQLDHLNLVGADVVEVS
AADM01000201.1     ERVPKDIPVYISVDIDVLDPSAAPGTGTMEVGGWLTRELISIIRKLEDLTLVGADIVEVS
XP_503530          DRV-GNTNVYISVDIDVLDPVYAPGTGTAEPGGYTTREFMQILDGLEGINIVGADVVEVA
EFW95653           ERIPKDYPTYVSVDIDCLDPGFAPGTGTIESGGLLPRELFYLLRNI-DVNLVGADIVEVS
XP_002552049       DIVPKDKPVYISVDIDVLDPSAAPGTGTVEVGGLLTRELISIIRQLDGLSLIGADVVEVS
XP_001523956       ETIPKDSPTYLSVDIDVLDPGFASGTGTQESGGWLPRELIHILRGVEELTIVGADVVEVA
XP_001482640       ATIPADTPTYISVDIDVLDPGFGSGTGTQEPGGWLPRELIYVLRHIDHLTIVGGDVVEVS
XP_004196483       KVVPANTPTYISVDIDVLDPAFGSGTGTQEPGGWLPRELIYILRGLENLDVVGADIVEVS
XP_001385334       DTVPDDSPTYISVDVDVLDPGFTSGTGTQEPGGWLPRELLHVLRSIEGLTVVGGDVVEVS
XP_003679661       QLLPKKQPLYISVDIDVLDPSAAPGTGTVEAGGWLTRELIYLLRSLEDYPIVGADVVEVS
AACE03000003.1     EHVPKDKPVYISVDIDVLDPSAAPGTGTIEVGGLLTRELISIIRQLEDLHLVGADVVEVS
                            * *** *  **    **** * **   **              * * ***
Consenso           ---P-D-P YISVDIDVLDP--APGTGT-E-GGW-RELI----R-LE-L--VGAD-VEVS
```

```
XP_456325          PPFD-PTEITSLAGAQIAYELITNMVKKGPIDPELIKHN-----LELSDKLT-QGQQL--
XP_002498240       PPFD-HADVTAMAAAQVAYEIITNMVKT-PLELETRKF----------------------
XP_716668          PAYD-IAEITATNGAQIAYEVLTSMVKRGNIDKSLVKSV-----VHVFD-----------
XP_461566          PAFD-TAEITATNGAQVVFEILTSMVKKGSVG-HLVKNNNPKELLEVKSKNDGKSTQY--
EIF45280           PAYD-QAEITATNAAQVVFELVTTMVKRG-------------------------------
BAO40383           PPFD-PSEITTLAGAQVAYELITNMVKKGPLDPDMVKRN-----LESKNANH-FQMQ---
AADM01000201.1     PAYD-SGDVTSLAAAQIAYELITNMVKKGPVADEIVEKN-----KQIAENLA-ENNHV--
XP_503530          PAYDGPGDVTLLAAAQVIDSLASLMVMNGPL-------S-----TRQSTK----------
EFW95653           PQYD-HAEITATNGAEVAYQLITSIVK---------------------------QGKK---
XP_002552049       PAYD-QSDITSTAASQIVYELITNMVKKGPLDPAMIQAN-----K--NSEMD-QGDKP--
XP_001523956       PAYD-VSEVTATNGAQMAFEILTSMVKKGNVDKNIVDRT-----IEI-------------
XP_001482640       PAFD-NAEITATNGAQVAYELLTSMVKKGRVD--LLVQR-----EEATPIIK-SSEVE--
XP_004196483       PAFD-IAEITATNGAQVVFEILTSIVKKG--------NS-----EKVADSSS-QKDKG--
XP_001385334       PAFD-TAEITSTNGAQIAFEIITSMVKKGPIDPAIVKKN-----KKELVKIT-HVNELEQ
XP_003679661       PPFD-QSEITAIAASQIAYELLTSMVKSGPIEPQMIQEN-----GLFNLRAL-QDNHV--
AACE03000003.1     PAYD-HADITSTAASQIVYELITNMVKKGPVDPAIVEAN---------------------
                   *  *     *               *
Consenso           P--D---EI--T-GAQ-AYELT--MVKKGP------------------------------
```

FIG. 2A, Continuación

```
XP_456325          -------LGFSSPT--------DELNDKIQ-----KE--QFV----------LQA-----
XP_002498240       ---------FSNM-----------------------------------------------
XP_716668          ------------------------------------------------------------
XP_461566          -------LDKQEIN--------RLIENKLQ-----EF--ENIKFNLLSEIEQLRST----
EIF45280           -----------KPL--------PVRED---------------------------------
BAO40383           ----------------------------Q-----NE--QNV----------LHI-----
AADM01000201.1     -------RVTEQGT--------TKILERLI-----QDANQAV----------LHATNPLL
XP_503530          ------------------------------------------------------------
EFW95653           ------------------------------------------------------------
XP_002552049       -------QNLAENE--------YQSLGEGA-----RQ--NFMESL-------LQGRV---
XP_001523956       ---------FES------------------------------------------------
XP_001482640       -------LTLEERL--------TSEIRQLK-----NM--QAL----------FN------
XP_004196483       -------SSAALQD--------DKIKQQKQDVLTRKN--PFV----------ISN-----
XP_001385334       EKKDFIDLEKAKKTIEQKLKELDELKSELS-----SQ--LLE----------LREIPF--
XP_003679661       -------DFSAAKP--------DTDYDKLL------------------------------
AACE03000003.1     ------------------------------------------------------------

  Consenso         ------------------------------------------------------------
```

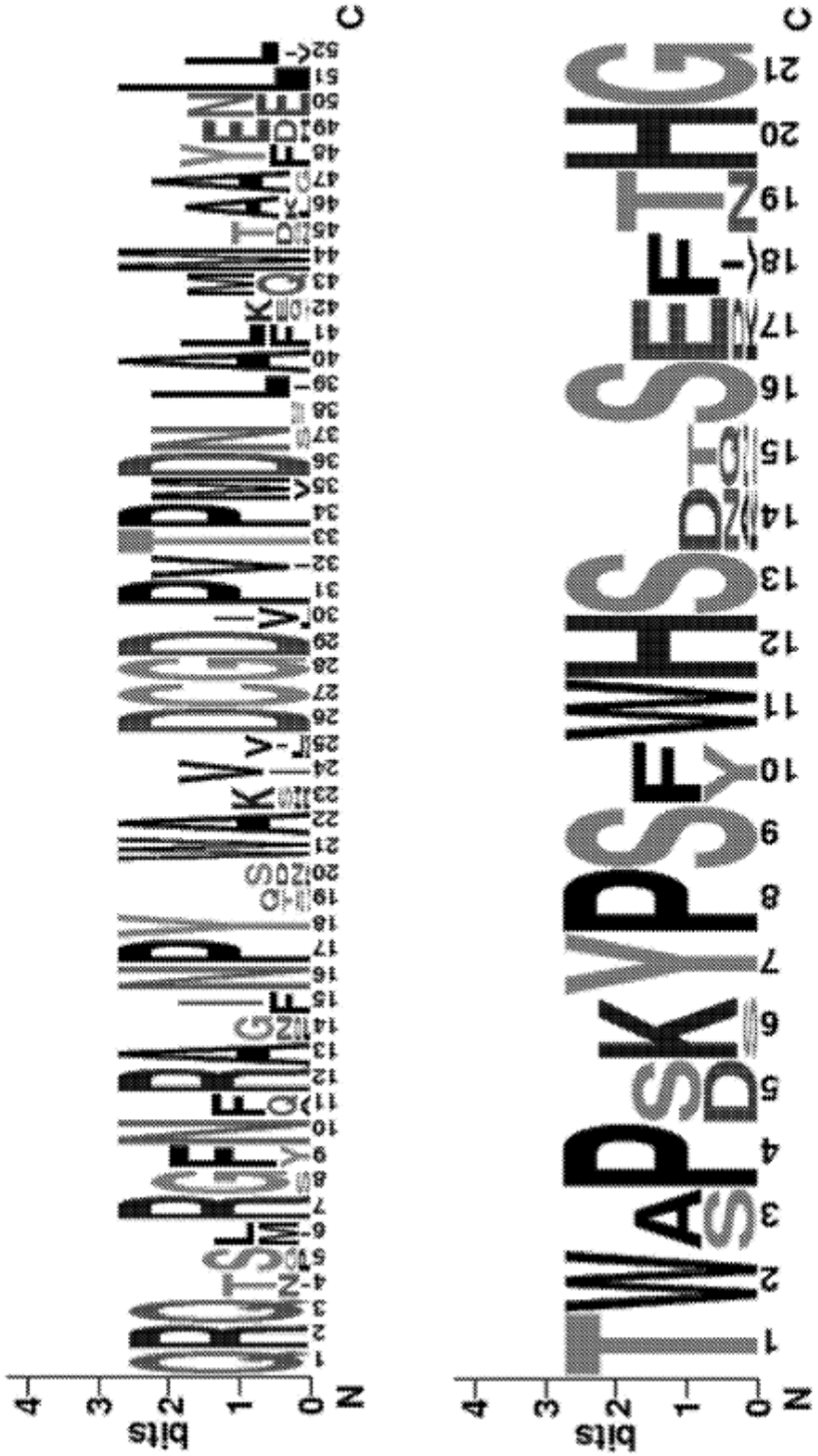

FIG. 2B

FIG. 3

```
  1 mkvagfilga liqfsltegh veqnenanlt emwgedwpfs giqtfahlph hkclidmekk
 61 fdigvigvpf dtavsfrgga rfgpqairka sqrqtsmrgf nfradinpyq dwasvvdcgd
121 vpvtpmdncl alkmmtaaye nllshesqts dnnlpprfvt lggdhsiilp alralrktyg
181 rlavihfdsh ldtwapskyp sfwhsdtsef thgsmlwiah neglltennn ihaglrtrls
241 gssfedyddd dkvgfhriea deimdggiks ivekikskip sdvpvyisvd idvldpsaap
301 gtgtmevggw mtreliriir eledlnlvga divevsppfd pteitslaga qiayelitnm
361 vkkgpidpel ikhnlelsdk ltqgqqllgf ssptdelndk iqkeqfvlqa
```

FIG. 4A

```
NP_289508          --------------MSTLGHQYDNSLVSNAFGFLRLPMNFQPYDSDADWVITGVPFDMATS
WP_000105576       --------------MSTLGHQYDNSLVSNAFGFLRLPMNFQPYDSDADWVITGVPFDMATS
YP_005016506       --------------MSTLGHQYDNSLVSNAFGFLRLPMNFMPYDSDADWVITGVPFDMATS
YP_001337000       --------------MSTLGHQYDNSLVSNAFGFLRLPMNFMPYESDADWVITGVPFDMATS
YP_003614749       --------------MSTLGHQYDNSLVSNAFGFLRLPMNFQPYDSDADWVITGVPFDMATS
YP_001455807       --------------MSTLGHQYDNSLVSNAFGFLRLPLNFQPYDSDADWVITGVPFDMATS
-P_004532666       --------------MSTLGHQYDNSLVSNAFGFLRLPLNFQPYDSDADWVITGVPFDMATS
WP_006734551       --------------MSTLGHQYDNSLVSNAFGFLRLPMNFQPYDSDADWVITGVPFDMATS
AHE29794           MNETLYGDGAIRRPSIYGSSVENT-YAGVLSFMR--RNYTRDLGGVDVAVCGVPLDLATT
                                *  *    *        * *   *        *    *** * **
Consenso                      MSTLGHQYDNSLVSNAFGFLRLPMNFQPYDSDADWVITGVPFDMATS

NP_289508          GRAGGRHGPAAIRQVSTNLAWEHNRFPWNFDMRERLNVVDCGDLVYAFGDAREMSEKLQA
WP_000105576       GRAGGRHGPAAIRQVSTNLAWEHNRFPWNFDMRERLNVVDCGDLVYAFGDAREMSEKLQA
YP_005016506       GRAGGRHGPAAIRQVSTNLAWEHNRFPWNFDMRERLNVVDCGDLVYAFGDAREMSEKLQA
YP_001337000       GRAGGRHGPAAIRQVSTNLAWEHNRFPWNFDMRERLNVVDCGDLVYAFGDAREMSEKLQA
YP_003614749       GRAGGRHGPAAIRQVSTNLAWEHNRFPWNFDMRERLNVVDCGDLVYAFGDAREMSEKLQA
YP_001455807       GRAGGRHGPAAIRQVSTNLAWEHNRFPWNFDMRERLNVVDCGDLVYAFGDAREMSEKLQA
-P_004532666       GRAGGRHGPAAIRQVSTNLAWEHYRFPWNFDMRERLNVVDCGDLVYAFGDAREMSEKLQA
WP_006734551       GRAGGRHGPAAIRQVSTNLAWEHHRFPWNFDMRERLNVVDCGDLVYAFGDAREMSEKLQA
AHE29794           FRSGARLGPAAVRAASVQLA-ELRPYPWGFDPFDDLAVIDYGDCWFDAHNPLSIKPAIVE
                    * * * **** *  *  ** *    ** **    * * * **
Consenso           GRAGGRHGPAAIRQVSTNLAWEHNRFPWNFDMRERLNVVDCGDLVYAFGDAREMSEKLQA

NP_289508          HAEKLLAAGKRMLSFGGDHFVTLPLLRAHAKHFGK-MALVHFDAHTDTYANGC--EFDHG
WP_000105576       HAEKLLASGKRMLSFGGDHFVTLPLLRAHAKHFGK-MALVHFDAHTDTYANGC--EFDHG
YP_005016506       HAEKLLAAGKRMLSFGGDHFVTLPLLRAHAKHFGK-MALVHFDAHTDTYANGC--EFDHG
YP_001337000       HAEKLLAAGKRMLSFGGDHFVTLPLLRAHAKHFGK-MALVHFDAHTDTYANGC--EFDHG
YP_003614749       HAEKLLAAGKRMLSFGGDHFVTLPLLRAHAKHFGK-MALVHFDAHTDTYANGC--EFDHG
YP_001455807       HAEKLLSAGKRMLSFGGDHFVTLPLLRAHAKHFGK-MALVHFDAHTDTYANGC--EFDHG
-P_004532666       HAERLLAAGKRMLSFGGDHFVTLPLLRAHAKHFGK-MALVHFDAHTDTYANGC--EFDHG
WP_006734551       HAEKLLSAGKRMLSFGGDHFVTLPLLRAHAKHFGK-MALVHFDAHTDTYANGC--EFDHG
AHE29794           HARTILQSGAAMLTLGGDHYITYPLLIAHAQRHGKPLSLIHFDAHCDTWADDAPDSLNHG
                   **   *  *  **  ****  * *** ***   **   * ***** ** *        **
Consenso           HAEKLLAAGKRMLSFGGDHFVTLPLLRAHAKHFGK-MALVHFDAHTDTYANGC  EFDHG
```

FIG. 4A, Continuación

```
NP_289508          TMFYTAPKEGLIDPNHSVQIGIRTEFDKDNGFTVLDACQVNDRSVDDVIAQVKQIVGDMP
WP_000105576       TMFYTAPKEGLIDPNHSVQIGIRTEFDKDNGFTVLDACQVNDRSVDDVIAQVKQIVGDMP
YP_005016506       TMFYTAPNEGLIDPNHSVQIGIRTEFDKDNGFTVLDAGQVNDRSVDDVIAQVKQIVGDMP
YP_001337000       TMFYTAPNEGLIDPNHSVQIGIRTEFDKDNGFTVLDAGQVNDRSVDDVIAQVKQIVGDMP
YP_003614749       TMFYTAPNEGLIDPNHSVQIGIRTEFDKDNGFTVLDACQVNDRGVDDIIAQVKQIVGDMP
YP_001455807       TMFYTAPKEGLIDPNHSVQIGIRTEFDKDNGFTVLDACQVNDRGVDDIIAQVNQIVGDMP
-P_004532666       TMFYTAPKEGLIDPNHSVQIGIRTEFDKDNGFTVLDACQVNDRGVDDIIAQVKQIVGDMP
WP_006734551       TMFYTAPKEGLIDPHHSVQIGIRTEFDKDNGFTVLDACQVNDRGVDDILAQVKQIVGDMP
AHE29794           TMFYKAVNEGLIDPKTSVQVGIRTWNDDYLGIHVLDAAWVHEHGPRATAERIASIVGGRP
                   **** *  ******  *** ****  *   *  ****  *              ***  *
Consenso           TMFYTAPKEGLIDPNHSVQIGIRTEFDKDNGFTVLDACQVNDRGVDDVIAQVKQIVGDMP

NP_289508          VYLTFDIDCLDPAFAPGTGTPVIGGLTSDRAIKLVRGLKDLNIVGMDVVEVAPAYDQSEI
WP_000105576       VYLTFDIDCLDPAFAPGTGTPVIGGLTSDRAIKLVRGLKDLNIVGMDVVEVAPAYDQSEI
YP_005016506       VYLTFDIDCLDPAFAPGTGTPVIGGLTSDRAIKLVRGLKDLNIVGMDVVEVAPAYDQSEI
YP_001337000       VYLTFDIDCLDPAFAPGTGTPVIGGLTSDRAIKLVRGLKDLNIVGMDVVEVAPAYDQSEI
YP_003614749       VYLTFDIDCLDPAFAPGTGTPVIGGLTSDRAIKLVRGLKDLNIVGMDVVEVAPAYDQSEI
YP_001455807       VYLTFDIDCLDPAFAPGTGTPVIGGLTSDRAIKLVRGLKDLNIVGMDVVEVAPAYDQSEI
-P_004532666       VYLTFDIDCLDPAFAPGTGTPVIGGLTSDRAIKLVRGLKDLNIVGMDVVEVAPAYDQSEI
WP_006734551       VYLTFDIDCLDPAFAPGTGTPVIGGLTSDRAIKLVRGLKDLNIVGMDVVEVAPAYDQSEI
AHE29794           AYLTFDIDCLDPAFAPGTGTPVAGGLSSAQALAIVRGLGGVNLIGADVVEVAPAYDQSEI
                    ********************** *** *  *   ****   *  * **************
Consenso           VYLTFDIDCLDPAFAPGTGTPVIGGLTSDRAIKLVRGLKDLNIVGMDVVEVAPAYDQSEI

NP_289508          TALAAATLALEMLYIQAAKKGE--
WP_000105576       TALAAATLALEMLYIQAAKKGE--
YP_005016506       TALAAATLALEMLYIQAAKKGE--
YP_001337000       TALAAATLALEMLYIQAAKKGE--
YP_003614749       TALAAATLALEMLYIQAAKKGE--
YP_001455807       TALAAATLALEMLYIQAAKKGE--
-P_004532666       TALAAATLALEMLYIQAAKKGE--
WP_006734551       TALAAATLALEMLYIQSAKKGE--
AHE29794           TAIAAAHVACDLLCLWRQRKTGAR
                   ** ***  *   *      *
Consenso           TALAAATLALEMLYIQAAKKGE
```

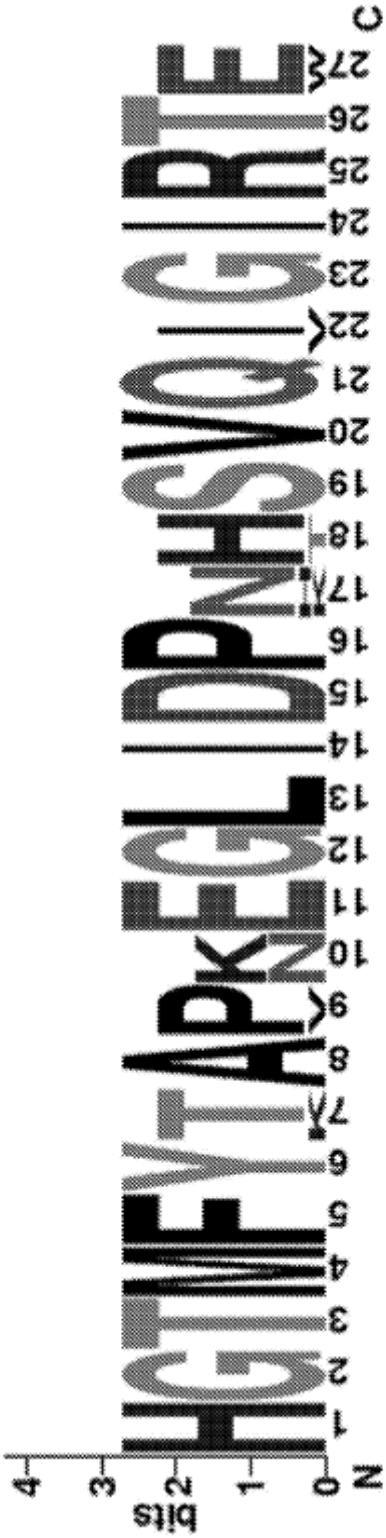

FIG. 4B

FIG. 5

```
  1 mstlghqydn slvsnafgfl rlpmnfqpyd sdadwvitgv pfdmatsgra ggrhgpaair
 61 qvstnlaweh nrfpwnfdmr erlnvvdcgd lvyafgdare mseklqahae kllaagkrml
121 sfggdhfvtl pllrahakhf gkmalvhfda htdtyangce fdhgtmfyta pkeglidpnh
181 svqigirtef dkdngftvld acqvndrsvd dviaqvkqiv gdmpvyltfd idcldpafap
241 gtgtpviggl tsdraiklvr glkdlnivgm dvvevapayd qseitalaaa tlalemlyiq
301 aakkge
```

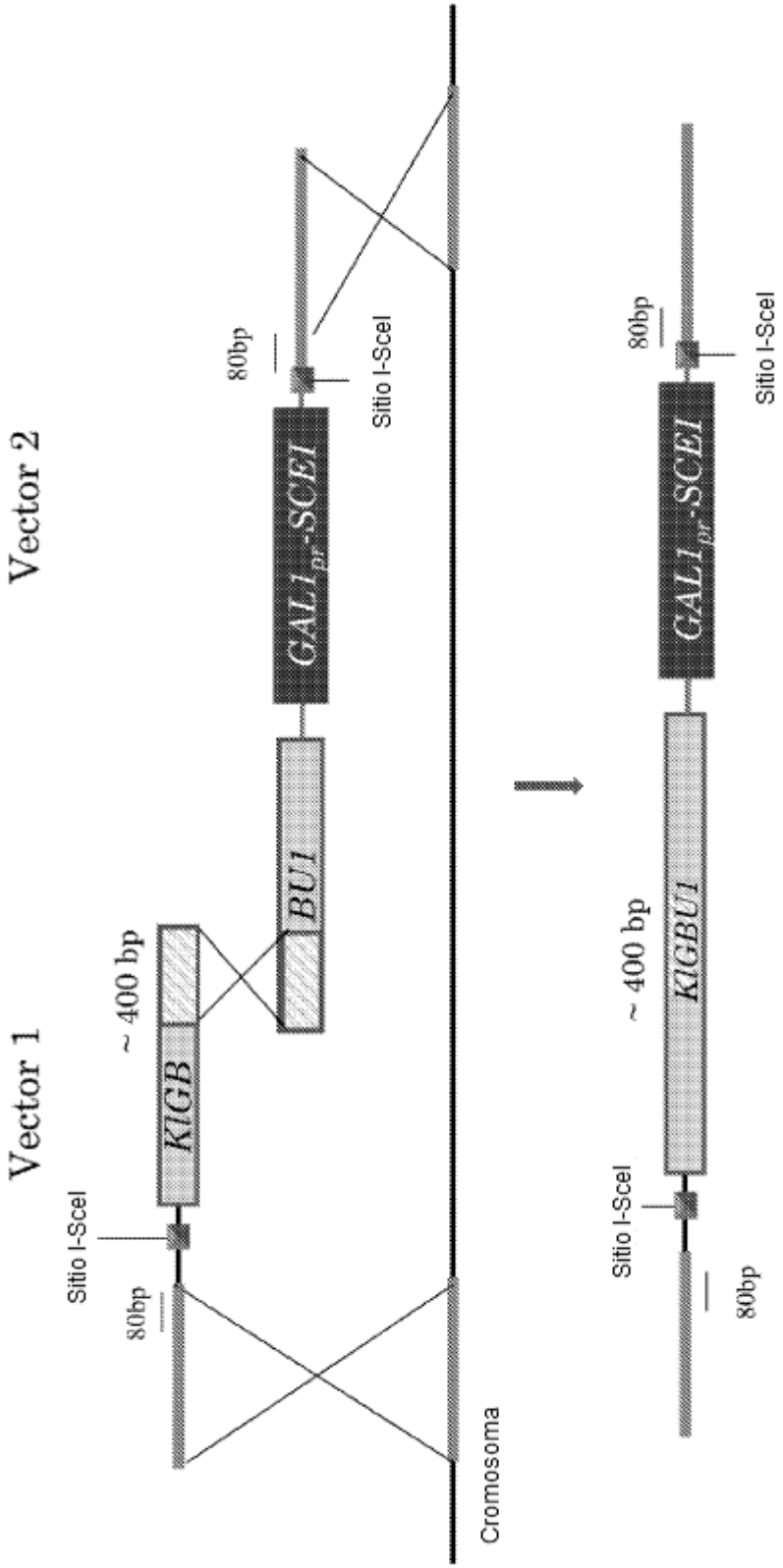
Vector 1
Vector 2
Sitio I-SceI
80bp
KIGB
~ 400 bp
BU1
GAL1pr-SCEI
80bp
Sitio I-SceI
Cromosoma
Sitio I-SceI
80bp
KIGBU1
~ 400 bp
GAL1pr-SCEI
80bp
Sitio I-SceI

FIG. 6

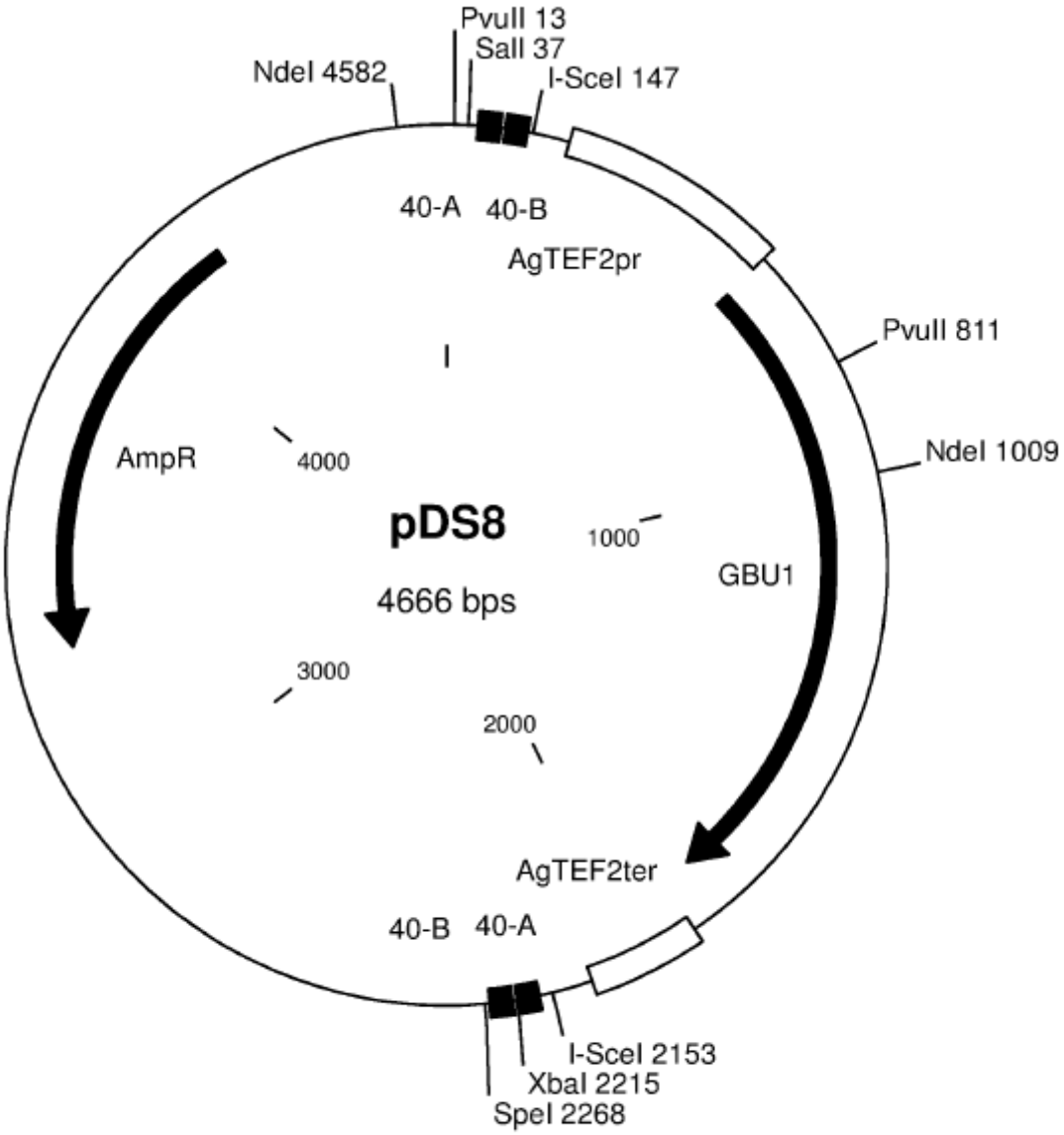
PvuII 13
SalI 37
NdeI 4582
I-SceI 147
40-A
40-B
AgTEF2pr
PvuII 811
NdeI 1009
AmpR
4000
pDS8
1000
GBU1
4666 bps
3000
2000
AgTEF2ter
40-B
40-A
I-SceI 2153
XbaI 2215
SpeI 2268

FIG. 7

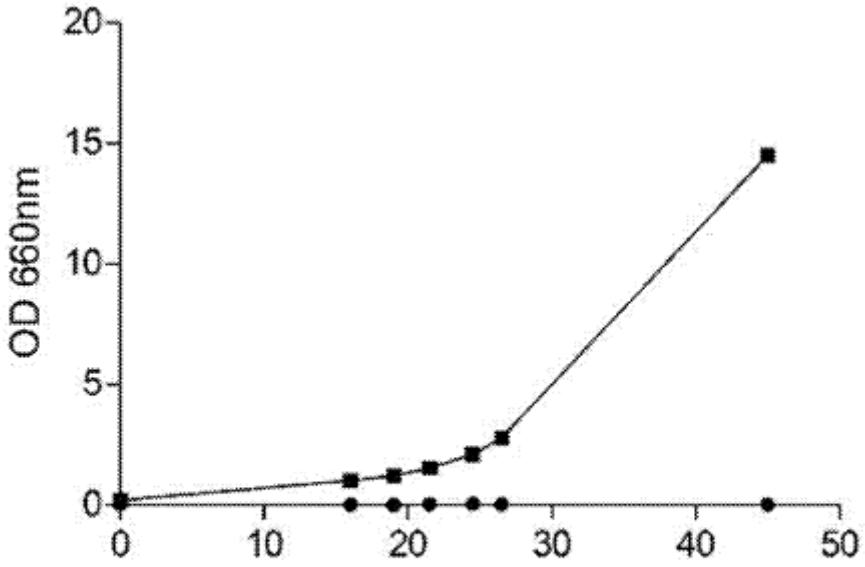

FIG. 8

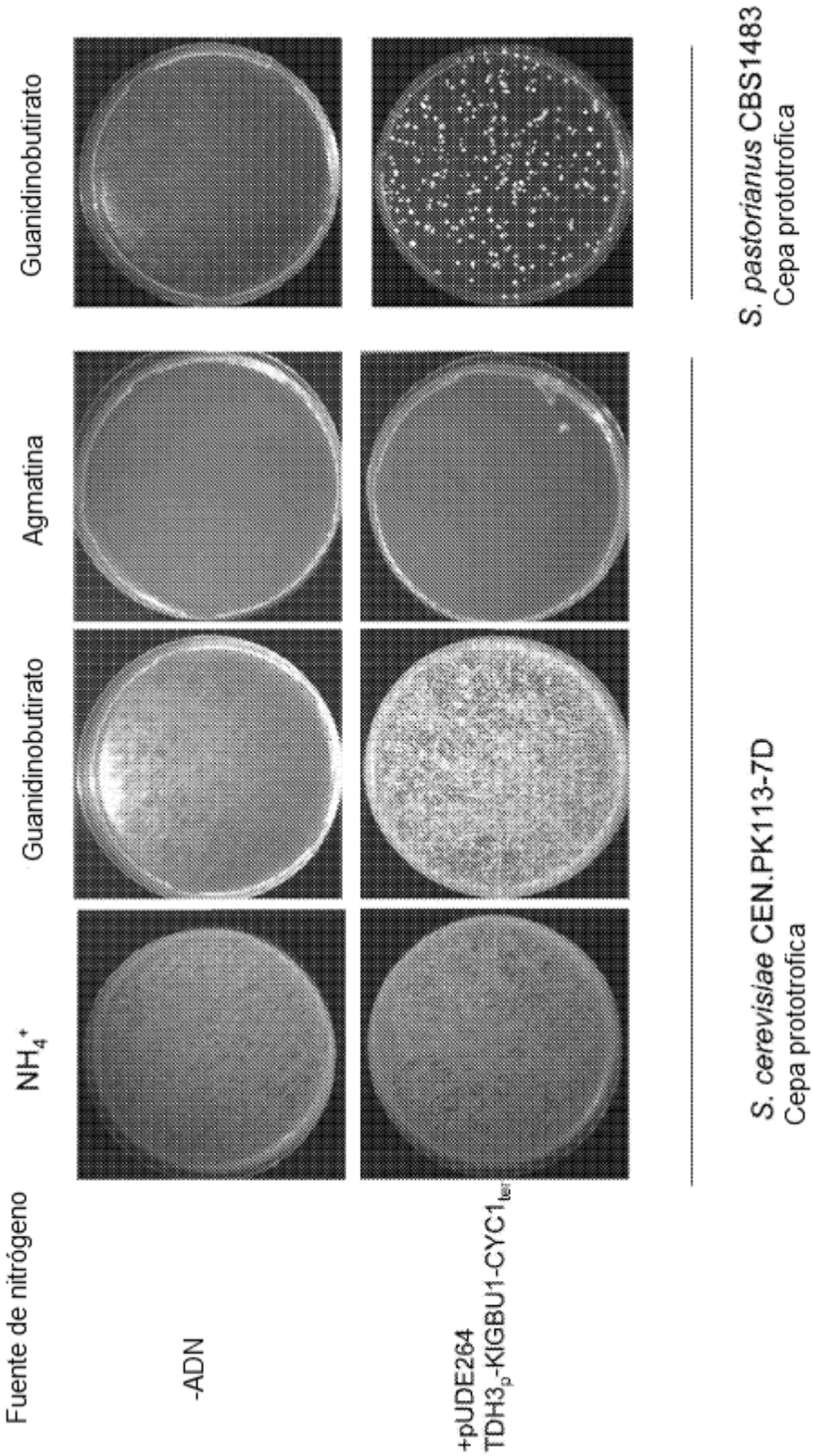
Fuente de nitrógeno
NH_4^+
Guanidinobutirato
Agmatina
Guanidinobutirato
-ADN
+pUDE264
TDH3$_p$-KIGBU1-CYC1$_{ter}$
S. cerevisiae CEN.PK113-7D
Cepa prototrofica
S. pastorianus CBS1483
Cepa prototrofica

FIG. 9

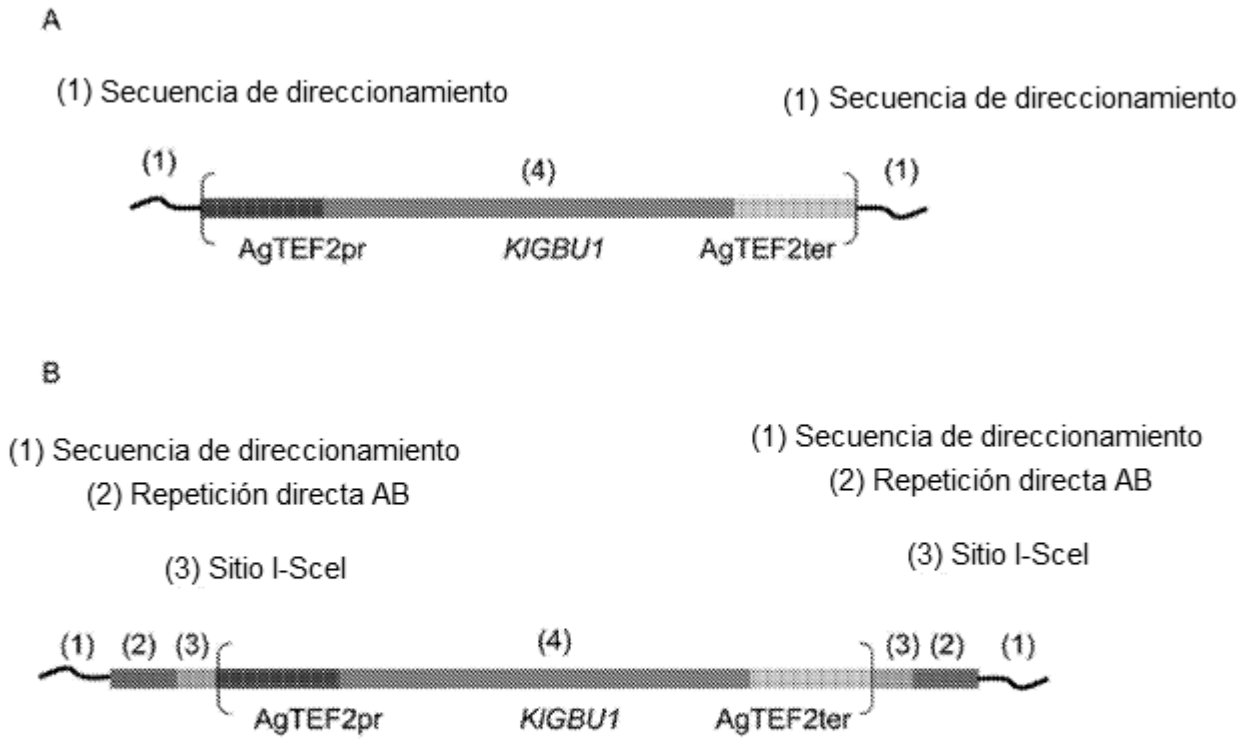
A
(1) Secuencia de direccionamiento
(1) Secuencia de direccionamiento
(1)
(4)
(1)
AgTEF2pr
KIGBU1
AgTEF2ter
B
(1) Secuencia de direccionamiento
(2) Repetición directa AB
(3) Sitio I-SceI
(1) Secuencia de direccionamiento
(2) Repetición directa AB
(3) Sitio I-SceI
(1) (2) (3)
(4)
(3) (2) (1)
AgTEF2pr
KIGBU1
AgTEF2ter

FIG. 10