

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 645 680**

51 Int. Cl.:

**C12P 7/06** (2006.01)

**C12P 7/10** (2006.01)

**C12N 9/04** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **22.11.2011 PCT/US2011/061955**

87 Fecha y número de publicación internacional: **31.05.2012 WO12071470**

96 Fecha de presentación y número de la solicitud europea: **22.11.2011 E 11843422 (4)**

97 Fecha y número de publicación de la concesión europea: **09.08.2017 EP 2643466**

54 Título: **Células de levadura y proceso para convertir acetaldehído en etanol**

30 Prioridad:

**22.11.2010 US 416169 P**

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

**07.12.2017**

73 Titular/es:

**CARGILL, INCORPORATED (50.0%)  
15407 McGinty Road West  
Wayzata, MN 55391, US y  
YI, JIAN (50.0%)**

72 Inventor/es:

**YI, JIAN y  
JESSEN, HOLLY, J.**

74 Agente/Representante:

**ELZABURU, S.L.P**

ES 2 645 680 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

Células de levadura y proceso para convertir acetaldehído en etanol

### Antecedentes

5 En años recientes se ha llevado a cabo un trabajo considerable para desarrollar métodos rentables para la generación de etanol a partir de biomasa. El uso de biomasa para generar etanol como combustible presenta varias ventajas respecto del uso de las fuentes de materias primas más tradicionales. Las materias primas potenciales son abundantes y diversas, el uso de estas materias primas no se desvía del suministro de alimentos, y exhiben potencialmente una huella de carbono menor.

10 Aunque la biomasa proporciona un sustrato atractivo para la producción de etanol, también presenta varios retos. En primer lugar, la biomasa contiene tanto celulosa, que se puede descomponer en la hexosa glucosa, como hemicelulosa, que se puede descomponer tanto en hexosas como en pentosas, tales como xilosa y arabinosa. Muchos de los microorganismos usados tradicionalmente en la fermentación etanólica son incapaces de fermentar hexosas y pentosas hasta etanol. En segundo lugar, a diferencia de las fuentes más tradicionales de materias primas para etanol (p.ej., maíz, azúcar de caña), la biomasa incluye componentes estructurales de las fuentes vegetales. Debido a que el material de la fuente es estructural y más difícil de descomponer, la biomasa requiere un mayor procesamiento para generar los monómeros de azúcares que funcionan como sustrato de fermentación. En 15 tercer lugar, el hidrolizado resultante del pretratamiento de la biomasa representa un medio adverso para los microorganismos de fermentación.

20 Varias especies de bacterias son capaces de fermentar pentosas hasta etanol, pero estas especies producen en general una mezcla de productos en vez de un único producto. Con frecuencia, uno o más de estos productos son perjudiciales para las bacterias. Además, las bacterias pueden exhibir velocidades de fermentación drásticamente reducidas en el medio adverso de un hidrolizado de materia vegetal.

25 En general, se considera que las levaduras son candidatos más atractivos que las bacterias para la fermentación etanólica a escala industrial. Sin embargo, muy pocas levaduras son capaces de fermentar pentosas hasta etanol. Se han introducido diversas modificaciones genéticas en diferentes especies de levaduras en un intento de superar este problema. Sin embargo, ninguna de estas cepas modificadas desarrolladas previamente ha demostrado ser completamente satisfactoria para la producción de etanol a gran escala a partir de biomasa. Por lo tanto, existe en la técnica la necesidad de cepas nuevas de levaduras modificadas genéticamente capaces de fermentar la biomasa hasta etanol.

30 El documento US 2010/0291653 describe las enzimas ADH1. Suwannarangsee et al, Appl. Microbiol. Biotechnol. (2010), 88:497-500 caracteriza la ADH1. El documento US 2009/0226989 describe la modificación genética de levaduras. Hughes et al, JALA (agosto de 2009) 200-212, describe cepas de *Saccharomyces* para mejorar la utilización de xilosa, y menciona la sobreexpresión de ADH1.

### Sumario

35 La invención se define en las reivindicaciones. La invención se refiere a una célula de levadura modificada genéticamente que sobreexpresa un polipéptido que comprende una secuencia de aminoácidos con al menos un 85% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°: 6, en la que dicho polipéptido es capaz de catalizar la conversión de acetaldehído hasta etanol. La invención también se refiere a una célula de levadura modificada genéticamente que sobreexpresa un primer polipéptido que comprende una secuencia de aminoácidos con al menos un 70% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°: 6 y que comprende una delección o alteración de un gen que codifica un segundo polipéptido que comprende una secuencia de aminoácidos con al menos un 70% de identidad de secuencia respecto de una secuencia de aminoácidos expuesta en SEQ ID N°: 2 o 4, en la que dicho primer polipéptido es capaz de catalizar la conversión de acetaldehído hasta etanol, y en la que dicho segundo polipéptido es capaz de catalizar la conversión de etanol hasta acetaldehído. En otros aspectos, la invención se refiere a un método para producir una célula de levadura modificada genéticamente que comprende deleccionar o alterar un gen endógeno que codifica un polipéptido que comprende una secuencia de aminoácidos con al menos un 70% de identidad de secuencia respecto de una secuencia de aminoácidos expuesta en SEQ ID N°: 2 o 4, en la que dicho polipéptido es capaz de catalizar la conversión de acetaldehído hasta etanol.

50 En la presente memoria se proporcionan polinucleótidos aislados de ADH1, ADHa, y ADHb de *I. orientalis*. Estos polinucleótidos pueden codificar la secuencia de aminoácidos expuesta en SEQ ID N°:2 (ADHa), SEQ ID N°:4 (ADHb), o SEQ ID N°:6 (ADH1). Los polinucleótidos pueden codificar una secuencia de aminoácidos con al menos alrededor del 90%, 95%, 96%, 97%, 98%, 99%, o 99,5% de identidad de secuencia respecto de las secuencias de aminoácidos expuestas en SEQ ID N°s:2, 4, o 6. Los polinucleótidos pueden codificar una secuencia de aminoácidos con menos de un 90% de identidad de secuencia respecto de las secuencias de aminoácidos expuestas en SEQ ID N°s:2, 4, o 6, en los que el polipéptido codificado tiene, sin embargo, la capacidad de catalizar la conversión de acetaldehído hasta etanol o viceversa. Los polinucleótidos pueden codificar una secuencia de aminoácidos con al menos un 70% de identidad de secuencia respecto de las secuencias de aminoácidos expuestas en SEQ ID N°s:2,

- 4, o 6. Los polinucleótidos proporcionados en la presente memoria pueden comprender la secuencia de ADN de la región codificante de SEQ ID N°:1 (ADHa), SEQ ID N°:3 (ADHb), o SEQ ID N°:5 (ADH1). Los polinucleótidos pueden comprender una secuencia de ADN con al menos alrededor del 90%, 95%, 96%, 97%, 98%, 99%, o 99,5% de identidad de secuencia respecto de la región codificante de las secuencias de ADN expuestas en SEQ ID N°s:1, 3, o 5.
- 5 Los polinucleótidos proporcionados en la presente memoria pueden comprender una secuencia de ADN con menos de un 90% de identidad de secuencia respecto de la región codificante de SEQ ID N°s: 1, 3, o 5, pero sin embargo codifican un polipéptido con la capacidad de catalizar la conversión de acetaldehído hasta etanol o viceversa. También se proporcionan en la presente memoria vectores que comprenden los polinucleótidos proporcionados en la presente memoria, así como células hospedadoras que comprenden estos vectores.
- 10 En la presente memoria se proporcionan polipéptidos aislados de ADH1, ADHa, y ADHb de *I. orientalis*. Estos polipéptidos pueden comprender la secuencia de aminoácidos expuesta en SEQ ID N°:2 (ADHa), SEQ ID N°:4 (ADHb), o SEQ ID N°:6 (ADH1). Los polipéptidos pueden comprender una secuencia de aminoácidos con al menos alrededor del 90%, 95%, 96%, 97%, 98%, 99%, o 99,5% de identidad de secuencia respecto de las secuencias de aminoácidos expuestas en SEQ ID N°s:2, 4, o 6. Los polipéptidos proporcionados en la presente memoria pueden comprender una secuencia de aminoácidos con menos de un 90% de identidad de secuencia respecto de las secuencias de aminoácidos expuestas en SEQ ID N°s:2, 4, o 6, pero sin embargo tienen la capacidad de catalizar la conversión de acetaldehído hasta etanol y viceversa.

- En la presente memoria se proporcionan en ciertas realizaciones métodos para sobreexpresar ADH1 de *I. orientalis* en una célula de levadura introduciendo uno o más polinucleótidos de ADH1 de *I. orientalis*. De forma similar, en la presente memoria se proporcionan en ciertas realizaciones células de levadura modificadas genéticamente que sobreexpresan un polipéptido de ADH1 de *I. orientalis*. En ciertas realizaciones, estas células de levadura comprenden un polinucleótido que codifica la secuencia de aminoácidos expuesta en SEQ ID N°:6. En otras realizaciones, las células de levadura comprenden un polinucleótido que codifica una secuencia de aminoácidos con al menos alrededor de un 90%, 95%, 96%, 97%, 98%, 99%, o 99,5% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°:6. En otras realizaciones, las células de levadura comprenden un polinucleótido que codifica una secuencia de aminoácidos con menos de un 90% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°:6, en las que el polipéptido codificado sin embargo tiene la capacidad de catalizar la conversión de etanol hasta acetaldehído. En ciertas realizaciones, las células de levadura comprenden un polinucleótido que comprende la secuencia de ADN de la región codificante de SEQ ID N°:5. En otras realizaciones, las células de levadura comprenden un polinucleótido que comprende una secuencia de ADN con al menos alrededor de un 90%, 95%, 96%, 97%, 98%, 99%, o 99,5% de identidad de secuencia respecto de la región codificante de la secuencia de ADN expuesta en SEQ ID N°:5. En otras realizaciones, las células de levadura comprenden un polinucleótido que comprende una secuencia de ADN con menos de un 90% de identidad de secuencia respecto de la región codificante de SEQ ID N°:5, pero que sin embargo codifica un polipéptido con la capacidad de catalizar la conversión de etanol hasta acetaldehído. En ciertas realizaciones, se puede obtener la sobreexpresión de ADH1 a través de la introducción de uno o más genes de ADH1 exógenos, la expresión incrementada de uno o más genes de ADH1 endógenos, o una combinación de las mismas.

- En la presente memoria se proporcionan en ciertas realizaciones métodos para disminuir la expresión de ADHa y/o ADHb de *I. orientalis* en una célula de levadura delecionando o alterando uno o más genes de ADHa y/o ADHb endógenos de *I. orientalis*. De forma similar, en la presente memoria se proporcionan en ciertas realizaciones células de levadura modificadas genéticamente que comprenden una delección o alteración de uno o más genes de ADHa y/o ADHb de *I. orientalis*. En ciertas realizaciones, estas células de levadura comprenden una delección o alteración de un polinucleótido que codifica la secuencia de aminoácidos expuesta en SEQ ID N°:2 (ADHa) o SEQ ID N°:4 (ADHb). En otras realizaciones, las células de levadura comprenden una delección o alteración de un polinucleótido que codifica una secuencia de aminoácidos con al menos alrededor de un 90%, 95%, 96%, 97%, 98%, 99%, o 99,5% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°s:2 o 4. En otras realizaciones, las células de levadura comprenden una delección o alteración de un polinucleótido que codifica una secuencia de aminoácidos con menos de un 90% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°s:2 o 4, en las que el polipéptido codificado sin embargo tiene la capacidad de catalizar la conversión de acetaldehído hasta etanol. En ciertas realizaciones, las células de levadura comprenden una delección o alteración de un polinucleótido que comprende la secuencia de ADN de la región codificante de SEQ ID N°s:1 o 3. En otras realizaciones, las células de levadura comprenden una delección o alteración de un polinucleótido que comprende una secuencia de ADN con al menos alrededor de un 90%, 95%, 96%, 97%, 98%, 99%, o 99,5% de identidad de secuencia respecto de la región codificante de las secuencias de ADN expuestas en SEQ ID N°s:1 o 3. En otras realizaciones, las células de levadura comprenden una delección o alteración de un polinucleótido que comprende una secuencia de ADN con menos de un 90% de identidad de secuencia respecto de la región codificante de SEQ ID N°:1 o 3, pero que sin embargo codifica un polipéptido con la capacidad de catalizar la conversión de acetaldehído hasta etanol. En ciertas realizaciones, la delección o alteración de uno o más genes de ADHa y/o ADHb se puede acoplar con la introducción de uno o más genes exógenos.

- 60 En la presente memoria se proporcionan en ciertas realizaciones células de levadura modificadas genéticamente que sobreexpresan un polipéptido de ADH1 de *I. orientalis* y que comprenden una delección o alteración de uno o más genes de ADHa y/o ADHb de *I. orientalis*.

5 En ciertas realizaciones de las células de levadura modificadas genéticamente proporcionadas en la presente memoria, las células de levadura pertenecen al clado *I. orientalis/P. fermentans*. En algunas de estas realizaciones, las células de levadura son *I. orientalis*. En ciertas realizaciones, las células de levadura pueden haber experimentado mutación y/o selección antes, durante, o después de la introducción de las modificaciones genéticas relacionadas con la sobreexpresión de ADH1 y/o la delección/alteración de ADHa/ADHb. En algunas de estas realizaciones, las células de levadura pueden exhibir un cierto grado de tolerancia hacia etanol, ácidos orgánicos, otros productos o subproductos de la fermentación, y/o diversos componentes de los medios que es mayor que la exhibida por las células de levadura de tipo natural de la misma especie.

10 En la presente memoria se proporcionan en ciertas realizaciones procesos de fermentación en los que las células de levadura modificadas genéticamente proporcionadas en la presente memoria se cultivan en un medio de fermentación que contiene xilosa. En algunas de estas realizaciones, el medio de fermentación contiene al menos alrededor de 10 g/L, 20 g/L, 30 g/L, 40 g/L, 50 g/L, 75 g/L, 100 g/L, o 125 g/L de xilosa. En ciertas realizaciones, la xilosa del medio de fermentación se obtiene de un hidrolizado de biomasa vegetal.

15 En la presente memoria se proporcionan en ciertas realizaciones métodos para producir etanol mediante el uso de las células de levadura modificadas genéticamente proporcionadas en la presente memoria. En algunas de estas realizaciones, las células se cultivan en un medio que contiene xilosa, y en algunas de estas realizaciones el medio contiene al menos alrededor de 10 g/L, 20 g/L, 30 g/L, 40 g/L, 50 g/L, 75 g/L, 100 g/L, o 125 g/L de xilosa. En ciertas realizaciones, la xilosa del medio se obtiene de un hidrolizado de biomasa vegetal.

### Breve descripción de las figuras

20 Figura 1: Rutas de fermentación de xilosa y arabinosa en levadura.

Figura 2: Ruta en levadura para la conversión de piruvato en etanol.

Figura 3: Funcionamiento de la cepa 3416 de delección de ADHa, la cepa 3859 de delección de ADHb, la cepa 3860 de delección de ADHa/ADHb, y la cepa original 3356 en medio definido con 20 g/L de dextrosa y 80 g/L de xilosa a pH 4,8.

25 Figura 4: Funcionamiento de la cepa 3416 de delección de ADHa, la cepa 3859 de delección de ADHb, la cepa 3860 de delección de ADHa/ADHb, y la cepa original 3356 en medio definido con 20 g/L de dextrosa y 80 g/L de xilosa a pH 5,1.

Figura 5: Funcionamiento de la cepa 3416 de delección de ADHa y la cepa 3489 de sobreexpresión de ADH1/delección de ADHa en medio definido con 20 g/L de dextrosa y 80 g/L de xilosa a pH 4,8.

30 Figura 6: Funcionamiento de la cepa 3416 de delección de ADHa y la cepa original 3356 en medio DMDX de CSH del 30% a pH 5,8.

Figura 7: Funcionamiento de la cepa 3416 de delección de ADHa y la cepa 3489 de sobreexpresión de ADH1/delección de ADHa en medio DMDX de CSH del 30% a pH 5,8.

35 Figura 8: Funcionamiento de las cepas 3489 y 4138 de sobreexpresión de ADH1/delección de ADHa y de las cepas 3922 y 12053 de sobreexpresión de ADH1/delección de ADHa/ADHb en medio CSH a pH 5,0.

Figura 9: Funcionamiento de la cepa 3416 de delección de ADHa y su cepa original 3356 en medio definido con 20 g/L de dextrosa, 80 g/L de xilosa, 10 g/L de arabinosa, y 10 g/L de acetato a pH 4,95.

Figura 10: Funcionamiento de la cepa 3416 de delección de ADHa y su cepa original 3356 en medio definido con 20 g/L de dextrosa, 80 g/L de xilosa, 10 g/L de arabinosa, y 10 g/L de acetato a pH 5,8.

40 Figura 11: Funcionamiento de la cepa 3489 de sobreexpresión de ADH1/delección de ADHa, su cepa 3416 de delección de ADHa original, y la cepa 3863 de control de sitio de inserción en medio definido con 20 g/L de dextrosa, 80 g/L de xilosa, 10 g/L de arabinosa, y 10 g/L de acetato a pH 4,95.

45 Figura 12: Funcionamiento de la cepa 3489 de sobreexpresión de ADH1/delección de ADHa, su cepa 3416 de delección de ADHa original, y la cepa 3863 de control de sitio de inserción en medio definido con 20 g/L de dextrosa, 80 g/L de xilosa, 10 g/L de arabinosa, y 10 g/L de acetato a pH 5,8.

Figura 13: Funcionamiento de la cepa 3859 de delección de ADHb y su cepa original 3356 en medio definido con 20 g/L de dextrosa, 80 g/L de xilosa, 10 g/L de arabinosa, y 10 g/L de acetato a pH 4,95.

Figura 14: Funcionamiento de la cepa 3859 de delección de ADHb y su cepa original 3356 en medio definido con 20 g/L de dextrosa, 80 g/L de xilosa, 10 g/L de arabinosa, y 10 g/L de acetato a pH 5,8.

50 Figura 15: Alineación de secuencias de aminoácidos de ADH1, ADH2, y ADH3 de *S. cerevisiae* con S141G2556, S141G9091, y S141G1202.

## Descripción detallada

Abreviaturas:

ADH, alcohol deshidrogenasa; ALD, acetaldehído deshidrogenasa; CSH, hidrolizado de rastrojo de maíz; DM, medio definido; D5P, D-xilulosa 5-fosfato; F6P, fructosa 6-fosfato; G3P, gliceraldehído 3-fosfato; HMF, hidroximetil furfural; ORF, marco de lectura abierto; OUR, velocidad de absorción de oxígeno; PPP, ruta de pentosa fosfato; RKI, ribosa-5-fosfato cetol-isomerasa; RPE, D-ribulosa-5-fosfato 3-epimerasa; TAL, transaldolasa; TKL, transcetolasa; XDH, xilitol deshidrogenasa; XK, xiluloquinasa; XR, xilosa reductasa; YP, extracto de levadura/peptona.

La especie de levadura ideal para la producción de etanol a escala industrial a partir de biomasa debería exhibir resistencia a medios de pH bajo, capacidad de fermentar tanto hexosas como pentosas hasta etanol, y resistencia hacia los compuestos inhibitorios presentes en un hidrolizado de materia vegetal y que se generan de la fermentación, lo que incluye acetato, hidroximetil furfural (HMF), furfural, productos fenólicos, aldehídos, cetonas, y el propio etanol.

*S. cerevisiae* y la mayoría de las demás especies de levaduras son capaces de fermentar hexosas hasta etanol. Sin embargo, la mayoría de las especies de levaduras son incapaces de metabolizar pentosas. Las que son capaces de metabolizar pentosas lo hacen por medio de una ruta compleja no fermentativa. Por ejemplo, las especies de levaduras que metabolizan xilosa, el azúcar predominante en la biomasa, reducen D-xilosa hasta xilitol mediante el uso de la xilosa reductasa (XR). El xilitol se oxida hasta D-xilulosa mediante la xilitol deshidrogenasa (XDH), y la D-xilulosa se fosforila mediante xiluloquinasa (XK) para producir D-xilulosa 5-fosfato (D5P). Esta ruta se ilustra en la Figura 1. La D5P resultante entra en la ruta de pentosa fosfato (PPP), que genera fructosa 6-fosfato (F6P) y gliceraldehído 3-fosfato (G3P), las cuales entran en el ciclo glicolítico.

El piruvato que se genera de la glicolisis se convierte en acetaldehído y CO<sub>2</sub> mediante la piruvato descarboxilasa. El acetaldehído resultante se puede reducir hasta etanol mediante la alcohol deshidrogenasa (ADH) o se puede convertir en ácido acético mediante la acetaldehído deshidrogenasa (ALD) (Figura 2).

La ruta de xilosa en levaduras es ineficaz, porque genera un desequilibrio redox. La conversión de xilosa en xilitol usa NADPH como cofactor, mientras la etapa de xilitol a xilulosa produce NADH. En condiciones anaerobias, se produce más NADH del que se puede reciclar, y se acumula xilitol. Los primeros intentos de modificar genéticamente levaduras para fermentar xilosa hasta etanol de manera más eficaz utilizaron genes de XR y XDH exógenos (documento WO95/13362; documento WO97/42307). Sin embargo, estos organismos modificados no produjeron etanol de manera eficaz. Los intentos posteriores intentaron evitar completamente el intermedio xilitol introduciendo un gen exógeno de D-xilosa isomerasa (XI) y delecionando XR y/o XDH (documento WO04/099381). XI convierte xilosa directamente en xilulosa, lo que evita la generación de un desequilibrio redox. Las rutas que utilizan XI para metabolizar xilosa son habituales en bacterias, pero infrecuentes en levaduras. Una *K. marxianus* modificada genéticamente que expresaba XI exógena de *Piromyces* y que sobreexpresaba XK, y con delecciones de los genes de XR y XDH endógenos, exhibió una utilización incrementada de xilosa y una producción incrementada de etanol (documento WO04/099381).

En *Saccharomyces*, la enzima principal para la producción de etanol a partir de acetaldehído es ADH1. La reacción inversa de etanol de vuelta a acetaldehído la cataliza principalmente la ADH2, que tiene una afinidad mayor hacia etanol que las otras ADHs, y es importante en el uso de etanol como fuente de carbono. Se ha informado previamente que la ADH1 se reprime transcripcionalmente en *Saccharomyces* en ausencia de una fuente de carbono fermentable, mientras la ADH2 se reprime por la glucosa (Denis J Biol Chem 258:1165 (1983)). Se han identificado los genes de tres ADHs adicionales (ADH3, ADH4, y ADH5) implicadas en el metabolismo de etanol en *Saccharomyces*, pero se desconocen sus papeles exactos.

La función y regulación de las ADHs en las especies de levaduras no está conservada. En *Kluyveromyces lactis*, se han identificado cuatro genes de ADH. Dos de estos genes de ADH son activos en el citoplasma, mientras los otros dos son activos en las mitocondrias. Se ha demostrado que una de las ADHs mitocondriales se induce por etanol en vez de reprimirse por glucosa, lo que se aproxima a la expresión constitutiva en las cepas de fermentación. En *Pichia stipitis*, se han caracterizado dos ADHs citoplasmáticas. La expresión de ADH1 de *P. stipitis* parece inducirse aproximadamente 10 veces por la limitación de oxígeno. Aunque la expresión de ADH2 de *P. stipitis* fue baja tanto en condiciones de limitación de oxígeno como en condiciones completamente aerobias, se incrementó mediante la alteración de ADH1, lo que indica una retroregulación de ADH2. Se han identificado tres ADHs citoplasmáticas en *Candida maltosa*, dos de las cuales (ADH2a y ADH2b) están localizadas en tándem entre sí en el genoma. La ADH1 de *C. maltosa* es responsable de la producción de etanol a partir de glucosa, mientras la ADH2a se reprime por glucosa. Sin embargo, estas dos enzimas funcionaron en la producción de etanol a partir de xilosa. La ADH2b de *C. maltosa* se expresa a un nivel inferior, y todavía está por determinar su función completa.

Tal como se describe en la presente memoria, se han identificado y caracterizado tres genes de ADH de *I. orientalis*. Los primeros dos genes, denominados en la presente memoria ADHa y ADHb, codifican proteínas ADH que se expresan a un nivel inferior en condiciones de glucosa que la enzima fermentativa principal de *I. orientalis*, y exhiben propiedades similares a ADH2 en ciertas condiciones, pero no en todas. Las secuencias de ADN de ADHa y ADHb

se exponen en SEQ ID N°s:1 y 3, respectivamente. La región codificante de ADHa (nucleótidos 1052 a 2182 en SEQ ID N°:1) codifica el polipéptido de ADHa expuesto en SEQ ID N°:2, mientras la región codificante de ADHb (nucleótidos 1001 a 2134 en SEQ ID N°:3) codifica el polipéptido de ADHb expuesto en SEQ ID N°:4. Los resultados experimentales proporcionados en la presente memoria establecen que la inactivación de la expresión de ADHa y/o ADHb incrementa el título de etanol y el consumo de xilosa en *I. orientalis* en los medios que contienen xilosa. El tercer gen de ADH descrito en la presente memoria, ADH1, es comparable funcionalmente con la ADH1 de *S. cerevisiae*. La secuencia de ADN de la región codificante de ADH1 se expone en SEQ ID N°:5, y la secuencia de aminoácidos del polipéptido codificado se expone en SEQ ID N°:6. Los resultados experimentales proporcionados en la presente memoria establecen que la sobreexpresión de ADH1 incrementa el título de etanol y el consumo de xilosa en *I. orientalis* en muchas condiciones. Por lo tanto, en la presente memoria se proporcionan polinucleótidos y polipéptidos de ADH1, ADHa, y ADHb, así como vectores que comprenden polinucleótidos de ADH1, ADHa, y/o ADHb, células hospedadoras que comprenden estos vectores, y métodos para expresar ADH1, ADHa, y/o ADHb a partir de estas células hospedadoras.

En la presente memoria se proporcionan polinucleótidos aislados de ADHa, ADHb, y ADH1. Estos polinucleótidos aislados pueden comprender una región codificante que codifica un polipéptido que tiene la secuencia de aminoácidos expuesta en SEQ ID N°:2, 4, o 6. Los polinucleótidos pueden comprender la región codificante de la secuencia de nucleótidos expuesta en SEQ ID N°s:1, 3, o 5. Los polinucleótidos pueden comprender una secuencia de nucleótidos con al menos un 90% de identidad de secuencia respecto de la región codificante de la secuencia de nucleótidos expuesta en SEQ ID N°s:1, 3, o 5. Los polinucleótidos pueden comprender una secuencia de nucleótidos que tiene al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98%, al menos un 99%, o al menos un 99,5% de identidad de secuencia respecto de la región codificante de la secuencia de nucleótidos expuesta en SEQ ID N°s:1, 3, o 5.

Los porcentajes de identidad de secuencia para las secuencias de nucleótidos o aminoácidos se pueden calcular mediante métodos conocidos en la técnica, tal como, por ejemplo, mediante el uso del programa informático BLAST (Basic Local Alignment Search Tool del National Center for Biological Information (NCBI)) versión 2.2.1 con los parámetros por defecto. Se considera que las secuencias que tienen un índice de identidad de al menos un 90%, mediante el uso del algoritmo BLAST versión 2.2.1 con los parámetros por defecto, tienen al menos un 90% de identidad de secuencia. El programa informático BLAST está disponible del NCBI, Bethesda, Maryland.

Los polinucleótidos aislados proporcionados en la presente memoria pueden comprender una región codificante que codifica un polipéptido que comprende una secuencia de aminoácidos con al menos un 90% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°s:2, 4, o 6. El polipéptido codificado puede comprender una secuencia de aminoácidos con al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98%, al menos un 99%, o al menos un 99,5% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°s:2, 4, o 6. Los polinucleótidos aislados pueden comprender una secuencia de nucleótidos con al menos un 90% de identidad de secuencia respecto de la región codificante de la secuencia de nucleótidos expuesta en SEQ ID N°s:1, 3, o 5. Los polinucleótidos aislados pueden comprender una secuencia de nucleótidos que tiene al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98%, al menos un 99%, o al menos un 99,5% de identidad de secuencia respecto de la región codificante de la secuencia de nucleótidos expuesta en SEQ ID N°s:1, 3, o 5.

Se proporcionan polinucleótidos aislados que pueden comprender una región codificante que codifica un polipéptido con un 70% o más de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°s:2, 4, o 6, en los que el polipéptido es capaz de catalizar la conversión de etanol hasta acetaldehído o viceversa. Tal como se usa en la presente memoria, se considera que un polipéptido tiene la capacidad de catalizar la conversión de acetaldehído hasta etanol si una célula de levadura de ensayo que sobreexpresa el polipéptido tiene al menos un 105% del incremento máximo del título de etanol durante el consumo de 20 g/L o más de xilosa en ausencia de glucosa en comparación con una célula de levadura de control, en el que la célula de levadura de control es genéticamente idéntica a la célula de levadura de ensayo, excepto por la expresión nativa del polipéptido. De forma similar, se considera que un polipéptido tiene la capacidad de catalizar la conversión de etanol hasta acetaldehído si una célula de levadura de ensayo con una delección del gen que codifica el polipéptido tiene al menos un 105% del incremento máximo del título de etanol durante el consumo de 20 g/L o más de xilosa en ausencia de glucosa en comparación con una célula de levadura de control, en el que la célula de levadura de control es genéticamente idéntica a la célula de levadura de ensayo pero sin la delección del gen que codifica el polipéptido. En un protocolo ejemplar para establecer si una célula de levadura de ensayo tiene al menos un 105% del incremento máximo del título de etanol durante el consumo de 20 g/L o más de xilosa en ausencia de glucosa frente a una célula de control, se usan cultivos nocturnos en YPD de las células de ensayo y de control para inocular 50 mL de medios YP que contienen 20 g/L de dextrosa y 80 g/L de xilosa a pH 4-6 en un matraz con deflectores de 125 ml a una DO<sub>600</sub> inicial de 0,2 en un espectrofotómetro modelo DU600 (Beckman Coulter) con una longitud de trayectoria de 1 cm. Las células se incuban a 30-37 °C y 100 rpm hasta que se agota la dextrosa y, tras el agotamiento de la dextrosa, se consumen al menos 20 g/L de xilosa.

Los polinucleótidos pueden comprender una región codificante que codifica un polipéptido con al menos un 70%, al menos un 75%, al menos un 80%, al menos un 85%, o al menos un 90% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°s:2 o 4, en el que el polipéptido es capaz de catalizar la

conversión de etanol hasta acetaldehído. Los polinucleótidos pueden comprender una secuencia de nucleótidos con al menos un 70%, al menos un 75%, al menos un 80%, al menos un 85%, o al menos un 90% de identidad de secuencia respecto de la región codificante de la secuencia de nucleótidos expuesta en SEQ ID N°s:1 o 3. Los polinucleótidos pueden comprender una región codificante que codifica un polipéptido con al menos un 70%, al menos un 75%, al menos un 80%, al menos un 85%, o al menos un 90% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°:6, en el que el polipéptido es capaz de catalizar la conversión de acetaldehído hasta etanol. El polinucleótido puede comprender una secuencia de nucleótidos con al menos un 70%, al menos un 75%, al menos un 80%, al menos un 85%, o al menos un 90% de identidad de secuencia respecto de la región codificante de la secuencia de nucleótidos expuesta en SEQ ID N°:5.

En la presente memoria se proporcionan construcciones que comprenden uno o más de los polinucleótidos aislados proporcionados en la presente memoria. El término "construcción", tal como se usa en la presente memoria, se refiere a una secuencia de ADN que se usa para transformar una célula. La construcción puede ser, por ejemplo, un plásmido o vector circular, una porción de un plásmido o vector circular (tal como un producto de digestión de una enzima de restricción), un plásmido o vector linealizado, o un producto de PCR preparado mediante el uso de un plásmido o vector como molde. Además de uno o más de los polinucleótidos proporcionados en la presente memoria, una construcción puede comprender uno o más elementos reguladores (p.ej., promotores, terminadores) unidos de forma operable a la secuencia polinucleotídica. La construcción puede comprender además uno o más componentes adicionales, que incluyen, por ejemplo, uno o más sitios de restricción y/o uno o más genes marcadores de selección, unidos opcionalmente a uno o más elementos reguladores. Un "gen marcador de selección" es un gen que codifica una proteína necesaria para la supervivencia y/o el crecimiento de la célula transformada en un medio de cultivo selectivo, y por lo tanto se puede usar para aplicar una presión de selección a la célula.

Tal como se usa en la presente memoria, el término "promotor" se refiere a una secuencia sin traducir localizada antes (es decir, en 5') del codón de inicio de la traducción de un gen (en general dentro de alrededor de 1 a 1000 pares de bases (pb), preferiblemente dentro de alrededor de 1 a 500 pb) que controla el inicio de la transcripción del gen. El término "terminador", tal como se usa en la presente memoria, se refiere a una secuencia sin traducir localizada después (es decir, en 3') del codón de terminación de la traducción de un gen (en general dentro de alrededor de 1 a 500 pb, preferiblemente dentro de alrededor de 1 a 300 pb, y especialmente dentro de alrededor de 1 a 100 pb) que controla el final de la transcripción del gen. Un promotor o terminador está "unido de forma operable" a un gen si su posición en el genoma respecto de la del gen es tal que el promotor o terminador, según el caso, lleva a cabo su función de control transcripcional. Los promotores y terminadores adecuados se describen, por ejemplo, en los documentos WO99/14335, WO00/71738, WO02/42471, WO03/102201, WO03/102152 y WO03/049525.

También se proporcionan en la presente memoria células hospedadoras que se han transformado con una o más de las construcciones proporcionadas en la presente memoria, así como métodos para expresar ADHa, ADHb, y/o ADH1 a partir de estas células hospedadoras. En algunas de estas realizaciones, las células hospedadoras son células de levadura o células bacterianas. En algunas de esas realizaciones en las que las células hospedadoras son células de levadura, las células de levadura son células de levadura Crabtree-negativas, y en algunas de estas realizaciones las células de levadura pertenecen a los géneros *Candida* o *Issatchenkia*.

En la presente memoria se proporcionan polipéptidos aislados de ADHa, ADHb, y ADH1. Estos polipéptidos pueden comprender la secuencia de aminoácidos expuesta en SEQ ID N°s:2, 4, o 6. Los polipéptidos pueden comprender una secuencia de aminoácidos con al menos un 90% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°s:2, 4, o 6. Los polipéptidos pueden comprender una secuencia de aminoácidos con al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98%, al menos un 99%, o al menos un 99,5% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°s:2, 4, o 6. Los polipéptidos pueden comprender una secuencia de aminoácidos con al menos un 70% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°s:2, 4, o 6, y también tienen la capacidad de catalizar la conversión *in vitro* de etanol hasta acetaldehído o viceversa. Los polipéptidos pueden comprender una secuencia de aminoácidos con al menos un 70% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°s:2 o 4, y son capaces de catalizar la conversión de etanol hasta acetaldehído. Los polipéptidos pueden comprender una secuencia de aminoácidos con al menos un 75%, al menos un 80%, al menos un 85%, o al menos un 90% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°s:2 o 4. Los polipéptidos pueden comprender una secuencia de aminoácidos con al menos un 70% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°:6, y son capaces de catalizar la conversión de acetaldehído hasta etanol. Los polipéptidos pueden comprender una secuencia de aminoácidos con al menos un 75%, al menos un 80%, al menos un 85%, o al menos un 90% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°:6.

Tal como se describe en la presente memoria, la delección o alteración de los genes de ADHa y/o ADHb en *I. orientalis* dio como resultado una cepa de levadura con una utilización de xilosa y un título de etanol incrementados frente a las cepas originales, tanto en un medio sintético como en un hidrolizado. Como se describe adicionalmente en la presente memoria, la sobreexpresión del gen de ADH1 en las cepas de *I. orientalis* en las que se ha delecionado o alterado ADHa y/o ADHb produjo una cepa de levadura que exhibió una utilización de xilosa y un título de etanol incrementados frente a una cepa original que tenía solamente la delección o alteración de ADHa y/o ADHb.

Como se discutió anteriormente, el papel funcional específico y la regulación de las ADHs no está conservado de manera generalizada en las especies de levaduras, y las ADHs de levadura exhiben una variación significativa con respecto a su actividad en presencia de glucosa, etanol, oxígeno, y otros reguladores potenciales. Además, la funcionalidad de las ADHs durante la fermentación de azúcares que no se fermentan de manera nativa por parte de una cepa hospedadora (p.ej., pentosas) se desconoce en gran medida o ha mostrado resultados divergentes de los descritos en la presente memoria. Por ejemplo, el documento WO10/039692 describió que la sobreexpresión de ADH1 no dio como resultado una producción de etanol incrementada en medios que contenían pentosas, a menos que también se sobreexpresase COX10. De forma similar, se demostró previamente que la sobreexpresión de ADH2 en *S. cerevisiae* no dio como resultado la disminución esperada del título de etanol (Maestre 2008). Por lo tanto, los efectos de la sobreexpresión de ADH1 y la delección de ADHa/ADHb sobre la utilización de xilosa y el título de etanol fueron inesperados. Como tal, en la presente memoria se proporcionan células de levadura modificadas genéticamente capaces de fermentar xilosa hasta etanol y que comprenden una o más modificaciones en un gen que codifica un polipéptido capaz de catalizar la conversión de acetaldehído hasta etanol o la conversión de etanol hasta acetaldehído. Estas modificaciones pueden incluir la delección o alteración de uno o más genes endógenos y/o la sobreexpresión de uno o más genes endógenos o exógenos. En ciertas realizaciones, las modificaciones incluyen una o más de delección o alteración de ADHa, delección o alteración de ADHb, y sobreexpresión de ADH1. También se proporcionan en la presente memoria métodos para producir las células de levadura modificadas genéticamente proporcionadas en la presente memoria y métodos para el uso de estas células de levadura modificadas genéticamente para producir etanol.

En la presente memoria se proporcionan, en ciertas realizaciones, células de levadura modificadas genéticamente que comprenden un genoma con una delección o alteración de uno o más genes endógenos que codifican ADHa y/o ADHb y/o una delección o alteración de uno o más elementos reguladores asociados a tales genes. "Delección o alteración", tal como se usa en la presente memoria con respecto a un gen, significa que la región codificante del gen se elimina completamente (delección) o se modifica de tal manera que el gen ya no es capaz de producir su polipéptido codificado, o produce un polipéptido con una actividad notablemente disminuida (alteración). "Delección o alteración", tal como se usa en la presente memoria con respecto a un elemento regulador, significa que el elemento regulador se elimina completamente o se modifica de tal manera que el gen al que está unido de forma operable ya no produce un polipéptido funcional, o produce un polipéptido con una actividad notablemente disminuida.

En ciertas realizaciones, las células de levadura modificadas genéticamente proporcionadas en la presente memoria comprenden una delección o alteración de uno o más genes que codifican un polipéptido que comprende la secuencia de aminoácidos expuesta en SEQ ID N°s:2 o 4 antes de la delección o alteración. En algunas de estas realizaciones, los genes delecionados o alterados comprendieron la región codificante de la secuencia de nucleótidos expuesta en SEQ ID N°:1 o 3 antes de la delección o alteración, mientras en otras realizaciones los genes delecionados o alterados comprendieron una secuencia de nucleótidos con al menos un 90% de identidad de secuencia respecto de la región codificante de la secuencia de nucleótidos expuesta en SEQ ID N°s:1 o 3 antes de la delección o alteración. En algunas de estas realizaciones, los genes delecionados o alterados comprendieron una secuencia de nucleótidos que tiene al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98%, al menos un 99%, o al menos un 99,5% de identidad de secuencia respecto de la región codificante de la secuencia de nucleótidos expuesta en SEQ ID N°s:1 o 3 antes de la alteración.

En ciertas realizaciones, los genes delecionados o alterados codificaron un polipéptido que comprendía una secuencia de aminoácidos con al menos un 90% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°s:2 o 4 antes de la delección o alteración. En algunas de estas realizaciones, el polipéptido codificado comprendió una secuencia de aminoácidos con al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98%, al menos un 99%, o al menos un 99,5% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°s:2 o 4. En ciertas realizaciones, los genes delecionados o alterados comprendieron una secuencia de nucleótidos con al menos un 90% de identidad de secuencia respecto de la región codificante de la secuencia de nucleótidos expuesta en SEQ ID N°s:1 o 3 antes de la delección o alteración. En algunas de estas realizaciones, los genes delecionados o alterados comprendieron una secuencia de nucleótidos que tiene al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98%, al menos un 99%, o al menos un 99,5% de identidad de secuencia respecto de la región codificante de la secuencia de nucleótidos expuesta en SEQ ID N°s:1 o 3 antes de la alteración.

En ciertas realizaciones, los genes delecionados o alterados codificaron un polipéptido con un 70% o más de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°s:2 o 4 antes de la delección o alteración, en el que el polipéptido codificado tuvo la capacidad de catalizar la conversión de etanol hasta acetaldehído *in vitro* o *in vivo*. En algunas de estas realizaciones, el polipéptido comprendió una secuencia de aminoácidos con al menos un 75%, al menos un 80%, al menos un 85%, o al menos un 90% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°s:2 o 4. En ciertas realizaciones, los genes delecionados o alterados comprendieron una secuencia de nucleótidos con al menos un 70% de identidad de secuencia respecto de la región codificante de la secuencia de nucleótidos expuesta en SEQ ID N°s:1 o 3 antes de la delección o alteración. En algunas de estas realizaciones, los genes delecionados o alterados comprendieron una secuencia de nucleótidos con al menos un 75%, al menos un 80%, al menos un 85%, o al menos un 90% de identidad de secuencia respecto de la región codificante de la secuencia de nucleótidos expuesta en SEQ ID N°s:1 o 3 antes de la delección o alteración.



La delección o alteración de un gen objetivo se puede llevar a cabo mediante cualquiera de varios métodos conocidos en la técnica. Por ejemplo, se puede transformar una célula con una construcción de delección. Se puede ensamblar una construcción de delección mediante el uso de dos secuencias de ADN objetivo clonadas del gen seleccionado para la delección o alteración desde sus regiones flanqueantes en posición anterior (5') o posterior (3'). Las dos secuencias de ADN del gen objetivo o sus regiones flanqueantes preferiblemente no están contiguas, pero pueden estar contiguas si se va a interponer material genético adicional (tal como un gen marcador de selección) entre ellas en la construcción de delección. En este contexto, "no contiguo" significa que las secuencias de ADN no están inmediatamente adyacentes entre sí en el genoma nativo, sino que están separadas por una región que se va a delecionar para delecionar o alterar el gen. Las secuencias "contiguas", tal como se usa en la presente memoria, están directamente adyacentes entre sí en el genoma nativo. Una de las secuencias clonadas puede incluir una región en 5' del promotor del gen objetivo, la totalidad o una porción de la región promotora, la totalidad o una porción de la secuencia codificante del gen objetivo, o cierta combinación de las mismas. La otra secuencia clonada puede incluir una región en 3' respecto del terminador del gen objetivo, la totalidad o una porción de la región terminadora, y/o la totalidad o una porción de la secuencia codificante del gen objetivo. Las dos secuencias objetivo clonadas se incorporan en la construcción de delección de forma que están orientadas en la misma dirección entre sí, tal como aparecen de manera nativa en el genoma de la célula hospedadora.

Se puede clonar un gen marcador de selección en la construcción de delección entre las dos secuencias del gen objetivo para permitir la selección de los transformantes. El gen marcador de selección se puede incorporar en la construcción de delección como parte de un casete de expresión que incluye opcionalmente uno o más elementos reguladores. Los transformantes eficaces contendrán el gen marcador de selección, que confiere a la célula transformada de manera eficaz al menos una característica que proporciona una base para la selección. Los genes marcadores de selección típicos codifican proteínas que (a) confieren resistencia a antibióticos u otras toxinas (p.ej., resistencia a bleomicina o zeomicina (p.ej., gen *ble* de *Streptoalloteichus hindustanus*), aminoglicósidos tales como G418 o kanamicina (p.ej., gen de resistencia de kanamicina del transposón Tn903), o higromicina (p.ej., gen de resistencia a antibióticos aminoglicósidos de *E. coli*)), (b) complementan deficiencias auxotróficas de la célula (p.ej., deficiencias de leucina (p.ej., gen LEU2 de *K. marxianus*), uracilo (p.ej., gen URA3 de *K. marxianus*, *S. cerevisiae*, o *I. orientalis*), o triptófano (p.ej., gen TRP de *K. marxianus*, *S. cerevisiae*, o *I. orientalis*)), (c) posibilitan que la célula sintetice nutrientes cruciales no disponibles en medios simples, o (d) confieren la capacidad a la célula de crecer con una fuente de carbono particular (p.ej., gen MEL5 de *S. cerevisiae*, que codifica la enzima alfa-galactosidasa (melibiasa) y confiere la capacidad de crecer con melibiosa como única fuente de carbono). Los marcadores de selección preferidos incluyen el gen de resistencia a zeocina, el gen de resistencia G418, el gen MEL5 y el gen de resistencia de higromicina. Otro marcador de selección preferido es un casete de gen de L-lactato:ferricitocromo c oxidorreductasa (*CYB2*), con tal de que la célula hospedadora carezca de manera nativa de tal gen o que su(s) gen(es) nativo(s) *CYB2* se delecionen o alteren primero.

Además de los genes marcadores de selección, se pueden incorporar otro u otros tipos de genes exógenos en una construcción de delección. Por ejemplo, se pueden clonar uno o más genes exógenos que codifican enzimas implicadas en una ruta de fermentación de etanol en la construcción de delección. Tras la transformación, la célula hospedadora expresará este gen exógeno en lugar del gen deleccionado o alterado. Como con los genes marcadores de selección, estos genes exógenos adicionales se pueden incorporar en la construcción de delección como parte de un casete de expresión que contiene opcionalmente uno o más elementos reguladores.

Esta construcción de delección se usa para transformar la célula hospedadora. Los métodos para transformar una célula de levadura con una construcción de ADN exógeno se describen, por ejemplo, en los documentos WO99/14335, WO00/71738, WO02/42471, WO03/102201, WO03/102152, y WO03/049525. La transformación se puede llevar a cabo mediante el uso de cualquier método conocido en la técnica, que incluye los métodos de electroporación y/o la transformación química (p.ej., cloruro de calcio, basados en acetato de litio, etc.). Se puede llevar a cabo una selección o cribado para identificar los transformantes eficaces. En los transformantes eficaces, un evento de recombinación homóloga en el locus del gen objetivo da como resultado la alteración o la delección del gen objetivo. Se puede delecionar la totalidad o una porción del gen objetivo nativo, su promotor y/o su terminador durante este evento de recombinación. Si la construcción de delección contiene material genético entre las dos secuencias clonadas del gen objetivo (p.ej., un casete de marcador de selección, casete de expresión), ese material genético se inserta en el genoma de la célula hospedadora en el locus del material deleccionado. Se puede llevar a cabo un análisis mediante PCR o un análisis de Southern para confirmar que ha tenido lugar la delección o delección/inserción deseada.

Cuando una construcción de delección comprende un gen marcador de selección, la construcción se puede diseñar de forma que el gen marcador se deleccione espontáneamente como resultado de un suceso de recombinación homóloga posterior. Una manera adecuada de conseguir esto es diseñar la construcción de delección de forma que el gen marcador de selección esté flanqueado por secuencias de repeticiones directas. Las secuencias de repeticiones directas son secuencias de ADN idénticas, nativas o no nativas para la célula hospedadora, y orientadas en la construcción en la misma dirección entre sí. Las secuencias de repeticiones directas tienen de manera ventajosa una longitud de alrededor de 50 a 1500 pb, y no tienen que codificar nada. La inclusión de las secuencias de repeticiones directas permite que se dé un suceso de recombinación homóloga, lo que da como resultado la delección del gen marcador de selección y una de las secuencias de repeticiones directas. Debido a que la recombinación homóloga se da con una frecuencia relativamente baja, puede ser necesario cultivar los transformantes durante

diversas rondas en medios no selectivos para permitir que la recombinación homóloga espontánea se dé en algunas de las células. Las células en las que se ha delecionado espontáneamente el gen marcador de selección se pueden seleccionar o cribar basándose en su pérdida de la característica de selección conferida por el gen marcador de selección. En ciertos casos, la expresión de una enzima recombinasa puede aumentar la recombinación entre los sitios repetidos.

En la presente memoria se proporcionan en ciertas realizaciones, como se define en las reivindicaciones, células de levadura modificadas genéticamente que comprenden una modificación genética que da como resultado la sobreexpresión de ADH1, lo que significa que las células expresan ADH1 a un nivel mayor que una célula nativa en al menos algunas condiciones. La modificación genética que da como resultado la sobreexpresión de ADH1 puede ser 1) la introducción de uno o más genes de ADH1 exógenos en una célula hospedadora; 2) la introducción de un elemento regulador exógeno que incrementa la expresión de un gen de ADH1 endógeno o exógeno en la célula hospedadora (p.ej., una secuencia promotora fuerte constitutiva o inducible); o 3) una modificación genética que activa o incrementa la actividad de un elemento regulador asociado con un gen de ADH1 exógeno o endógeno; o cualquier combinación de lo anterior. Por lo tanto, en la presente memoria se proporcionan en ciertas realizaciones células de levadura modificadas genéticamente que comprenden uno o más genes de ADH1 exógenos o endógenos. También se proporcionan en la presente memoria células de levadura modificadas genéticamente que comprenden uno o más promotores exógenos que incrementan la expresión de un gen de ADH1 exógeno o endógeno.

En ciertas realizaciones, se proporcionan células de levadura modificadas genéticamente que comprenden una o más copias de un gen de ADH1 exógeno. En algunas de estas realizaciones, las células comprenden además una o más copias de un gen de ADH1 endógeno. En estas realizaciones, la introducción de uno o más genes de ADH1 exógenos en la célula incrementa el número de copias del gen de ADH1. Se puede expresar igualmente ADH1 a partir de genes de ADH1 tanto endógenos como exógenos, o los genes de ADH1 endógenos y exógenos se pueden expresar a niveles diferentes. Por ejemplo, los genes de ADH1 exógenos se pueden expresar a un nivel mayor que los genes de ADH1 endógenos.

"Endógeno", tal como se usa en la presente memoria con respecto a componentes genéticos tales como genes, secuencias promotoras y terminadoras, significa que el componente genético está presente en una localización particular del genoma de una forma nativa de una célula de levadura particular. "Exógeno", tal como se usa en la presente memoria con respecto a componentes genéticos, significa que el componente genético no está presente en una localización particular en el genoma de una forma nativa de una célula de levadura particular. "Nativo", tal como se usa en la presente memoria con respecto a una célula de levadura, se refiere a una célula de levadura de tipo natural de una especie de levadura particular. "Nativo", tal como se usa en la presente memoria con respecto a una ruta metabólica, se refiere a una ruta metabólica que existe y es activa en una célula de levadura nativa.

Un componente genético exógeno puede tener una secuencia nativa o no nativa. Un componente genético exógeno con una secuencia nativa comprende una secuencia idéntica (aparte de las mutaciones de individuo a individuo que no afectan a la función) a un componente genético que está presente en el genoma de una célula nativa (es decir, el componente genético exógeno es idéntico a un componente genético endógeno). Sin embargo, el componente exógeno está presente en una localización diferente en el genoma de la célula hospedadora que el componente endógeno. Por ejemplo, se puede insertar un gen de ADH1 exógeno que es idéntico a un gen de ADH1 endógeno en una célula de levadura, lo que da como resultado una célula modificada con un número no nativo (incrementado) de copias del gen de ADH1. Un componente genético exógeno con una secuencia no nativa comprende una secuencia que no se halla en el genoma de una célula nativa. Por ejemplo, se puede insertar un gen exógeno de una especie particular en una célula de levadura de otra especie. Un gen exógeno se integra preferiblemente en el genoma de la célula hospedadora de una manera funcional, lo que significa que es capaz de producir una proteína activa en la célula hospedadora. Sin embargo, el gen exógeno se puede introducir en la célula como parte de un vector que se mantiene de manera estable en el citoplasma del hospedador.

En ciertas realizaciones, las células de levadura modificadas genéticamente proporcionadas en la presente memoria que sobreexpresan ADH1 comprenden un gen de ADH1 exógeno o endógeno que codifica un polipéptido que comprende la secuencia de aminoácidos de SEQ ID N°:6. En algunas de estas realizaciones, el gen de ADH1 comprende la secuencia de nucleótidos expuesta en SEQ ID N°:5. En otras realizaciones, el gen de ADH1 comprende una secuencia de nucleótidos con al menos un 90% de identidad de secuencia respecto de la secuencia de nucleótidos expuesta en SEQ ID N°:5. En algunas de estas realizaciones, los polinucleótidos comprenden una secuencia de nucleótidos que tiene al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98%, al menos un 99%, o al menos un 99,5% de identidad de secuencia respecto de las secuencias de nucleótidos expuestas en cualquiera de SEQ ID N°:5.

En ciertas realizaciones, las células de levadura modificadas genéticamente proporcionadas en la presente memoria que sobreexpresan ADH1 comprenden un gen de ADH1 exógeno o endógeno que codifica un polipéptido que comprende una secuencia de aminoácidos con al menos un 90% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°:6. En algunas de estas realizaciones, el polipéptido comprende una secuencia de aminoácidos con al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98%, al menos un 99%, o al menos un 99,5% de identidad de secuencia respecto de la secuencia de aminoácidos de SEQ

ID N°:6. En ciertas realizaciones, el gen de ADH1 comprende una secuencia de nucleótidos con al menos un 90% de identidad de secuencia respecto de la secuencia de nucleótidos expuesta en SEQ ID N°:5. En algunas de estas realizaciones, el gen de ADH1 comprende una secuencia de nucleótidos con al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98%, al menos un 99%, o al menos un 99,5% de identidad de secuencia respecto de la secuencia de nucleótidos expuesta en SEQ ID N°:5.

En ciertas realizaciones, las células de levadura modificadas genéticamente proporcionadas en la presente memoria que sobreexpresan ADH1 comprenden un gen de ADH1 exógeno o endógeno que codifica un polipéptido que comprende una secuencia de aminoácidos con al menos un 70% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°:6, en el que el polipéptido codificado tiene la capacidad de catalizar la conversión de acetaldehído hasta etanol *in vitro* o *in vivo*. En algunas de estas realizaciones, el polipéptido codificado comprende una secuencia de aminoácidos con al menos un 75%, al menos un 80%, al menos un 85%, o al menos un 90% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°:6. En ciertas realizaciones, el gen de ADH1 comprende una secuencia de nucleótidos con al menos un 70% de identidad de secuencia respecto de la secuencia de nucleótidos expuesta en SEQ ID N°:5. En algunas de estas realizaciones, el gen de ADH1 comprende una secuencia de nucleótidos con al menos un 75%, al menos un 80%, al menos un 85%, o al menos un 90% de identidad de secuencia respecto de la secuencia de nucleótidos expuesta en SEQ ID N°:5.

En las células de levadura proporcionadas en la presente memoria que comprenden una o más copias de un gen de ADH1 exógeno, el gen puede estar unido de forma operable a uno o más elementos reguladores, tales como un promotor o terminador. En ciertas realizaciones, estos elementos reguladores pueden ser nativos para la célula hospedadora, es decir, se puede insertar un gen exógeno en una célula de levadura de forma que esté bajo control transcripcional de un promotor y/o terminador endógenos. En otras realizaciones, los elementos reguladores pueden ser exógenos. En estas realizaciones, los elementos reguladores se pueden haber introducido en la célula como parte de la construcción de expresión del gen de ADH1 exógeno. Los promotores unidos a uno o más genes de ADH1 exógenos pueden ser promotores fuertes, tales como promotores constitutivos o inducibles. En ciertas realizaciones, los promotores o terminadores exógenos pueden ser idénticos, o al menos idénticos a lo largo de sus porciones funcionales, a las secuencias promotoras y terminadoras nativas. En otras realizaciones, los promotores y terminadores exógenos pueden comprender una secuencia de nucleótidos que exhibe un grado relativamente elevado de identidad de secuencia respecto de secuencias promotoras o terminadoras nativas. Por ejemplo, un gen de ADH1 exógeno puede estar unido de forma operable a un promotor o terminador exógeno con al menos un 50%, al menos un 60%, al menos un 70%, al menos un 80%, al menos un 90%, o al menos un 95% de identidad de secuencia respecto de un promotor o terminador nativo. El promotor o terminador nativo hacia el que el promotor o terminador exógeno exhibe este grado elevado de identidad de secuencia puede estar unido de manera nativa a un gen de ADH1 endógeno, a otro gen implicado en la producción de etanol, o a un gen no relacionado. En las realizaciones en las que se insertan múltiples genes exógenos en una célula hospedadora, cada gen exógeno puede estar bajo control de un promotor y/o terminador diferente, o dos o más genes exógenos pueden estar bajo control del mismo promotor y/o terminador.

En las realizaciones en las que las células de levadura proporcionadas en la presente memoria comprenden una o más copias de un gen de ADH1 exógeno, el gen exógeno se puede introducir por medio de cualquier método conocido en la técnica. El gen de ADH1 exógeno se puede integrar en el genoma de la célula hospedadora de una manera aleatoria o selectiva. En las realizaciones en las que el gen se integra de una manera selectiva, se puede integrar en los loci de un gen particular, de forma que la integración del gen exógeno se acopla a la delección o alteración de un gen nativo. Por ejemplo, la introducción del gen de ADH1 exógeno se puede acoplar a la delección de uno o más genes implicados en una ruta de producción de etanol, tal como un gen de ADHa o ADHb. De manera alternativa, el gen exógeno se puede integrar en una porción del genoma que no corresponde a un gen.

La integración selectiva puede utilizar una construcción de delección, como se describió anteriormente. En estos métodos, se incorpora un gen de ADH1 en la construcción entre las dos secuencias objetivo clonadas. El gen de ADH1 se puede incorporar en la construcción solo o como parte de un casete de expresión que comprende uno o más elementos reguladores, tales como promotores y/o terminadores. Cuando la construcción comprende un gen marcador de selección, el gen marcador de selección o casete y el gen de ADH1 o casete pueden estar contiguos o no contiguos. En las realizaciones en las que la integración del gen de ADH1 exógeno se va a acoplar con la delección o alteración de un gen objetivo, las secuencias objetivo se obtienen del gen objetivo y/o sus regiones flanqueantes. En las realizaciones en las que la integración del gen de ADH1 exógeno no se acopla a la delección o alteración de un gen objetivo, las secuencias objetivo se seleccionan de forma que ningún gen se extienda en la región entre las secuencias objetivo. Tras la transformación de la célula hospedadora, el gen de ADH1 se inserta en un sitio objetivo mediante recombinación homóloga.

Se puede introducir más de una copia de un gen de ADH1 exógeno en la célula de levadura. Por ejemplo, se pueden introducir de una a diez copias del gen de ADH1. Cuando se introducen múltiples copias de un gen de ADH1, las copias pueden ser idénticas o pueden variar con respecto a la secuencia exacta del gen de ADH1. Las diferentes copias del gen de ADH1 exógeno se pueden integrar en el genoma de la célula de levadura en una única localización de forma que estén adyacentes entre sí, o se pueden integrar en diferentes localizaciones. Cada copia del gen de ADH1 se puede unir al mismo o a diferentes promotores, terminadores, y/o marcadores de selección.

En ciertas realizaciones, se proporcionan células de levadura modificadas genéticamente que comprenden uno o más promotores exógenos unidos de forma operable a uno o más genes de ADH1 endógenos. En estas realizaciones, el promotor exógeno puede sustituir o complementar a un promotor nativo asociado al gen de ADH1 endógeno. La incorporación de los promotores exógenos da como resultado la expresión incrementada de ADH1 frente a las células nativas.

Aunque la delección o alteración de ADHa y/o ADHb o la sobreexpresión de ADH1 sola es suficiente para incrementar la utilización de xilosa y el título de etanol, la combinación de ambas modificaciones dio como resultado un incremento mayor que cada modificación sola. Por lo tanto, en ciertas realizaciones, se proporcionan células de levadura modificadas genéticamente que comprenden tanto un genoma con una delección o alteración de uno o más genes endógenos que codifican ADHa y/o ADHb como una modificación genética que da como resultado la sobreexpresión de ADH1. En ciertas realizaciones, la modificación genética que da como resultado la sobreexpresión de ADH1 es la presencia de una o más copias de un gen de ADH1 exógeno.

Las células de levadura modificadas genéticamente proporcionadas en la presente memoria se pueden seleccionar de una diversidad de especies de levaduras. En ciertas realizaciones, las células de levadura modificadas genéticamente proporcionadas en la presente memoria son células de levadura distintas de *Saccharomyces*. En algunas de estas realizaciones, las células de levadura son células de levadura Crabtree-negativas, y en algunas de estas realizaciones las células de levadura pertenecen al clado *I. orientalis/P. fermentans*. El clado *I. orientalis/P. fermentans* es el clado más terminal que contiene al menos la especie *I. orientalis*, *P. galeiformis*, *P. sp. YB-4149* (denominación de la NRRL), *C. ethanolica*, *P. deserticola*, *P. membranifaciens*, y *P. fermentans*. Los miembros del clado *I. orientalis/P. fermentans* se identifican mediante el análisis del dominio D1/D2 variable del ADN ribosómico 26S de las especies de levaduras, mediante el uso del método descrito por Kurtzman y Robnett en "Identification and Phylogeny of Ascomycetous Yeasts from Analysis of Nuclear Large Subunit (26S) Ribosomal DNA Partial Sequences", *Antonie van Leeuwenhoek* 73:331-371, 1998, (véase en especial la pág. 349). El análisis del dominio D1/D2 variable del ADN ribosómico 26S de cientos de ascomicetos ha revelado que el clado *I. orientalis/P. fermentans* contiene especies muy relacionadas. Los miembros del clado *I. orientalis/P. fermentans* exhiben una similitud mayor en el dominio D1/D2 variable del ADN ribosómico 26S respecto de otros miembros del clado que una especie de levadura fuera del clado. Por lo tanto, se pueden identificar otros miembros del clado *I. orientalis/P. fermentans* mediante la comparación de los dominios D1/D2 de su respectivo ADN ribosómico y comparándolo con el de otros miembros del clado y especies estrechamente relacionadas fuera del clado, mediante el uso de los métodos de Kurtzman y Robnett. En ciertas realizaciones, las células de levadura modificadas genéticamente proporcionadas en la presente memoria pertenecen al género *Issatchenkia*, y en algunas de estas realizaciones las células de levadura son *I. orientalis*. Cuando se caracterizó por primera vez, a la especie *I. orientalis* se le asignó el nombre *Pichia kudriavzevii*. El anamorfo (forma asexual) de *I. orientalis* se conoce como *Candida krusei*. Se han enumerado numerosos sinónimos adicionales para la especie *I. orientalis* en otra parte (Kurtzman y Fell, *The Yeasts, a Taxonomic Study*. Sección 35. *Issatchenkia Kudryavtsev*, págs. 222-223 (1998)). *I. orientalis* y otros miembros del clado *I. orientalis/P. fermentans* exhiben algunas características que las hacen ideales para la fermentación de etanol a partir de biomasa, lo que incluye la tolerancia al pH bajo, al etanol, a la temperatura elevada (40 °C o más), y a diversos inhibidores presentes en el hidrolizado.

Como se expone más adelante en los ejemplos, se llevó a cabo el análisis de la expresión de ADH1, ADHa, y ADHb mediante el uso de una cepa de *I. orientalis* (cepa 1822) que previamente se había seleccionado por la resistencia a ácido 2-hidroxipropiónico. Por lo tanto, en ciertas realizaciones, las células de levadura modificadas genéticamente proporcionadas en la presente memoria pueden haber experimentado mutación y/o selección hacia la resistencia a etanol, ácidos orgánicos, otros productos de fermentación o subproductos, o componentes del medio tales como acetato. La selección se puede llevar a cabo antes, durante, o después de la introducción de las modificaciones genéticas relacionadas con ADH1, ADHa, y/o ADHb mediante el uso de métodos conocidos en la técnica. Por ejemplo, la selección se puede llevar a cabo mediante el uso de un quimiostato. Un quimiostato es un dispositivo que permite un cultivo continuo de microorganismos (p.ej., levaduras), en el que se puede controlar independientemente la velocidad específica de crecimiento y el número de células. Un cultivo continuo es básicamente un sistema de flujo de volumen constante al cual se le añade medio continuamente y en el cual se puede llevar a cabo una extracción continua de cualquier rebosamiento. Una vez que tal sistema está en equilibrio, el número de células y el estado de los nutrientes permanecen constantes, y el sistema está en un estado estacionario. Un quimiostato permite el control de la densidad de la población y la velocidad específica de crecimiento de un cultivo por medio de la tasa de dilución y la alteración de la concentración de un nutriente limitante, tal como una fuente de carbono o nitrógeno. Mediante la alteración de las condiciones a medida que un cultivo crece (p.ej., disminuyendo la concentración de una fuente secundaria de carbono necesaria para el crecimiento de la cepa del inóculo, entre otros), se seleccionarán los microorganismos de la población que son capaces de crecer más rápido en las condiciones alteradas y superarán a los microorganismos que no funcionan tan bien en las nuevas condiciones. En general, tal selección requiere el incremento o disminución progresiva de al menos un componente del cultivo a lo largo del desarrollo del crecimiento del cultivo en quimiostato. El funcionamiento de los quimiostatos y su uso en la evolución dirigida de microorganismos se conoce bien en la técnica (véase, p.ej., Novick *Proc Natl Acad Sci USA* 36:708-719 (1950), Harder *J Appl Bacteriol* 43:1-24 (1977)).

En ciertas realizaciones, las células de levadura proporcionadas en la presente memoria comprenden una o más modificaciones genéticas además de la sobreexpresión de ADH1 y/o delección o alteración de ADHa/ADHb. Estas

modificaciones genéticas adicionales pueden incluir una o más de las siguientes: sobreexpresión de XI; sobreexpresión de XK; delección o alteración de uno o más genes que codifican un polipéptido con actividad de XR; delección o alteración de uno o más genes que codifican polipéptidos con actividad de XDH; sobreexpresión de uno o más genes en la ruta no oxidativa de pentosas fosfato (transaldolasa (TAL), transcetolasa (TKL), D-ribulosa-5-fosfato 3-epimerasa (RPE), ribulosa 5-fosfato cetol-isomerasa (RKI)); expresión de uno o más genes en una ruta de consumo de arabinosa; expresión de un transportador de pentosas; y delección o alteración de uno o más genes implicados en la conversión de acetaldehído hasta ácido acético, tal como ALD.

En la presente memoria se proporcionan en ciertas realizaciones, como se define en las reivindicaciones, métodos para producir una célula de levadura modificada genéticamente capaz de fermentar xilosa hasta etanol deleccionando o alterando uno o más genes endógenos que codifican ADHa y/o ADHb. En ciertas realizaciones, los genes deleccionados o alterados codificaron un polipéptido que comprendía la secuencia de aminoácidos expuesta en SEQ ID N°s:2 o 4 antes de la delección o alteración. En algunas de estas realizaciones, los genes deleccionados o alterados comprendieron la región codificante de la secuencia de nucleótidos expuesta en SEQ ID N°:1 o 3 antes de la delección o alteración, mientras en otras realizaciones los genes deleccionados o alterados comprendieron una secuencia de nucleótidos con al menos un 90%, al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98%, al menos un 99%, o al menos un 99,5% de identidad de secuencia respecto de la región codificante de la secuencia de nucleótidos expuesta en SEQ ID N°s:1 o 3 antes de la delección o alteración. En ciertas realizaciones, los genes deleccionados o alterados codificaron un polipéptido que comprendía una secuencia de aminoácidos con al menos un 90%, al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98%, al menos un 99%, o al menos un 99,5% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°s:2 o 4 antes de la delección o alteración, y en algunas de estas realizaciones el gen deleccionado o alterado comprendió una secuencia de nucleótidos con al menos un 90%, al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98%, al menos un 99%, o al menos un 99,5% de identidad de secuencia respecto de la región codificante de la secuencia de nucleótidos expuesta en SEQ ID N°s:1 o 3. En ciertas realizaciones, los genes deleccionados o alterados codificaron un polipéptido que comprendía una secuencia de aminoácidos con al menos un 70% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°s:2 o 4 antes de la delección o alteración, en el que el polipéptido fue capaz de catalizar la conversión de etanol hasta acetaldehído. En algunas de estas realizaciones, el polipéptido codificado comprendió una secuencia de aminoácidos con al menos un 75%, al menos un 80%, al menos un 85%, o al menos un 90% de identidad de secuencia respecto de la secuencia de aminoácidos de SEQ ID N°s:2 o 4. En algunas de estas realizaciones, los genes deleccionados o alterados comprendieron una secuencia de nucleótidos con al menos un 70%, al menos un 75%, al menos un 80%, al menos un 85%, o al menos un 90% de identidad de secuencia respecto de la región codificante de la secuencia de nucleótidos expuesta en SEQ ID N°s:1 o 3 antes de la delección o alteración. En ciertas realizaciones, se introducen una o más modificaciones genéticas adicionales en las células de levadura además de la delección o alteración de uno o más genes de ADHa y/o ADHb. En algunas de estas realizaciones, las células se modifican para sobreexpresar ADH1. En algunas de estas realizaciones, la sobreexpresión de ADH1 se lleva a cabo introduciendo una o más copias de un gen de ADH1 exógeno. En otras realizaciones, la sobreexpresión se lleva a cabo mediante la expresión creciente de una o más copias endógenas del gen de ADH1 que ya están presentes en la célula.

En la presente memoria se proporcionan en ciertas realizaciones métodos para producir una célula de levadura modificada genéticamente capaz de fermentar xilosa hasta etanol introduciendo una modificación genética que da como resultado la sobreexpresión de ADH1. ADH1 se puede sobreexpresar a partir de uno o más genes exógenos, uno o más genes endógenos, o una combinación de los mismos. Por lo tanto, en ciertas realizaciones, estos métodos comprenden introducir uno o más genes de ADH1 exógenos en una célula de levadura hospedadora, de forma que la célula comprende una o más copias de un gen de ADH1 exógeno. En ciertas realizaciones, el gen de ADH1 que se sobreexpresa codifica un polipéptido que comprende la secuencia de aminoácidos de SEQ ID N°:6. En algunas de estas realizaciones, el gen de ADH1 que se sobreexpresa comprende la secuencia de nucleótidos expuesta en SEQ ID N°:5. En otras realizaciones, el gen de ADH1 comprende una secuencia de nucleótidos con al menos un 90%, al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98%, al menos un 99%, o al menos un 99,5% de identidad de secuencia respecto de la secuencia de nucleótidos expuesta en SEQ ID N°:5. En ciertas realizaciones, el gen de ADH1 que se sobreexpresa codifica un polipéptido que comprende una secuencia de aminoácidos con al menos un 90%, al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98%, al menos un 99%, o al menos un 99,5% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°:6. En algunas de estas realizaciones, el gen de ADH1 que se sobreexpresa comprende una secuencia de nucleótidos con al menos un 90% de identidad de secuencia respecto de la secuencia de nucleótidos expuesta en SEQ ID N°:5. En ciertas realizaciones, el gen de ADH1 que se sobreexpresa codifica un polipéptido que comprende una secuencia de aminoácidos con al menos un 70% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°:6, en el que el polipéptido es capaz de catalizar la conversión de acetaldehído hasta etanol. En algunas de estas realizaciones, el polipéptido comprende una secuencia de aminoácidos con al menos un 75%, al menos un 80%, al menos un 85%, o al menos un 95% de identidad de secuencia respecto de la secuencia de aminoácidos de SEQ ID N°:6. En algunas de estas realizaciones, el gen de ADH1 que se sobreexpresa comprende una secuencia de nucleótidos con al menos un 70%, al menos un 75%, al menos un 80%, al menos un 85%, o al menos un 90% de identidad de secuencia respecto de la secuencia de nucleótidos expuesta en SEQ ID N°:5. En ciertas realizaciones, se introducen una o más modificaciones genéticas adicionales en las células de levadura además de las modificaciones que dan como resultado la sobreexpresión de

ADH1. En algunas de estas realizaciones, las células se modifican delecionando o alterando uno o más genes de ADHa o ADHb.

5 En ciertas realizaciones, se proporcionan procesos de fermentación en los que se cultiva una célula de levadura modificada genéticamente tal como se proporciona en la presente memoria en condiciones de fermentación. En ciertas realizaciones, las células de levadura comprenden un genoma con una delección o alteración de uno o más genes que codifican ADHa y/o ADHb. En otras realizaciones, las células de levadura comprenden una modificación genética que da como resultado la sobreexpresión de ADH1, y en algunas de estas realizaciones las células de levadura comprenden una o más copias de un gen de ADH1 exógeno. En ciertas realizaciones, las células de levadura comprenden una combinación de modificaciones genéticas que dan como resultado la sobreexpresión de ADH1 y la delección o alteración de uno o más genes que codifican ADHa y/o ADHb. En algunas de estas realizaciones, el proceso de fermentación da como resultado la producción de etanol.

10 En ciertas realizaciones, se proporcionan métodos para producir etanol cultivando una célula de levadura modificada genéticamente como se proporciona en la presente memoria con una o más pentosas y/o hexosas. En ciertas realizaciones, las células de levadura comprenden un genoma con una delección o alteración de uno o más genes que codifican ADHa y/o ADHb. En otras realizaciones, las células de levadura comprenden una modificación genética que da como resultado la sobreexpresión de ADH1. En algunas de estas realizaciones, las células de levadura comprenden una o más copias de un gen de ADH1 exógeno. En ciertas realizaciones, las células de levadura comprenden una combinación de modificaciones genéticas que dan como resultado la sobreexpresión de ADH1 y la delección o alteración de uno o más genes que codifican ADHa y/o ADHb.

15 En ciertas realizaciones de los procesos y métodos proporcionados en la presente memoria, los medios usados para cultivar las células de levadura modificadas genéticamente proporcionadas en la presente memoria comprenden uno o más azúcares distintos de glucosa que son fermentables por parte de las células. En algunas de estas realizaciones, los azúcares distintos de glucosa pueden ser xilosa, xilano, otro oligómero de xilosa, y/o arabinosa. Estos azúcares distintos de glucosa pueden ser hidrolizados de una biomasa que contiene hemicelulosa, tal como un hidrolizado de biomasa vegetal. Los medios pueden comprender además glucosa y/u oligómeros o polímeros de glucosa. Cuando hay azúcares multiméricos, puede ser necesario añadir enzimas al caldo de fermentación para digerir estos azúcares hasta el azúcar monomérico correspondiente.

20 En ciertas realizaciones del proceso y los métodos proporcionados en la presente memoria, el medio usado para cultivar las células de levadura modificadas genéticamente proporcionadas en la presente memoria es un medio que contiene xilosa, y en algunas de estas realizaciones la xilosa deriva de un hidrolizado de biomasa vegetal. En ciertas realizaciones, la xilosa puede estar presente en el medio a una concentración de alrededor de 0 a alrededor de 150 g/L al comienzo de la fermentación (es decir, en o antes del punto en el que se añaden las células al medio) y/o en diversos momentos durante el proceso de fermentación. En algunas de estas realizaciones, la xilosa puede estar presente en el medio a una concentración de al menos alrededor de 10 g/L, 20 g/L, 30 g/L, 40 g/L, 50 g/L, 75 g/L, 100 g/L, o 125 g/L. En ciertas realizaciones, los medios pueden comprender uno o más azúcares además de xilosa, que incluyen una o más pentosas y/o hexosas. En algunas de estas realizaciones, la xilosa puede constituir alrededor del 10 a alrededor del 95% del contenido total de azúcares del medio al comienzo de la fermentación y/o en diversos momentos durante el proceso de fermentación. En algunas de estas realizaciones, la xilosa puede constituir al menos alrededor del 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, o 90% del contenido total de azúcares del medio. En ciertas realizaciones, las células de levadura modificadas genéticamente pueden fermentar uno o más de los azúcares adicionales presentes en los medios hasta etanol.

25 En ciertas realizaciones del proceso y los métodos proporcionados en la presente memoria, el medio es un medio sintético tal como un extracto de levadura/medio de peptona, y en algunas de estas realizaciones el medio puede contener acetato. En otras realizaciones, el medio es un medio sintético definido, y en algunas de estas realizaciones el medio puede contener acetato. En ciertas realizaciones, el medio comprende cierto porcentaje de hidrolizado de biomasa, tal como hidrolizado de rastrojo de maíz. En estas realizaciones, el hidrolizado puede estar presente en el medio en cualquier porcentaje de alrededor del 10% al 100% del volumen total de medio. En algunas de estas realizaciones, el hidrolizado se puede haber pretratado. Por ejemplo, el hidrolizado se puede haber pretratado con uno o más ácidos o enzimas para descomponer parcialmente la materia prima. En ciertas realizaciones, el hidrolizado es un hidrolizado sin destoxificar. En las realizaciones en las que el medio comprende un hidrolizado a menos del 100%, el resto del medio puede comprender uno o más agentes diluyentes que incluyen un medio sintético o agua.

30 En ciertas realizaciones, el cultivo de las células proporcionadas en la presente memoria para producir etanol se puede dividir en fases. Por ejemplo, el proceso de cultivo de células se puede dividir en una fase de cultivo, una fase de producción, y una fase de recuperación. Alguien de experiencia habitual en la técnica reconocerá que se pueden variar estas condiciones basándose en factores tales como la especie de levadura a usar, la ruta de fermentación específica utilizada por la levadura, el rendimiento deseado, u otros factores.

35 En ciertas realizaciones de los procesos y métodos proporcionados en la presente memoria, las células se cultivan a una temperatura de alrededor de 20 °C a alrededor de 60 °C. En algunas de estas realizaciones, la fermentación tiene lugar a una temperatura que oscila de alrededor de 30 °C a alrededor de 50 °C, y en algunas de estas

realizaciones la fermentación tiene lugar a una temperatura de alrededor de 35 °C a alrededor de 45 °C. La temperatura se puede variar a lo largo del proceso de fermentación.

La fermentación se puede llevar a cabo de manera aerobia, microaerobia, sustancialmente anaerobia, o anaerobia. Si se desea, se puede variar la velocidad de absorción de oxígeno a lo largo de la fermentación como control del proceso (véase, p.ej., el documento WO03/102200). En ciertas realizaciones preferidas, la fermentación puede tener lugar en condiciones microaeróbicas, que se caracterizan por una velocidad de absorción de oxígeno de alrededor de 2 a alrededor de 25 mmol/L/h.

### Ejemplos

Ejemplo 1: Identificación de los genes de ADH1, ADHa, y ADHb de *I. orientalis*:

Se usó la secuencia de aminoácidos de ADH2 de *S. cerevisiae* para llevar a cabo una búsqueda con Blast del genoma de *I. orientalis* de tipo natural. Se identificaron tres supuestos homólogos: los marcos de lectura abiertos (ORFs) S141G9091, S141G1202, y S141G2556. S141G9091 tuvo la secuencia de ADN expuesta en SEQ ID N°:1. La región codificante de S141G9091 (nucleótidos 1052 a 2182 de SEQ ID N°:1) codifica la secuencia polipeptídica expuesta en SEQ ID N°:2. S141G1202 tuvo la secuencia de ADN expuesta en SEQ ID N°:3. La región codificante de S141G1202 (nucleótidos 1001 a 2134 de SEQ ID N°:3) codifica la secuencia polipeptídica expuesta en SEQ ID N°:4. La región codificante de S141G2556 tuvo la secuencia de ADN expuesta en SEQ ID N°:5, y codifica la secuencia polipeptídica expuesta en SEQ ID N°:6.

Las alineaciones de los homólogos de *I. orientalis* con los homólogos de ADH caracterizados de *Saccharomyces* y otras especies de levaduras demostraron que todos los homólogos fueron aproximadamente iguales en similitud a los homólogos de ADH1, ADH2 y ADH3. La Figura 15 muestra una alineación de secuencias de aminoácidos de S141G9091 (SEQ ID N°:2), S141G1202 SEQ ID N°:4, y S141G2556 (SEQ ID N°:6) con ADH1 (SEQ ID N°: 13), ADH2 (SEQ ID N°:14), y ADH3 (SEQ ID N°:15) de *S. cerevisiae*. La Tabla 1 resume el porcentaje de identidad entre las secuencias de aminoácidos de S141G9091, S141G1202, y S141G2556 y ADH1, ADH2, y ADH3 de *S. cerevisiae*. S141G9091 y S141G1202 poseen una prolongación N-terminal que puede ser indicativa de una secuencia de selección de orgánulo.

Tabla 1: Porcentaje de identidad entre secuencias de aminoácidos

	S141G2556	S141G9091	S141G1202
ScADH1	73%	70%	71%
ScADH2	74%	69%	71%
ScADH3	70%	71%	75%

Debido a las similitudes en la homología entre las comparaciones por parejas, se analizó la expresión de ARN en la cepa 1822 de *I. orientalis* para identificar qué homólogo fue la ADH fermentativa principal y que pudiera estar implicado en el consumo de etanol. La cepa 1822 es una cepa resistente a ácido 2-hidroxipropiónico que se obtuvo evolucionando la cepa ATCC PTA-6658 de *I. orientalis* en un quimiostato con limitación de glucosa. Durante este proceso, el sistema se alimentó con 15 g/L de dextrosa en un medio DM, y se hizo funcionar a una tasa de dilución de 0,06 h<sup>-1</sup> a pH 3,0 con ácido 2-hidroxipropiónico añadido en el medio de alimentación. Las condiciones se mantuvieron con una velocidad de transferencia de oxígeno baja de aproximadamente 2 mmol L<sup>-1</sup>h<sup>-1</sup>, y la concentración de oxígeno disuelto permaneció constante al 0% de saturación de aire. La concentración de ácido 2-hidroxipropiónico en el medio de alimentación se incrementó en incrementos de 5 g/L aproximadamente cada dos semanas desde una concentración inicial de 30 g/L hasta una concentración final de 60 g/L. Los aislamientos de colonias individuales del tiempo final se caracterizaron en dos ensayos con matraces de agitación. En el primer ensayo, las cepas se caracterizaron por su capacidad de fermentar glucosa hasta etanol en presencia de 25 g/L de ácido 2-hidroxipropiónico libre. En el segundo ensayo, se midieron las velocidades de crecimiento de los aislamientos en presencia de 25, 32 y 45 g/L de ácido 2-hidroxipropiónico total sin ajuste del pH. La cepa 1822 representó un único aislamiento que se seleccionó basándose en las velocidades medidas de fermentación y crecimiento.

Para obtener biomasa para el análisis de la expresión, se centrifugó un cultivo nocturno de la cepa 1822 de *I. orientalis* cultivada en medio YPD (medio basado en YP (10 g/L de extracto de levadura y 20 g/L de peptona) que contenía 100 g/L de dextrosa), se lavó, y se usó para inocular matraces de 50 mL (50 mL de medio YP en matraces de 250 mL) que contenían un 2% de etanol o 2% de glucosa. Los cultivos se cultivaron a 37 °C y 250 rpm hasta una DO<sub>600</sub> de 2,0, y se centrifugaron muestras de 10 mL y se congelaron en nitrógeno líquido. El ARN se aisló y se usó para obtener cADN mediante el uso de una transcriptasa inversa (Promega). Se llevó a cabo una PCR cuantitativa mediante el uso de cebadores específicos para cada homólogo (S141G9091: SEQ ID N°:9 (directo), SEQ ID N°:10 (inverso); S141G1202: SEQ ID N°: 11 (directo), SEQ ID N°:12 (inverso); S141G2556: SEQ ID N°:7 (directo), SEQ ID N°:8 (inverso)). Los resultados se resumen en la Tabla 2. Uno de los tres homólogos (S141G9091) mostró expresión

solamente con etanol como sustrato. Los otros dos homólogos (S141G1202 y S141G2556) mostraron expresión con los sustratos de etanol y glucosa, aunque el nivel de expresión de S141G1202 fue mucho menor que el de S141G2556.

Tabla 2: Valores de C(t) en cultivos con glucosa y etanol

	Glucosa	Etanol
S141G1202	33,1, 33,2	34,2, 34,0
S141G9091	N/A	28,3, 28,5
S141G2556	27,3, 27,4	28,4, 28,6
actina	36,5	35,8/37,4

5

El análisis de la expresión con micromatriz (Nimblegen) se llevó a cabo con la cepa 3556 fermentadora de xilosa (derivada de la cepa 1822) cultivada en fermentadores en medio YP con glucosa, xilosa, o una mezcla de glucosa y xilosa como fuente de carbono. La concentración de oxígeno disuelto en estas fermentaciones se midió mediante el uso de un electrodo de oxígeno disuelto polarográfico. La concentración de oxígeno disuelto se expresa como porcentaje de la concentración saturada de oxígeno en el medio de fermentación con aire a una presión ambiental de 1 atmósfera. Las muestras para la extracción de ARN se tomaron dos horas después de que el oxígeno disuelto alcanzase el cero por ciento para los cultivos cultivados con glucosa, cinco horas después de que el oxígeno disuelto alcanzase el cero por ciento para los cultivos cultivados con una mezcla de glucosa-xilosa, y diez horas después de que el oxígeno disuelto alcanzase el cero por ciento para los cultivos cultivados con xilosa. Los niveles normalizados de expresión para los tres loci se muestran en la Tabla 3.

10

15

Tabla 3: Niveles normalizados de expresión en cultivos de glucosa, xilosa, y glucosa/xilosa

	Glucosa	Xilosa	Glucosa + Xilosa
S141G1202	1551	54362	808
S141G9091	17092	15446	25412
S141G2556	39781	32743	47484

Basándose en los niveles y patrones de expresión, se concluyó que S141G2556 representa la enzima ADH fermentativa principal. S141G2556 se denominó, por lo tanto, ADH1 de *I. orientalis*. Los otros dos homólogos exhibieron una expresión baja en presencia de glucosa al menos en ciertas condiciones, un comportamiento más coherente con un papel en el consumo de etanol. Sin embargo, debido a que este comportamiento no fue coherente en los estudios de expresión, estos homólogos se denominaron ADHa (S141G9091) y ADHb (S141G1202) de *I. orientalis*.

20

Ejemplo 2: Caracterización de ADHa mediante el uso de inactivaciones de genes:

Para confirmar el papel de ADHa en el metabolismo del etanol, se desarrolló una cepa de *I. orientalis* con ambas copias de ADHa inactivadas. Las regiones aguas arriba y aguas abajo de ADHa (~0,5-1 Kb) se amplificaron a partir de ADN genómico y se clonaron en un vector TOPO separadas por un sitio NotI. El producto aguas arriba se digirió con KpnI y NotI, y el producto aguas abajo se digirió con NotI y ApaI. El vector TOPO se digirió con ApaI y KpnI y se purificó en gel, y los dos productos de PCR digeridos se ligaron en el vector TOPO. La reacción de ligadura se transformó en *E. coli*, y el ADN plasmídico de las colonias individuales se cribó en busca de la secuencia correcta de ADN. Se insertó un fragmento NotI que portaba el casete de selección URA3 de *I. orientalis* en el vector TOPO para crear los vectores pHJJ27 (orientación 1) y pHJJ28 (orientación 2). El casete de selección URA3 consiste en el gen URA3 y sus elementos reguladores flanqueados por secuencias repetitivas directas para permitir el reciclaje y la reutilización del marcador.

30

pHJJ27 se digirió con ApaI y KpnI para liberar el fragmento de integración, y el ADN linealizado resultante se transformó en la cepa 3098 de *I. orientalis* (derivado ura- de la cepa 3082), que contuvo cuatro copias de un gen exógeno que codificaba XI de *B. thetaiotaomicron*, dos copias de un gen exógeno nativo que codificaba XK, y dos copias de un gen exógeno nativo que codificaba TAL, junto con deleciones en los loci de XR y XDH. Se incorporaron genes de XI exógenos porque *I. orientalis* carece de una ruta nativa para fermentar la xilosa. La inserción en el locus de ADHa se confirmó en la cepa resultante (cepa 3274) mediante PCR en ambas uniones de integración. La cepa 3274 se cultivó durante la noche en YPD y se colocó en placas con medios FOA. El fenotipo ura- se confirmó colocando en placas con medios ScD-ura, y la retención de la integración se confirmó mediante PCR. La cepa ura- resultante se etiquetó como 3284.

40



pHJJ28 se digirió con Apal y KpnI para liberar el fragmento de integración, y el ADN linealizado resultante se transformó en la cepa 3284 de *I. orientalis*. La cepa 3085 se identificó como dos copias que contenían los loci inactivados/no de tipo natural.

- 5 Se estudió el cultivo y la fermentación de la cepa 3085 frente a la cepa original 3082 en un matraz de agitación. El medio se basó en YP (10 g/L de extracto de levadura y 20 g/L de peptona) con 0,5 g/L de MgSO<sub>4</sub>, oligoelementos, y vitaminas, y se llevó a pH 5,1 con H<sub>2</sub>SO<sub>4</sub>. Las fermentaciones se llevaron a cabo con 50 mL de medio en matraces de 125 mL a 37 °C con agitación a 100 RPM. Se demostró que la delección de ADHa tuvo poco impacto sobre la utilización de dextrosa o xilosa en los medios YP que contenían 20 g/L de dextrosa, 40 g/L de xilosa, y 9 g/L de acetato (YP20D40X9Ac), mientras aumentó la tasa de etanol en un 16% y la tasa específica de etanol en un 10%.  
10 Se observaron efectos mayores para la cepa con la inactivación mediante el uso de medios YP que contenían 60 g/L de xilosa (YP60X). En estas condiciones, la cepa 3085 tuvo una tasa de utilización de xilosa de 0,46 g/L/hr en comparación con 0,14 para la cepa original 3082, y una tasa de producción de etanol de 0,46 g/L/hr en comparación con 0,024 para la cepa original. Estos resultados indican que ADHa está implicada en el consumo de etanol, y se expresa de manera más elevada durante el cultivo con xilosa que cuando hay presente glucosa.  
15

Ejemplo 3: Caracterización de ADHb mediante el uso de inactivaciones de genes:

- Se generaron dos cepas diferentes con inactivación de ADHb mediante el uso de métodos similares a los descritos anteriormente para ADHa. La primera cepa (3859) contuvo una doble inactivación de ADHb, mientras la segunda cepa (3860) contuvo una doble inactivación de ADHb y ADHa. Las fermentaciones en matraces de agitación  
20 demostraron que las cepas 3859 y 3860 exhibieron una utilización de xilosa y un título de etanol mejorados en medios YP 20D:80X (medio basado en YP que contenía 20 g/L de dextrosa y 80 g/L de xilosa) a un pH de 4,8 frente a la cepa original 3356 y la cepa 3416 con inactivación de ADHa (discutida más adelante) (Figura 3). Ambas cepas también exhibieron una utilización de xilosa y un título de etanol mejorados en medios YP 20D:80X que contuvieron acetato a un pH de 5,1 (Figura 4). En estos experimentos se consumió toda la dextrosa antes de 19 horas. Estos  
25 resultados establecen que las cepas con inactivación de ADHa y ADHb son capaces de fermentar xilosa hasta etanol tanto en presencia como en ausencia de acetato.

Ejemplo 4: Generación de cepas adicionales con inactivación de ADHa:

Se desarrollaron tres cepas adicionales de *I. orientalis* modificadas genéticamente, en las que se inactivó el gen que codifica ADHa.

- 30 La primera cepa con inactivación de ADHa (cepa 3416) contuvo cuatro copias de un gen exógeno que codificaba XI de *B. thetaiotaomicron*, dos copias de un gen exógeno nativo que codificaba XK, y un complemento completo de enzimas PPP exógenas nativas (dos copias de TAL, TKL, RKL, y RPE. La cepa 3416 expresó niveles normales de ADH1 endógena.

- Las otras dos cepas con inactivación de ADHa (cepas 3489 y 3490) se generaron integrando dos copias extra de un gen de ADH1 que comprendían la región codificante expuesta en SEQ ID N°:5 bajo control del promotor glicolítico fuerte de TDH3 (gliceraldehído 3-fosfato deshidrogenasa) en el genoma de la cepa 3416. La secuencia del gen de ADH1 de *I. orientalis* identificada en el Ejemplo 1 se amplificó a partir de ADN genómico mediante el uso de la ADN polimerasa Pfu y cebadores que incorporaban un sitio de restricción XbaI en el extremo 5' y un sitio PstI en el extremo 3'. El fragmento purificado del gel resultante se digirió con PstI y XbaI y se ligó en el vector pHJJ7 digerido  
40 de forma similar. pHJJ7 contiene un inserto con el promotor de TDH3 de *I. orientalis*, XI de *B. thetaiotaomicron*, el terminador de PDC de *I. orientalis*, y un casete marcador de URA3 de *I. orientalis* (P<sub>TDH3</sub>-BtXI-T<sub>PDC</sub>-URA3), y el gen de XI se libera con la digestión con PstI/XbaI. Así, la ligadura dio como resultado el gen de ADH1 unido al promotor de TDH3 y un marcador URA3. El vector resultante (pHJJ60) se digirió con NotI, y el fragmento que contenía el inserto P<sub>TDH3</sub>-ADH1-T<sub>PDC</sub>-URA3 se purificó en gel. Se identificó un gen (S141G8160) homólogo a un gen de L-xilulosa reductasa de *A. monospora* en *I. orientalis*. Se ha descubierto que la enzima codificada por este gen es activa en la producción de D-xilulosa a partir de D-arabitol en una cepa de *I. orientalis* que no fermenta pentosas. La delección de este gen puede ser útil para reducir la formación de xilitol a partir de D-xilulosa por medio de la actividad de xilitol deshidrogenasa, lo que hace que S141G8160 sea un sitio de inserción beneficioso. Se amplificaron las regiones en las regiones aguas arriba y aguas abajo de S141G8160 mediante el uso de diferentes grupos de  
45 cebadores, y los fragmentos resultantes se insertaron en el vector PCR2.1-TOPO con un sitio NotI entre los fragmentos. Esta construcción se transformó en *E. coli*, y las colonias que tuvieron plásmidos con los insertos deseados se identificaron mediante PCR. Se identificó un inserto que no tenía ningún error de secuencia, y el vector con este inserto se denominó pHJJ63. pHJJ63 se digirió con NotI, y el inserto P<sub>TDH3</sub>-ADH1-T<sub>PDC</sub>-URA3 se ligó en el sitio NotI. La ligadura se transformó en *E. coli*, y se identificaron las colonias que contuvieron plásmidos con el inserto en orientación 1 (pHJJ61) u orientación 2 (pHJJ62).  
55

Los fragmentos de integración de pHJJ61 y pHJJ62 (P<sub>TDH3</sub>-ADH1-T<sub>PDC</sub>-URA3 con flancos de S141G8160) se liberaron mediante digestión con enzimas de restricción. El ADN linealizado de pHJJ61 se transformó en yACN77, el derivado ura- de la cepa 3416. Se confirmaron las colonias individuales que tenían una integración en el sitio

- 5 S141G8160 mediante PCR. Se identificaron dos cepas que contenían una copia del fragmento de integración de ADH1 (cepas yHJJ76 e yHJJ77). yHJJ76 se cultivó durante la noche en medio YPD y se colocó en placas en medio ScD-FOA para seleccionar la pérdida del gen URA3. Se purificaron colonias individuales en YPD y se transfirieron a medios ScD-ura e YPD para confirmar el fenotipo ura-. Las colonias ura- que habían conservado la copia integrada de ADH1 se identificaron mediante PCR. Estos derivados ura- de yHJJ76 se denominaron yHJJ80 e yHJJ81. El ADN linealizado de pHJJ62 se transformó en pHJJ81, y se purificaron colonias individuales en medios ScD-ura. Se confirmó mediante PCR que las cepas 3489 y 3490 contuvieron cada una dos copias del fragmento de integración de ADH1 en el sitio S141G160.
- 10 Se construyeron dos cepas adicionales que contuvieron una deleción de ADHb. Para la primera cepa, se integraron dos copias del vector de deleción de ADHb, como se describió previamente, en el derivado ura- de la cepa 4138, un mutante tolerante a etanol de la cepa 3489 obtenido mediante mutagénesis química y selección. Esta nueva cepa se denominó cepa 12053. Para la segunda cepa, se integraron dos copias del vector de deleción de ADHb, como se describió previamente, en el derivado ura- de la cepa 3489. La cepa resultante se denominó cepa 3922.
- 15 Las diversas cepas de sobreexpresión de ADH1 y/o deleción de ADHa/b generadas en este ejemplo y en los ejemplos previos se resumen en la Tabla 4 (derivados ura- y de copias individuales no incluidos).

Tabla 4: Cepas de *I. orientalis*

Nombre de la cepa	Descripción	Cepa original
3082	Cepa original con genes de XI, XK, y TAL exógenos, deleciones de XR y XDH	
3085	Deleción de ADHa	3082
3356	Cepa original con genes de XI, XK, TAL, TKL, RKI, y RPI exógenos, deleciones de XR y XDH	
3416	Deleción de ADHa	3356
3489, 3490	Deleción de ADHa Sobreexpresión de ADH1 Deleción de S141G8160	3416
3859	Deleción de ADHb	3356
3860	Deleción de ADHa Deleción de ADHb	3416
3863	Deleción de ADHa Deleción de S141G8160	3416
3922	Deleción de ADHa Deleción de ADHb Sobreexpresión de ADH1 Deleción de S141G8160	3489
4138	Cepa tolerante a etanol Deleción de ADHa	3489

Nombre de la cepa	Descripción	Cepa original
	Sobreexpresión de ADH1 Delección de S141G8160	
12053	Cepa tolerante a etanol Delección de ADHa Delección de ADHb Sobreexpresión de ADH1 Delección de S141G8160	4138

Ejemplo 5: Utilización de xilosa y título de etanol en cepas con inactivación de ADHa en medios sintéticos:

Tres de las cepas con inactivación de ADHa generadas en el Ejemplo 4 (cepas 3416, 3489, y 3490) se ensayaron por su capacidad de producir etanol a partir de un medio basado en YP con mezcla de azúcares en fermentaciones en matraces de agitación. Las tres cepas se cultivaron durante la noche en medio YPD en tubos Falcon, y estos cultivos se usaron para inocular 50 mL de medio en matraces con deflectores de 125 mL hasta una DO<sub>600</sub> inicial de 0,2. Los medios de los matraces de agitación contuvieron 20 g/L de dextrosa y 80 g/L de xilosa, pH 4,8. Los matraces se incubaron a 40 °C y 100 rpm. Se tomaron muestras para el análisis mediante HPLC después de 0, 8, 23, 32, 47, y 57 horas. Se acidificaron 500 µl de muestra con 50 µl de ácido sulfúrico, se centrifugaron, y se filtró el sobrenadante. También se midió el pH y la DO<sub>600</sub> de cada muestra.

Las tres cepas consumieron toda la dextrosa antes de 9 horas, tuvieron velocidades de crecimiento similares, y exhibieron la capacidad de fermentar xilosa hasta etanol. Sin embargo, las dos cepas que sobreexpresaron ADH1 (cepas 3489 y 3490) exhibieron una utilización de xilosa un 25% mayor y un título de etanol un 23% mayor que la cepa original que no sobreexpresó ADH1 (cepa 3416) (Figura 5). Además, las cepas que sobreexpresaban ADH1 produjeron ligeramente más arabitol y glicerol y ligeramente menos xilitol que la cepa original.

Ejemplo 6: Utilización de xilosa y título de etanol en cepas con inactivación de ADHa en medios de hidrolizado:

Tres de las cepas con inactivación de ADHa del Ejemplo 4 (cepas 3416, 3489, y 3490) se ensayaron a continuación con respecto a su capacidad de producir etanol en diversos medios de hidrolizado. Se usaron aspas de siembra de biomasa de placas de YPD para inocular matraces con deflectores de 250 mL que contenían 100 mL de medio definido (DMDX) o medio basado en YP (YPDX) que tenían 20 g/L de dextrosa y 80 g/L de xilosa y un pH ajustado a alrededor de 5,0. El medio definido contuvo urea como fuente de nitrógeno y tampón MES 0,2 M. Las células se incubaron a 250 rpm y 37 °C durante 15-24 horas, y se recogieron en la fase de crecimiento exponencial media-tardía. Los cultivos se mezclaron con un 80% de reserva de glicerol y se separaron en alícuotas de 1 mL. Se transfirieron de 50 a 400 µl de cada alícuota a 100 mL de medio en un matraz de agitación de 250 mL, se incubaron a 250 rpm y 37 °C durante 15-24 horas, y se recogieron en la fase de crecimiento exponencial media-tardía. Se recogieron muestras de 35 a 40 mL y se inocularon en recipientes de fermentación por cargas que contenían diversos medios de hidrolizado. Las muestras se recogieron a intervalos de 4 a 8 horas a lo largo de la fermentación, y se analizó la DO<sub>600</sub> mediante el uso de un espectrofotómetro y los niveles de sustratos y productos mediante el uso de análisis de HPLC.

La cepa 3416 con inactivación de ADHa exhibió un incremento del 60% del título de etanol y un incremento del 50% del consumo de xilosa frente a la cepa original 3356 en un medio DMDX de hidrolizado de rastrojo de maíz (CSH) del 30% a pH 5,8 (Figura 6). Estos resultados confirman que la inactivación de la expresión de ADHa incrementa el título de etanol en *I. orientalis*.

El incremento del título de etanol y del consumo de xilosa fue aún mayor en la cepa con inactivación de ADHa que sobreexpresó ADH1. La cepa 3489 exhibió aproximadamente un incremento del 40% en la utilización de xilosa y un incremento del 10% en el título de etanol frente a la cepa 3416 en el medio DMDX de CSH del 30% a pH 5,8 (Figura 7). La diferencia en el título de etanol entre las cepas 3416 y 3489 fue aún más notable (incremento del 30%) en un medio de hidrolizado del 15% (DMDX con un 15% de CSH, 5 g/L de ácido acético) a pH 4,9. En el medio YP 20D:80X a pH 4,9, la cepa que sobreexpresaba ADH1 mostró un incremento del 10% del título de etanol.

Ejemplo 7: Utilización de xilosa y título de etanol en cepas con inactivación de ADHb en medios de hidrolizado:

5 Las cepas con inactivación de ADHb del Ejemplo 4 (cepas 3922 y 12053), así como las cepas 3489 y 4138 con inactivación de ADHa, se ensayaron con respecto a la capacidad de producir etanol en un medio de hidrolizado de rastrojo de maíz licuado. Este medio contuvo un 20% de sólidos con una base de medio definido a pH 5,0. El hidrolizado se trató con celulasa (15 mg/g de glucano) durante 6 horas a 50 °C antes del uso en la fermentación. Los niveles iniciales de azúcares en el medio fueron 13 g/L de glucosa y 24 g/L de xilosa. Los matraces con deflectores se usaron a 100 rpm y 37 °C, y se usó cal para el ajuste del pH. La delección de ADHb proporcionó un incremento modesto pero coherente del título de etanol en estas condiciones (Figura 8).

Ejemplo 8: Incorporación de copias adicionales del gen de ADH1:

10 Se modificarán genéticamente adicionalmente una o más de las cepas de *I. orientalis* modificadas genéticamente descritas en los ejemplos anteriores incorporando copias adicionales del gen de ADH1. Las cepas resultantes pueden contener tres, cuatro, o más copias del gen de ADH1, una o más de las cuales pueden estar conectadas a un promotor fuerte.

15 Se generarán cepas con inactivación de ADHa y/o ADHb que contengan tres copias exógenas del gen de ADH1 unidas a un promotor fuerte. Se ensayará la capacidad de estas cepas de fermentar xilosa hasta etanol en diversos medios, que incluyen tanto medios sintéticos como medios de CSH. Se espera que estas cepas exhiban una utilización de xilosa y un título de etanol que sean iguales o mejores que los de las cepas correspondientes que contienen dos copias exógenas del gen de ADH1.

Ejemplo 9: Sobreexpresión de ADH1 en ausencia de inactivación de ADHa/ADHb:

20 Se desarrollará una cepa de *I. orientalis* modificada genéticamente que comprenda copias intactas de los genes de ADHa y ADHb, pero que sobreexpresé ADH1. La cepa de levadura resultante se ensayará con respecto a su capacidad de fermentar un medio que contiene xilosa hasta etanol, y se espera que muestre un consumo de xilosa y título de etanol incrementados frente a una cepa original que no sobreexpresé ADH1. También se espera que esta cepa soporte el efecto aditivo de la sobreexpresión de ADH1 y la delección de AHD2a/ADHb.

Ejemplo 10: Incorporación de modificaciones genéticas adicionales en cepas de levadura con inactivación de ADHa/ADHb y sobreexpresión de ADH1:

30 Se incorporarán una o más modificaciones genéticas adicionales en una o más de las cepas de levadura modificadas genéticamente descritas en los ejemplos previos. Estas modificaciones genéticas adicionales pueden incluir la introducción de uno o más genes exógenos de la ruta de arabinosa o genes de transportadores de azúcares, o la delección o ruptura de uno o más genes que codifican enzimas implicadas en rutas de fermentación no deseadas o en la producción de subproductos. Las cepas de levaduras resultantes se ensayarán con respecto a su capacidad de fermentar un medio que contiene xilosa hasta etanol, y se espera que una o más de estas cepas puedan exhibir un consumo de xilosa y título de etanol mejorados frente a sus cepas originales.

35 Ejemplo 11: Ensayo de *I. orientalis* modificada genéticamente en diversos medios:

Algunas de las cepas de levadura modificadas genéticamente descritas en los ejemplos previos se ensayaron con respecto a su capacidad de fermentar xilosa hasta etanol en medios de azúcares mixtos en fermentadores a escala de laboratorio. La caracterización se llevó a cabo en un reactor de cultivo por cargas de etapa simple de 2 L que contenía 1,5 L de un medio definido. Los medios para ambos protocolos contuvieron fuentes de carbono en forma de 40 20 g/L de dextrosa, 80 g/L de xilosa, y 10 g/L de arabinosa, así como 10 g/L de ácido acético glacial. Se añadieron sales en forma de 3,0 g/L de fosfato potásico monobásico y 0,5 g/L de sulfato magnésico heptahidrato. Se prepararon disoluciones de reserva de sales, oligoelementos minerales, vitaminas, y agente antiespumante, y se esterilizaron mediante filtración por separado. Los azúcares y el agua se esterilizaron en autoclave en el recipiente de fermentación, y los demás componentes se añadieron de manera aséptica al medio tras la esterilización. El primer protocolo por cargas se hizo a pH 4,95. El medio de carga se neutralizó antes de la inoculación y se mantuvo al pH seleccionado como objetivo mediante el uso de ácido sulfúrico 2 M y cal al 15%. Este medio de carga contuvo 45 2,25 g/L de sal de urea como fuente de nitrógeno. El segundo protocolo de carga se neutralizó hasta pH 5,8 con 2 g de cal al 15% e hidróxido amónico del 15%, y este último también sirvió como fuente de nitrógeno.

Otras condiciones de fermentación fueron coherentes con ambos protocolos. La temperatura se mantuvo a 37 °C, y se alcanzó una ventilación hasta una velocidad de absorción de oxígeno (OUR) objetivo de 5 mmol/L/h inyectando 50 aire a través del medio de carga a un caudal de 0,25 slpm y una velocidad de agitación constante de 450 rpm. Los niveles de oxígeno se monitorizaron mediante el uso de un electrodo de O<sub>2</sub> incorporado en el recipiente. Cada reactor de cultivo por cargas preparado se inoculó hasta una densidad celular seleccionada como objetivo de 0,15 g/L de peso seco de células con hasta 50 mL de un cultivo nocturno cultivado en un medio similar. Se tomaron 55 muestras para el análisis de HPLC al comienzo de la fermentación y una o dos veces al día posteriormente.

Los efectos de las modificaciones genéticas en estas cepas variaron con los medios usados (véanse las Figuras 9-14). En todos los casos, la dextrosa se consumió en aproximadamente 24 horas. Ambas cepas de delección (3416 y 3859) mostraron un incremento significativo del título de etanol y de la utilización de xilosa en las fermentaciones a pH 5,8 respecto de sus cepas originales (Figuras 10 y 14). Para la cepa de delección de ADHa (3416), también se observó un beneficio de la delección para la utilización de xilosa más tarde en la fermentación a pH 4,95 (Figura 9). La cepa de delección de ADHa/sobreexpresión de ADH1 (3489), por otra parte, mostró un beneficio significativo tanto en el consumo de xilosa como en el título de etanol en la fermentación a pH 4,95 (Figura 11) respecto de su cepa original 3416. La cepa 3489 también funcionó mucho mejor que su cepa original a pH mayor. Sin embargo, gran parte de este beneficio parece ser atribuible a la delección en el sitio de inserción, tal como se demuestra mediante el funcionamiento mejorado de la cepa 3863 de control del sitio de inserción (Figura 12). Las cepas que tuvieron la delección S141G8160 (3489 y 3863) exhibieron una producción inferior de xilitol con ambos medios de fermentación. En el medio de pH 4,95, los niveles de xilitol a las 114 horas fueron 3,2, 2,7, y 2,4 g/L para las cepas 3416, 3863, y 3489, respectivamente. Para el medio de pH 5,8, los valores respectivos a las 94 horas fueron 2,4, 1,8, y 1,5 g/L.

Ejemplo 12: Sobreexpresión de otros genes de ADH1 exógenos en *I. orientalis*:

Se generarán cepas que sobreexpresen fuentes alternativas de un gen de ADH1 unido a un promotor fuerte. Se ensayará la capacidad de estas cepas de fermentar xilosa hasta etanol en diversos medios, que incluyen tanto medios sintéticos como medios de CSH. Se espera que estas cepas de ADH1 puedan exhibir una utilización de xilosa y un título de etanol que sea mayor que los de las cepas originales. Las fuentes del gen de ADH1 pueden incluir *S. cerevisiae*, *P. stipitis*, *K. lactis*, y/o *C. maltosa*.

20

LISTA DE SECUENCIAS

<110> Jessen, Holly J Li, Jian

5 <120> COMPOSICIONES Y MÉTODOS PARA AUMENTAR EL TÍTULO DE ETANOL A PARTIR DE BIOMASA  
 <130> 33449.8054.WO00

<140>  
 10 <141> 22 Noviembre 2011

<150> 61/416,169  
 <151> 22 Noviembre 2010

15 <160> 15

<170> PatentIn versión 3.5

<210> 1  
 20 <211> 3179  
 <212> ADN  
 <213> *Issatchenkia orientalis*

<220>  
 25 <221> CDS  
 <222> (1052)..(2182)

<400> 1  
 gatttggacc tacaagggtc tgtaaagagt atgaacactt ctggggagga ggaatggaac  
 60  
 agtgatgacg atgatgatga agaaagtgac gaaagtaacg aaagtgatta ctattcttac  
 120  
 gatgaaggcg aagaaacaga tgatagtgag ggagcccaag agggagagga agacgaaaat  
 180  
 gaacgaatca ttgaagctct aagtagtggt gttggtgaac tcaagatgga ctctttaggt  
 240  
 aattatattc ttgaatagtt gtgtaaagcg aatatgcaaa tagatttggt ttataattat  
 300  
 gcatctcttt gaaagagggt tagaggcaaa gttcttgcac acaatattgt gattggttta  
 360  
 atgtcattct tgattttcat aaagagatta aaaaaaaaaa aaaaaaactt ataaaattga  
 420  
 gtagaacccat ttatatataa gacaaagatt gtctgtatta gtcctcaaca cactaacct  
 480  
 tacatactta gggtaaattt gctaatagag tgatatgttc atgagaactc caacgacaac  
 540

# ES 2 645 680 T3

acaaccacct atttgacaaa caaacacccat tgtcgcacgc tgcgcgcctt agaagtagaa  
600

agaaagggaa atgacattaa gagaatcata ccccggtgcc gtaacgccga aaaaatcaca  
660

cccggtcccc cacaccttaa aacctcaacc gcttaacacc gccacaccct ttctctttat  
720

aaagcgcggt tgcattactc attcttctta taaaccgcac cccccaaaac gcggaatago  
780

ttcaaccccc caatcagata tgagtttccc gggaaaccgc cttttcccga cagccccaca  
840

aggggttggt ctataaaaga ggacgttttc cccgtcatcg agattgaaga ttcttacagg  
900

cccatattatt caaattggag ttgattcttc ttgtctttac tttctttctc tctttttctt  
960

ctttttttaa tattatcttt tgtaagcctt ggttccttaa gttgaactct cttttcttgt  
1020

gactctccta tatagatacg ccttgccaaa t atg ttt gca tca acc ttc aga  
1072

Met Phe Ala Ser Thr Phe Arg  
1 5

agt caa gct gta aga gct gca aga ttt act aga ttc caa tcc act ttt  
1120

Ser Gln Ala Val Arg Ala Ala Arg Phe Thr Arg Phe Gln Ser Thr Phe  
10 15 20

gcc att cct gag aag caa atg ggt gtt atc ttc gaa act cat ggt ggt  
1168

Ala Ile Pro Glu Lys Gln Met Gly Val Ile Phe Glu Thr His Gly Gly  
25 30 35

cct tta caa tac aag gaa att cca gtt cca aaa cca aaa cca act gaa  
1216

Pro Leu Gln Tyr Lys Glu Ile Pro Val Pro Lys Pro Lys Pro Thr Glu  
40 45 50 55

att tta atc aat gtt aaa tac tct ggt gtc tgc cat acc gat tta cac  
1264

Ile Leu Ile Asn Val Lys Tyr Ser Gly Val Cys His Thr Asp Leu His  
60 65 70

gca tgg aaa ggt gac tgg cca tta cca gca aag tta ccc cta gtt ggt  
1312

Ala Trp Lys Gly Asp Trp Pro Leu Pro Ala Lys Leu Pro Leu Val Gly  
75 80 85

ggt cac gaa ggt gcg ggc att gtt gtt gcg aaa ggt tct gca gtt acc  
1360

Gly His Glu Gly Ala Gly Ile Val Val Ala Lys Gly Ser Ala Val Thr  
90 95 100

# ES 2 645 680 T3

```

aac ttt gag att ggc gat tat gct ggt att aag tgg tta aac ggt tca
1408
Asn Phe Glu Ile Gly Asp Tyr Ala Gly Ile Lys Trp Leu Asn Gly Ser
  105                110                115

tgt atg tca tgt gaa ttc tgt gaa caa ggt gat gaa tct aac tgt gaa
1456
Cys Met Ser Cys Glu Phe Cys Glu Gln Gly Asp Glu Ser Asn Cys Glu
120                125                130                135

cat gcc gat ttg agt ggt tat act cat gat ggt tct ttc caa caa tat
1504
His Ala Asp Leu Ser Gly Tyr Thr His Asp Gly Ser Phe Gln Gln Tyr
  140                145                150

gcc act gct gac gct att caa gct gca aag atc cca aag ggt acc gac
1552
Ala Thr Ala Asp Ala Ile Gln Ala Ala Lys Ile Pro Lys Gly Thr Asp
  155                160                165

tta tct gaa gtt gcg cca att tta tgt gct ggt gtt act gtc tat aaa
1600
Leu Ser Glu Val Ala Pro Ile Leu Cys Ala Gly Val Thr Val Tyr Lys
  170                175                180

gct ttg aaa act gct gat tta aga gca ggt caa tgg gtt gcg att tct
1648
Ala Leu Lys Thr Ala Asp Leu Arg Ala Gly Gln Trp Val Ala Ile Ser
  185                190                195

ggt gcc gct ggt ggt cta ggt tct ctt gct gtc caa tat gca aag gca
1696
Gly Ala Ala Gly Gly Leu Gly Ser Leu Ala Val Gln Tyr Ala Lys Ala
200                205                210                215

atg ggt cta aga gtt tta ggt atc gat ggt ggt gaa ggt aaa aag gaa
1744
Met Gly Leu Arg Val Leu Gly Ile Asp Gly Gly Glu Gly Lys Lys Glu
  220                225                230

ctt ttt gaa caa tgt ggt ggt gat gtg ttt atc gat ttc acc aga tac
1792
Leu Phe Glu Gln Cys Gly Gly Asp Val Phe Ile Asp Phe Thr Arg Tyr
  235                240                245

cca aga gat gca cct gaa aag atg gtt gct gat att aag gct gca act
1840
Pro Arg Asp Ala Pro Glu Lys Met Val Ala Asp Ile Lys Ala Ala Thr
  250                255                260

aac ggt ttg ggt cca cac ggt gtt atc aat gtc tct gtc tcc cca gct
1888
Asn Gly Leu Gly Pro His Gly Val Ile Asn Val Ser Val Ser Pro Ala
  265                270                275

gct atc tct caa tca tgt gac tat gtt aga gca act ggt aag gtt gtc
1936

```



# ES 2 645 680 T3

Ala Ile Ser Gln Ser Cys Asp Tyr Val Arg Ala Thr Gly Lys Val Val  
 280 285 290 295

ctt gtc ggt atg cca tct ggt gct gtc tgt aag tct gat gtc ttc act  
 1984

Leu Val Gly Met Pro Ser Gly Ala Val Cys Lys Ser Asp Val Phe Thr  
 300 305 310

cat gtt gtt aaa tcc tta caa att aaa ggt tct tat gtt ggt aac aga  
 2032

His Val Val Lys Ser Leu Gln Ile Lys Gly Ser Tyr Val Gly Asn Arg  
 315 320 325

gca gat acc aga gaa gct ttg gaa ttc ttt aat gaa ggt aag gtc aga  
 2080

Ala Asp Thr Arg Glu Ala Leu Glu Phe Phe Asn Glu Gly Lys Val Arg  
 330 335 340

tct cca atc aag gtt gtc cca tta tct act tta cct gaa att tac gaa  
 2128

Ser Pro Ile Lys Val Val Pro Leu Ser Thr Leu Pro Glu Ile Tyr Glu  
 345 350 355

ttg atg gag caa ggt aag att tta ggt aga tac gtt gtt gat act tct  
 2176

Leu Met Glu Gln Gly Lys Ile Leu Gly Arg Tyr Val Val Asp Thr Ser  
 360 365 370 375

aaa taa tgaagatgaa gaaaacagca aactttttat gactaccccc aaccatctaa  
 2232  
 Lys

cgatttatga tctatatata gttttctaga acatccattt atttattcac ttactcatgt  
 2292

atttatatta tataatacaa aataactaat tacaatgtgt acattttttt ttttcattac  
 2352

cataatgtat gcggtgagcc tcttgcacct tctttattag gaaatcagtt gaaaaatttc  
 2412

cggattgtct ttattattgg cccatttttt tttggtcaca cctttatttt tgtacacttc  
 2472

tcgggcaaag caaaaactat agtaccgat aggcctttat aaaactccag tgtgtatgat  
 2532

tttagttggt gtgccatcta cacgttctct tagtttcttt atcatgtcac agaaagcaag  
 2592

catgcaaacc cttacaaaaa ataacaacat acaaatgcct aaacaactgg actataatga  
 2652

tggtgagtca gttacgaaaa gagcaagtgg gttaatcga tttcgtaagg gacagtctga  
 2712

ES 2 645 680 T3

ggaagactac aattttcaaa aggagcagtt ctgggtccacg ggtccttttag tacagaatca  
2772

cacatttggtg actgaatttg ttgaaaagtt tattgaaaac acaattagtg aagattattc  
2832

aatcacagat agatcgaaaa tagaacgtga aacaatcata cacggattgg agaagctgta  
2892

ttttcaaagg gaatatgagc gatgtctaaa agatgttcaa ctattgaagg acaatatcga  
2952

taagttcaat cctaatttgg atcttaatga aaagaattta taatgagctg aattatattt  
3012

cttggatgtg catcaaaaag atccatgaga gtaacgaaaa gaaactgggg gaaatctaata  
3072

aatttacaat ttcaatatac acttctatat cctttaatgt aatggcttta taaataaaca  
3132

cgaacttcta cagcaccgac gtttcttttt cttaccagct cctcttc  
3179

<210> 2

<211> 376

<212> PRT

<213> *Issatchenkia orientalis*

5

<400> 2

Met Phe Ala Ser Thr Phe Arg Ser Gln Ala Val Arg Ala Ala Arg Phe  
1 5 10 15

Thr Arg Phe Gln Ser Thr Phe Ala Ile Pro Glu Lys Gln Met Gly Val  
20 25 30

Ile Phe Glu Thr His Gly Gly Pro Leu Gln Tyr Lys Glu Ile Pro Val  
35 40 45

Pro Lys Pro Lys Pro Thr Glu Ile Leu Ile Asn Val Lys Tyr Ser Gly  
50 55 60

Val Cys His Thr Asp Leu His Ala Trp Lys Gly Asp Trp Pro Leu Pro  
65 70 75 80

Ala Lys Leu Pro Leu Val Gly Gly His Glu Gly Ala Gly Ile Val Val  
85 90 95

Ala Lys Gly Ser Ala Val Thr Asn Phe Glu Ile Gly Asp Tyr Ala Gly  
100 105 110

ES 2 645 680 T3

Ile Lys Trp Leu Asn Gly Ser Cys Met Ser Cys Glu Phe Cys Glu Gln  
 115 120 125

Gly Asp Glu Ser Asn Cys Glu His Ala Asp Leu Ser Gly Tyr Thr His  
 130 135 140

Asp Gly Ser Phe Gln Gln Tyr Ala Thr Ala Asp Ala Ile Gln Ala Ala  
 145 150 155 160

Lys Ile Pro Lys Gly Thr Asp Leu Ser Glu Val Ala Pro Ile Leu Cys  
 165 170 175

Ala Gly Val Thr Val Tyr Lys Ala Leu Lys Thr Ala Asp Leu Arg Ala  
 180 185 190

Gly Gln Trp Val Ala Ile Ser Gly Ala Ala Gly Gly Leu Gly Ser Leu  
 195 200 205

Ala Val Gln Tyr Ala Lys Ala Met Gly Leu Arg Val Leu Gly Ile Asp  
 210 215 220

Gly Gly Glu Gly Lys Lys Glu Leu Phe Glu Gln Cys Gly Gly Asp Val  
 225 230 235 240

Phe Ile Asp Phe Thr Arg Tyr Pro Arg Asp Ala Pro Glu Lys Met Val  
 245 250 255

Ala Asp Ile Lys Ala Ala Thr Asn Gly Leu Gly Pro His Gly Val Ile  
 260 265 270

Asn Val Ser Val Ser Pro Ala Ala Ile Ser Gln Ser Cys Asp Tyr Val  
 275 280 285

Arg Ala Thr Gly Lys Val Val Leu Val Gly Met Pro Ser Gly Ala Val  
 290 295 300

Cys Lys Ser Asp Val Phe Thr His Val Val Lys Ser Leu Gln Ile Lys  
 305 310 315 320

Gly Ser Tyr Val Gly Asn Arg Ala Asp Thr Arg Glu Ala Leu Glu Phe  
 325 330 335

Phe Asn Glu Gly Lys Val Arg Ser Pro Ile Lys Val Val Pro Leu Ser  
 340 345 350

Thr Leu Pro Glu Ile Tyr Glu Leu Met Glu Gln Gly Lys Ile Leu Gly  
 355 360 365

Arg Tyr Val Val Asp Thr Ser Lys  
 370 375

<210> 3

# ES 2 645 680 T3

<211> 3134  
<212> ADN  
<213> *Issatchenkia orientalis*

5 <220>  
<221> CDS  
<222> (1001)..(2134)

<400> 3

atgtatttgg agatttcgaa aagagtttgt atagagtctg taattgggtg tgtatttcaa  
60

gaccacttt aaactgcgcc attaggagag ggagaggggg gggggggggg ggaagacggt  
120

gaagtgtata caggatcgaa gaatagaagt tgtgtgtgtg ttttattacc cgtttcgatg  
180

ggattcccag aagtggatac tatactgtct gcaatgcact aactctaaa aaagtattat  
240

acattacat acattagcaa atcaccaata ctctgcactg tttcagtgtg tgcacattgc  
300

taccaattg ggaaattgca gggaaaatga gacccccctt ccattccgta ttacgtaaga  
360

caatatcagg gctgccgaat toggcagaaa agccgagccg gccgagtcct cttgcacgga  
420

gtgtgtccga aaagggcagc tctgcagtgg gggagaggag gtcgcacgtc tatgcggtgt  
480

tggcatggcc tgtgcgtgta cctgtcccct ccttgggcat cccccactgc gcgccttctc  
540

cattgggggc tgcgggcact ccgcgccggt aatacaggag gggggggggg aaagcttaag  
600

attagagcgg gtacagtcag tgggtgtatt gacccattt ctgtcagtat aaacccccg  
660

ttgagccgcc ggtttggtg tttatggata aaatTTTTT tccccgatg gagaagattg  
720

# ES 2 645 680 T3

agggggagaaa ggaatgggaa aaaggccaga gccatctcca cagcgggaatc cgaccgttaa  
780

tggggtgaaa cacccccacc aggtagagca ggaagaatgg ggaacaagg tggagagatg  
840

gtcattgttg ggaatagtgg gaaaatgagg gggaagagaa tgactataaa atgggaaggg  
900

ggccaagt atccaagcag tccatttaga gaagggaaaa taaagctata gatagaaacc  
960

aaccaaacia ccaacaatt aaacaaacia ttaaacgaac atg tta tcc aag acc  
1015

Met Leu Ser Lys Thr  
1 5

atc act gct gca ttg agg ggc aat aca act cgt act gca ttc aga atc  
1063

Ile Thr Ala Ala Leu Arg Gly Asn Thr Thr Arg Thr Ala Phe Arg Ile  
10 15 20

aat gcc att aga agt tta gcg atc cca gct att cca gag aca caa aag  
1111

Asn Ala Ile Arg Ser Leu Ala Ile Pro Ala Ile Pro Glu Thr Gln Lys  
25 30 35

ggg gtt atc ttt tat gag aac gga ggt gaa cta ttt tac aag gac att  
1159

Gly Val Ile Phe Tyr Glu Asn Gly Gly Glu Leu Phe Tyr Lys Asp Ile  
40 45 50

cca gtt cca aag cca aag cca aat gag att ttg gtg aat gtc aag tat  
1207

Pro Val Pro Lys Pro Lys Pro Asn Glu Ile Leu Val Asn Val Lys Tyr  
55 60 65

tct ggt gtt tgt cat acc gat tta cac gca tgg aaa ggt gac tgg cct  
1255

Ser Gly Val Cys His Thr Asp Leu His Ala Trp Lys Gly Asp Trp Pro  
70 75 80 85

ttg gcg acc aag ttg cca ttg gtt ggt gga cat gaa ggt gcc gga gtt  
1303

Leu Ala Thr Lys Leu Pro Leu Val Gly Gly His Glu Gly Ala Gly Val  
90 95 100

ggt gtt gct aag ggg gac aat gtc acc aac ttt gaa att ggc gat tat  
1351

Val Val Ala Lys Gly Asp Asn Val Thr Asn Phe Glu Ile Gly Asp Tyr  
105 110 115

gcc ggt atc aag tgg ttg aat ggt tca tgt atg ggg tgt gaa ttt tgc  
1399

Ala Gly Ile Lys Trp Leu Asn Gly Ser Cys Met Gly Cys Glu Phe Cys  
120 125 130

# ES 2 645 680 T3

caa caa ggt gca gag cca aac tgt cca cag gcc gac ttg agt ggt tac  
 1447  
 Gln Gln Gly Ala Glu Pro Asn Cys Pro Gln Ala Asp Leu Ser Gly Tyr  
     135                            140                            145

acc cat gac ggg tcc ttt caa caa tat gcc act gcc gat gct gtt cag  
 1495  
 Thr His Asp Gly Ser Phe Gln Gln Tyr Ala Thr Ala Asp Ala Val Gln  
 150                            155                            160                            165

gca gcc aag att cct cag ggc act gat ttg gct caa gtt gcg cca att  
 1543  
 Ala Ala Lys Ile Pro Gln Gly Thr Asp Leu Ala Gln Val Ala Pro Ile  
                             170                            175                            180

tta tgt gca ggt att act gtc tat aag gct tta aag act gca gaa tta  
 1591  
 Leu Cys Ala Gly Ile Thr Val Tyr Lys Ala Leu Lys Thr Ala Glu Leu  
                             185                            190                            195

aga cca ggt caa tgg gtt gcc att tct ggt gct gct gga ggt tta ggt  
 1639  
 Arg Pro Gly Gln Trp Val Ala Ile Ser Gly Ala Ala Gly Gly Leu Gly  
                             200                            205                            210

tct ctt gct gtt caa tat gcc aag gcc atg ggt ttg aga gtt ttg ggt  
 1687  
 Ser Leu Ala Val Gln Tyr Ala Lys Ala Met Gly Leu Arg Val Leu Gly  
                             215                            220                            225

att gat ggt ggt gag gag aag ggc aag ttt gca aag tct ctt gga gct  
 1735  
 Ile Asp Gly Gly Glu Glu Lys Gly Lys Phe Ala Lys Ser Leu Gly Ala  
 230                            235                            240                            245

gaa gtt ttc att gat ttc acc aaa tcc aag gac att gtc aag gat acc  
 1783  
 Glu Val Phe Ile Asp Phe Thr Lys Ser Lys Asp Ile Val Lys Asp Ile  
                             250                            255                            260

caa gag gcc acc aat ggt ggt cca cat ggt gtc att aat gtt tct gtt  
 1831  
 Gln Glu Ala Thr Asn Gly Gly Pro His Gly Val Ile Asn Val Ser Val  
                             265                            270                            275

tct cca gct gct att tct caa agt acc cag tat gtc aga acc ttg ggt  
 1879  
 Ser Pro Ala Ala Ile Ser Gln Ser Thr Gln Tyr Val Arg Thr Leu Gly  
                             280                            285                            290

aag gtt gtc ctt gtt gga tta cca gcg cat gct gta tgc gag tct tcg  
 1927  
 Lys Val Val Leu Val Gly Leu Pro Ala His Ala Val Cys Glu Ser Ser  
                             295                            300                            305

gtt ttc gac cat gtt gtc aag tcg att caa att aga ggc tct tat gtt  
 1975  
 Val Phe Asp His Val Val Lys Ser Ile Gln Ile Arg Gly Ser Tyr Val

# ES 2 645 680 T3

```

310                               315                               320                               325
ggt aac aqg qaa gat act agt qag gct att gat ttt ttc acc aqg ggt
2023
Gly Asn Arg Glu Asp Thr Ser Glu Ala Ile Asp Phe Phe Thr Arg Gly
                               330                               335                               340
tta gtg aag tca cca att aag att gtt ggt ttg agt gag ttg cca aag
2071
Leu Val Lys Ser Pro Ile Lys Ile Val Gly Leu Ser Glu Leu Pro Lys
                               345                               350                               355
atc tat gaa ttg atg gag caa ggt aag att tta ggc aga tat gtt gtt
2119
Ile Tyr Glu Leu Met Glu Gln Gly Lys Ile Leu Gly Arg Tyr Val Val
                               360                               365                               370
gac act tog aaa tga tgggctgact tgggtgtact ggtgtgacgt ttttatgtgt
2174
Asp Thr Ser Lys
                               375
atattgatat gcatggggga tgtatagtga tgaggagtag agtatataac gaaatgaaat
2234
gaaataatat gatatgataa gataagatga gatcaaatac gataatataa gatgcgacat
2294
gaggagtcca atgtagcata ctacacgatg ctgcagtaca actctgatac gctagactat
2354
actatacaaa actgtagtag actatacgtt agtgtagtag ccagaaacaa cactgcttta
2414
tagtacaata caactctata atactatagt atactatgcc aaaccacgta ataccataat
2474
atgctccacg acatggtaga atgtgctata cttcatacta ttataccata tatactccga
2534
tatattattg atatactatt ttatactata ataccatacc acacaacact acattacaac
2594
gagcaacctt accataaatg tcagttatgt ggcccggaga ctctctcgag gagcgtgttc
2654
accctgttgt agacgttctg cacatcctct ccgagcaggg cacg-gctcc catagtggga
2714
ggggcctctt ccaagggcga cccgcggcgc cccgcaccaa gaagcgcctg ttccttgagc
2774
gcatgtgcaa tattgagaag ggtgtctatg ctgcgaagaa cgggtgtctgt gtcggcagca
2834
gcagcagcgg cgtctgctcc ctggcgggaa cgtgtcttcc ccgctaaggg gagcacagca
2894

```

ES 2 645 680 T3

agaatatcat gtaatgcagc aagagcattc tgagttgaag tatcgatttt cgatgccata  
2954

ttgtatgtgt attgtattaa gtgtgtattg tottaagtgt gtaagagaca tttatttgtg  
3014

tcaacaatag cgacgccact gaaaacctca aatatcgtat ttattaatcc ccttcccccc  
3074

agcgcagatc gtcccgtcga tttctattgt ttgggcatta tcagcgcgcg gacggcgacg  
3134

<210> 4

<211> 377

5 <212> PRT

<213> *Issatchenkia orientalis*

<400> 4

Met Leu Ser Lys Thr Ile Thr Ala Ala Leu Arg Gly Asn Thr Thr Arg  
1 5 10 15

Thr Ala Phe Arg Ile Asn Ala Ile Arg Ser Leu Ala Ile Pro Ala Ile  
20 25 30

Pro Glu Thr Gln Lys Gly Val Ile Phe Tyr Glu Asn Gly Gly Glu Leu  
35 40 45

Phe Tyr Lys Asp Ile Pro Val Pro Lys Pro Lys Pro Asn Glu Ile Leu  
50 55 60

Val Asn Val Lys Tyr Ser Gly Val Cys His Thr Asp Leu His Ala Trp  
65 70 75 80

Lys Gly Asp Trp Pro Leu Ala Thr Lys Leu Pro Leu Val Gly Gly His  
85 90 95

Glu Gly Ala Gly Val Val Val Ala Lys Gly Asp Asn Val Thr Asn Phe  
100 105 110

Glu Ile Gly Asp Tyr Ala Gly Ile Lys Trp Leu Asn Gly Ser Cys Met  
115 120 125

Gly Cys Glu Phe Cys Gln Gln Gly Ala Glu Pro Asn Cys Pro Gln Ala  
130 135 140

Asp Leu Ser Gly Tyr Thr His Asp Gly Ser Phe Gln Gln Tyr Ala Thr  
145 150 155 160



# ES 2 645 680 T3

Ala Asp Ala Val Gln Ala Ala Lys Ile Pro Gln Gly Thr Asp Leu Ala  
165 170 175

Gln Val Ala Pro Ile Leu Cys Ala Gly Ile Thr Val Tyr Lys Ala Leu  
180 185 190

Lys Thr Ala Glu Leu Arg Pro Gly Gln Trp Val Ala Ile Ser Gly Ala  
195 200 205

Ala Gly Gly Leu Gly Ser Leu Ala Val Gln Tyr Ala Lys Ala Met Gly  
210 215 220

Leu Arg Val Leu Gly Ile Asp Gly Gly Glu Glu Lys Gly Lys Phe Ala  
225 230 235 240

Lys Ser Leu Gly Ala Glu Val Phe Ile Asp Phe Thr Lys Ser Lys Asp  
245 250 255

Ile Val Lys Asp Ile Gln Glu Ala Thr Asn Gly Gly Pro His Gly Val  
260 265 270

Ile Asn Val Ser Val Ser Pro Ala Ala Ile Ser Gln Ser Thr Gln Tyr  
275 280 285

Val Arg Thr Leu Gly Lys Val Val Leu Val Gly Leu Pro Ala His Ala  
290 295 300

Val Cys Glu Ser Ser Val Phe Asp His Val Val Lys Ser Ile Gln Ile  
305 310 315 320

Arg Gly Ser Tyr Val Gly Asn Arg Glu Asp Thr Ser Glu Ala Ile Asp  
325 330 335

Phe Phe Thr Arg Gly Leu Val Lys Ser Pro Ile Lys Ile Val Gly Leu  
340 345 350

Ser Glu Leu Pro Lys Ile Tyr Glu Leu Met Glu Gln Gly Lys Ile Leu  
355 360 365

Gly Arg Tyr Val Val Asp Thr Ser Lys  
370 375

<210> 5  
<211> 1053  
5 <212> ADN  
<213> *Issatchenkia orientalis*

<220>  
<221> CDS  
10 <222> (1)..(1053)

<400> 5

# ES 2 645 680 T3

```

atg tct tac gaa atc cca caa aca caa aag gcc tgt gtc ttt tac gaa
48
Met Ser Tyr Glu Ile Pro Gln Thr Gln Lys Ala Cys Val Phe Tyr Glu
1           5           10           15

aac ggc ggc cca atc aca tac aag gac att cca gtt cca aag cca aaa
96
Asn Gly Gly Pro Ile Thr Tyr Lys Asp Ile Pro Val Pro Lys Pro Lys
          20           25           30

cct act gag att tta gtc aag gtt ctg tac tct ggt gtc tgc cac acc
144
Pro Thr Glu Ile Leu Val Lys Val Leu Tyr Ser Gly Val Cys His Thr
          35           40           45

gac ttg cac gca tgg aag ggt gac tgg cct cta gct acc aag ttg cca
192
Asp Leu His Ala Trp Lys Gly Asp Trp Pro Leu Ala Thr Lys Leu Pro
          50           55           60

ttg gtt ggt ggt cac gaa ggt gcc ggt gtt gtt gtt gcc aag ggt gaa
240
Leu Val Gly Gly His Glu Gly Ala Gly Val Val Val Ala Lys Gly Glu
65           70           75           80

aac gtc acc tct ttt gag att ggt gat tac gca ggt atc aag tgg ttg
288
Asn Val Thr Ser Phe Glu Ile Gly Asp Tyr Ala Gly Ile Lys Trp Leu
          85           90           95

aat ggt tca tgt atg ggt tgt gaa ttc tgt gaa caa ggt gct gaa cca
336
Asn Gly Ser Cys Met Gly Cys Glu Phe Cys Glu Gln Gly Ala Glu Pro
          100          105          110

aac tgt cct aag gcc gac ttg agt ggt tac acc cac gac ggt tcc ttc
384
Asn Cys Pro Lys Ala Asp Leu Ser Gly Tyr Thr His Asp Gly Ser Phe
          115          120          125

caa cag tat gct act gct gac gct att caa gct gca cac atc tcc aag
432
Gln Gln Tyr Ala Thr Ala Asp Ala Ile Gln Ala Ala His Ile Ser Lys
          130          135          140

gaa acc gac ttg gct ggt gtt gct cca atc ttg tgt gca ggt gtc act
480

```

ES 2 645 680 T3

Glu Thr Asp Leu Ala Gly Val Ala Pro Ile Leu Cys Ala Gly Val Thr  
145 150 155 160

gtc tac aag gct tta aag act gca gac ctt aga gca ggt gaa tgg gtt  
528

Val Tyr Lys Ala Leu Lys Thr Ala Asp Leu Arg Ala Gly Glu Trp Val  
165 170 175

tgt att tcc ggt gca gct ggt ggt tta ggt tct ctt gct att caa tat  
576

Cys Ile Ser Gly Ala Ala Gly Gly Leu Gly Ser Leu Ala Ile Gln Tyr  
180 185 190

gca aag gct atg ggt ctg aga gtt gtt ggt att gac ggt ggt gac gaa  
624

Ala Lys Ala Met Gly Leu Arg Val Val Gly Ile Asp Gly Gly Asp Glu  
195 200 205

aag aag gaa ttg tgt aaa tcc ctt ggt gct gaa gca ttt att gat ttc  
672

Lys Lys Glu Leu Cys Lys Ser Leu Gly Ala Glu Ala Phe Ile Asp Phe  
210 215 220

aca aag acc aag gat atc gtc aag gct gtc caa gag gca acc aat ggt  
720

Thr Lys Thr Lys Asp Ile Val Lys Ala Val Gln Glu Ala Thr Asn Gly  
225 230 235 240

ggt cca cat ggt gtc atc aat gtc tct gtc tct gaa gct gca att tct  
768

Gly Pro His Gly Val Ile Asn Val Ser Val Ser Glu Ala Ala Ile Ser  
245 250 255

caa tct tgt gaa tac gtt aga cct cta ggt aag gtt gtt ctt gtt ggt  
816

Gln Ser Cys Glu Tyr Val Arg Pro Leu Gly Lys Val Val Leu Val Gly  
260 265 270

tta cca gca ggc gca caa gtc aaa act ggt gtc ttt gaa gcc gtt gtc  
864

Leu Pro Ala Gly Ala Gln Val Lys Thr Gly Val Phe Glu Ala Val Val  
275 280 285

aag tct att gaa att aag ggt tct tat gtc ggt aac aga aag gat acc  
912

Lys Ser Ile Glu Ile Lys Gly Ser Tyr Val Gly Asn Arg Lys Asp Thr  
290 295 300

gcc gaa gca ctt gac ttc tac act aga ggc ttg gtc aag tct cca ttc  
960

Ala Glu Ala Leu Asp Phe Tyr Thr Arg Gly Leu Val Lys Ser Pro Phe  
305 310 315 320

aag att gtc ggt tta tcc gaa ttg cca aaa gtc ttt gaa ctc atg gaa  
1008

Lys Ile Val Gly Leu Ser Glu Leu Pro Lys Val Phe Glu Leu Met Glu  
325 330 335

cag ggt aag att tta ggt aga atg gtc tta gac acc tcc aaa taa  
1053

Gln Gly Lys Ile Leu Gly Arg Met Val Leu Asp Thr Ser Lys  
340 345 350

- 5 <210> 6
- <211> 350
- <212> PRT
- <213> *Issatchenkia orientalis*

# ES 2 645 680 T3

<400> 6

Met Ser Tyr Glu Ile Pro Gln Thr Gln Lys Ala Cys Val Phe Tyr Glu  
 1 5 10 15

Asn Gly Gly Pro Ile Thr Tyr Lys Asp Ile Pro Val Pro Lys Pro Lys  
 20 25 30

Pro Thr Glu Ile Leu Val Lys Val Leu Tyr Ser Gly Val Cys His Thr  
 35 40 45

Asp Leu His Ala Trp Lys Gly Asp Trp Pro Leu Ala Thr Lys Leu Pro  
 50 55 60

Leu Val Gly Gly His Glu Gly Ala Gly Val Val Val Ala Lys Gly Glu  
 65 70 75 80

Asn Val Thr Ser Phe Glu Ile Gly Asp Tyr Ala Gly Ile Lys Trp Leu  
 85 90 95

Asn Gly Ser Cys Met Gly Cys Glu Phe Cys Glu Gln Gly Ala Glu Pro  
 100 105 110

Asn Cys Pro Lys Ala Asp Leu Ser Gly Tyr Thr His Asp Gly Ser Phe  
 115 120 125

Gln Gln Tyr Ala Thr Ala Asp Ala Ile Gln Ala Ala His Ile Ser Lys  
 130 135 140

Glu Thr Asp Leu Ala Gly Val Ala Pro Ile Leu Cys Ala Gly Val Thr  
 145 150 155 160

Val Tyr Lys Ala Leu Lys Thr Ala Asp Leu Arg Ala Gly Glu Trp Val  
 165 170 175

Cys Ile Ser Gly Ala Ala Gly Gly Leu Gly Ser Leu Ala Ile Gln Tyr

# ES 2 645 680 T3

	180	185	190
Ala Lys Ala Met Gly Leu Arg Val Val Gly Ile Asp Gly Gly Asp Glu	195	200	205
Lys Lys Glu Leu Cys Lys Ser Leu Gly Ala Glu Ala Phe Ile Asp Phe	210	215	220
Thr Lys Thr Lys Asp Ile Val Lys Ala Val Gln Glu Ala Thr Asn Gly	225	230	235
Gly Pro His Gly Val Ile Asn Val Ser Val Ser Glu Ala Ala Ile Ser	245	250	255
Gln Ser Cys Glu Tyr Val Arg Pro Leu Gly Lys Val Val Leu Val Gly	260	265	270
Leu Pro Ala Gly Ala Gln Val Lys Thr Gly Val Phe Glu Ala Val Val	275	280	285
Lys Ser Ile Glu Ile Lys Gly Ser Tyr Val Gly Asn Arg Lys Asp Thr	290	295	300
Ala Glu Ala Leu Asp Phe Tyr Thr Arg Gly Leu Val Lys Ser Pro Phe	305	310	315
Lys Ile Val Gly Leu Ser Glu Leu Pro Lys Val Phe Glu Leu Met Glu	325	330	335
Gln Gly Lys Ile Leu Gly Arg Met Val Leu Asp Thr Ser Lys	340	345	350

- <210> 7
- <211> 23
- 5 <212> ADN
- <213> Secuencia Artificial
  
- <220>
- 10 <223> Cebador PCR
  
- <400> 7
- cttgtgaata cgtttagacct cta
- 23
  
- <210> 8
- 15 <211> 20
- <212> ADN
- <213> Secuencia Artificial
  
- <220>
- 20 <223> Cebador PCR
  
- <400> 8
- gagacttgac caagcctcta
- 20

<210> 9  
 <211> 23  
 <212> ADN  
 <213> Secuencia Artificial  
 5  
 <220>  
 <223> Cebador PCR  
 <400> 9  
 catgtgacta tgtttagagca act  
 10 23  
 <210> 10  
 <211> 21  
 <212> ADN  
 15 <213> Secuencia Artificial  
 <220>  
 <223> Cebador PCR  
 20 <400> 10  
 ggagatctga cottaccttc a  
 21  
 <210> 11  
 <211> 20  
 25 <212> ADN  
 <213> Secuencia Artificial  
 <220>  
 <223> Cebador PCR  
 30 <400> 11  
 cccagtatgt cagaaccttg  
 20  
 <210> 12  
 35 <211> 20  
 <212> ADN  
 <213> Secuencia Artificial  
 <220>  
 40 <223> Cebador PCR  
 <400> 12  
 gtgacttcac taaaccctg  
 20  
 45 <210> 13  
 <211> 348  
 <212> PRT  
 <213> Saccharomyces cerevisiae  
 50 <400> 13

# ES 2 645 680 T3

Met Ser Ile Pro Glu Thr Gln Lys Gly Val Ile Phe Tyr Glu Ser His  
 1 5 10 15

Gly Lys Leu Glu Tyr Lys Asp Ile Pro Val Pro Lys Pro Lys Ala Asn  
 20 25 30

Glu Leu Leu Ile Asn Val Lys Tyr Ser Gly Val Cys His Thr Asp Leu  
 35 40 45

His Ala Trp His Gly Asp Trp Pro Leu Pro Val Lys Leu Pro Leu Val  
 50 55 60

Gly Gly His Glu Gly Ala Gly Val Val Val Gly Met Gly Glu Asn Val  
 65 70 75 80

Lys Gly Trp Lys Ile Gly Asp Tyr Ala Gly Ile Lys Trp Leu Asn Gly  
 85 90 95

Ser Cys Met Ala Cys Glu Tyr Cys Glu Leu Gly Asn Glu Ser Asn Cys  
 100 105 110

Pro His Ala Asp Leu Ser Gly Tyr Thr His Asp Gly Ser Phe Gln Gln  
 115 120 125

Tyr Ala Thr Ala Asp Ala Val Gln Ala Ala His Ile Pro Gln Gly Thr  
 130 135 140

Asp Leu Ala Gln Val Ala Pro Ile Leu Cys Ala Gly Ile Thr Val Tyr  
 145 150 155 160

Lys Ala Leu Lys Ser Ala Asn Leu Met Ala Gly His Trp Val Ala Ile  
 165 170 175

Ser Gly Ala Ala Gly Gly Leu Gly Ser Leu Ala Val Gln Tyr Ala Lys  
 180 185 190

# ES 2 645 680 T3

Ala Met Gly Tyr Arg Val Leu Gly Ile Asp Gly Gly Glu Gly Lys Glu  
 195 200 205

Glu Leu Phe Arg Ser Ile Gly Gly Glu Val Phe Ile Asp Phe Thr Lys  
 210 215 220

Glu Lys Asp Ile Val Gly Ala Val Leu Lys Ala Thr Asp Gly Gly Ala  
 225 230 235 240

His Gly Val Ile Asn Val Ser Val Ser Glu Ala Ala Ile Glu Ala Ser  
 245 250 255

Thr Arg Tyr Val Arg Ala Asn Gly Thr Thr Val Leu Val Gly Met Pro  
 260 265 270

Ala Gly Ala Lys Cys Cys Ser Asp Val Phe Asn Gln Val Val Lys Ser  
 275 280 285

Ile Ser Ile Val Gly Ser Tyr Val Gly Asn Arg Ala Asp Thr Arg Glu  
 290 295 300

Ala Leu Asp Phe Phe Ala Arg Gly Leu Val Lys Ser Pro Ile Lys Val  
 305 310 315 320

Val Gly Leu Ser Thr Leu Pro Glu Ile Tyr Glu Lys Met Glu Lys Gly  
 325 330 335

Gln Ile Val Gly Arg Tyr Val Val Asp Thr Ser Lys  
 340 345

<210> 14

<211> 348

5 <212> PRT

<213> *Saccharomyces cerevisiae*

<400> 14

Met Ser Ile Pro Glu Thr Gln Lys Ala Ile Ile Phe Tyr Glu Ser Asn  
 1 5 10 15

Gly Lys Leu Glu His Lys Asp Ile Pro Val Pro Lys Pro Lys Pro Asn  
 20 25 30

Glu Leu Leu Ile Asn Val Lys Tyr Ser Gly Val Cys His Thr Asp Leu



# ES 2 645 680 T3

35	40	45																					
His	Ala	Trp	His	Gly	Asp	Trp	Pro	Leu	Pro	Thr	Lys	Leu	Pro	Leu	Val								
50						55					60												
Gly	Gly	His	Glu	Gly	Ala	Gly	Val	Val	Val	Gly	Met	Gly	Glu	Asn	Val								
65					70					75					80								
Lys	Gly	Trp	Lys	Ile	Gly	Asp	Tyr	Ala	Gly	Ile	Lys	Trp	Leu	Asn	Gly								
				85					90					95									
Ser	Cys	Met	Ala	Cys	Glu	Tyr	Cys	Glu	Leu	Gly	Asn	Glu	Ser	Asn	Cys								
			100					105					110										
Pro	His	Ala	Asp	Leu	Ser	Gly	Tyr	Thr	His	Asp	Gly	Ser	Phe	Gln	Glu								
		115						120					125										
Tyr	Ala	Thr	Ala	Asp	Ala	Val	Gln	Ala	Ala	His	Ile	Pro	Gln	Gly	Thr								
	130					135					140												
Asp	Leu	Ala	Glu	Val	Ala	Pro	Ile	Leu	Cys	Ala	Gly	Ile	Thr	Val	Tyr								
145					150					155					160								
Lys	Ala	Leu	Lys	Ser	Ala	Asn	Leu	Arg	Ala	Gly	His	Trp	Ala	Ala	Ile								
				165					170						175								
Ser	Gly	Ala	Ala	Gly	Gly	Leu	Gly	Ser	Leu	Ala	Val	Gln	Tyr	Ala	Lys								
			180					185					190										
Ala	Met	Gly	Tyr	Arg	Val	Leu	Gly	Ile	Asp	Gly	Gly	Pro	Gly	Lys	Glu								
		195					200					205											
Glu	Leu	Phe	Thr	Ser	Leu	Gly	Gly	Glu	Val	Phe	Ile	Asp	Phe	Thr	Lys								
		210				215					220												
Glu	Lys	Asp	Ile	Val	Ser	Ala	Val	Val	Lys	Ala	Thr	Asn	Gly	Gly	Ala								
225					230					235					240								
His	Gly	Ile	Ile	Asn	Val	Ser	Val	Ser	Glu	Ala	Ala	Ile	Glu	Ala	Ser								
				245					250					255									
Thr	Arg	Tyr	Cys	Arg	Ala	Asn	Gly	Thr	Val	Val	Leu	Val	Gly	Leu	Pro								
			260					265					270										

ES 2 645 680 T3

Ala Gly Ala Lys Cys Ser Ser Asp Val Phe Asn His Val Val Lys Ser  
 275 280 285

Ile Ser Ile Val Gly Ser Tyr Val Gly Asn Arg Ala Asp Thr Arg Glu  
 290 295 300

Ala Leu Asp Phe Phe Ala Arg Gly Leu Val Lys Ser Pro Ile Lys Val  
 305 310 315 320

Val Gly Leu Ser Ser Leu Pro Glu Ile Tyr Glu Lys Met Glu Lys Gly  
 325 330 335

Gln Ile Ala Gly Arg Tyr Val Val Asp Thr Ser Lys  
 340 345

<210> 15

<211> 375

5 <212> PRT

<213> *Saccharomyces cerevisiae*

<400> 15

Met Leu Arg Thr Ser Thr Leu Phe Thr Arg Arg Val Gln Pro Ser Leu  
 1 5 10 15

Phe Ser Arg Asn Ile Leu Arg Leu Gln Ser Thr Ala Ala Ile Pro Lys  
 20 25 30

Thr Gln Lys Gly Val Ile Phe Tyr Glu Asn Lys Gly Lys Leu His Tyr  
 35 40 45

Lys Asp Ile Pro Val Pro Glu Pro Lys Pro Asn Glu Ile Leu Ile Asn  
 50 55 60

Val Lys Tyr Ser Gly Val Cys His Thr Asp Leu His Ala Trp His Gly  
 65 70 75 80

Asp Trp Pro Leu Pro Val Lys Leu Pro Leu Val Gly Gly His Glu Gly  
 85 90 95

Ala Gly Val Val Val Lys Leu Gly Ser Asn Val Lys Gly Trp Lys Val  
 100 105 110

Gly Asp Leu Ala Gly Ile Lys Trp Leu Asn Gly Ser Cys Met Thr Cys  
 115 120 125

# ES 2 645 680 T3

Glu Phe Cys Glu Ser Gly His Glu Ser Asn Cys Pro Asp Ala Asp Leu  
 130 135 140

Ser Gly Tyr Thr His Asp Gly Ser Phe Gln Gln Phe Ala Thr Ala Asp  
 145 150 155 160

Ala Ile Gln Ala Ala Lys Ile Gln Gln Gly Thr Asp Leu Ala Glu Val  
 165 170 175

Ala Pro Ile Leu Cys Ala Gly Val Thr Val Tyr Lys Ala Leu Lys Glu  
 180 185 190

Ala Asp Leu Lys Ala Gly Asp Trp Val Ala Ile Ser Gly Ala Ala Gly  
 195 200 205

Gly Leu Gly Ser Leu Ala Val Gln Tyr Ala Thr Ala Met Gly Tyr Arg  
 210 215 220

Val Leu Gly Ile Asp Ala Gly Glu Glu Lys Glu Lys Leu Phe Lys Lys  
 225 230 235 240

Leu Gly Gly Glu Val Phe Ile Asp Phe Thr Lys Thr Lys Asn Met Val  
 245 250 255

Ser Asp Ile Gln Glu Ala Thr Lys Gly Gly Pro His Gly Val Ile Asn  
 260 265 270

Val Ser Val Ser Glu Ala Ala Ile Ser Leu Ser Thr Glu Tyr Val Arg  
 275 280 285

Pro Cys Gly Thr Val Val Leu Val Gly Leu Pro Ala Asn Ala Tyr Val  
 290 295 300

Lys Ser Glu Val Phe Ser His Val Val Lys Ser Ile Asn Ile Lys Gly  
 305 310 315 320

Ser Tyr Val Gly Asn Arg Ala Asp Thr Arg Glu Ala Leu Asp Phe Phe  
 325 330 335

Ser Arg Gly Leu Ile Lys Ser Pro Ile Lys Ile Val Gly Leu Ser Glu  
 340 345 350

Leu Pro Lys Val Tyr Asp Leu Met Glu Lys Gly Lys Ile Leu Gly Arg  
 355 360 365

Tyr Val Val Asp Thr Ser Lys  
 370 375

## REIVINDICACIONES

- 5 1. Una célula de levadura modificada genéticamente que sobreexpresa un polipéptido que comprende una secuencia de aminoácidos con al menos un 85% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°: 6, en la que dicho polipéptido es capaz de catalizar la conversión de acetaldehído hasta etanol.
2. Una célula de levadura modificada genéticamente según la reivindicación 1, que sobreexpresa un polipéptido que comprende una secuencia de aminoácidos con al menos un 90% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°: 6.
- 10 3. Una célula de levadura modificada genéticamente que sobreexpresa un primer polipéptido que comprende una secuencia de aminoácidos con al menos un 70% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°: 6 y que comprende una delección o alteración de un gen que codifica un segundo polipéptido que comprende una secuencia de aminoácidos con al menos un 70% de identidad de secuencia respecto de una secuencia de aminoácidos expuesta en SEQ ID N°: 2 o 4, en la que dicho primer polipéptido es capaz de catalizar la conversión de acetaldehído hasta etanol, y en la que dicho segundo polipéptido es capaz de catalizar la conversión de etanol hasta acetaldehído.
- 15 4. Una célula de levadura modificada genéticamente de la reivindicación 3, que sobreexpresa un primer polipéptido que comprende una secuencia de aminoácidos con al menos un 90% de identidad de secuencia respecto de la secuencia de aminoácidos expuesta en SEQ ID N°: 6 y que comprende una delección o alteración de un gen que codifica un segundo polipéptido que comprende una secuencia de aminoácidos con al menos un 90% de identidad de secuencia respecto de una secuencia de aminoácidos expuesta en SEQ ID N°: 2 o 4.
- 20 5. La célula de levadura modificada genéticamente de la reivindicación 4, en la que dicho segundo polipéptido comprende una secuencia de aminoácidos expuesta en SEQ ID N°: 2 o 4.
- 25 6. Un método para producir una célula de levadura modificada genéticamente que comprende delecionar o alterar un gen endógeno que codifica un polipéptido que comprende una secuencia de aminoácidos con al menos un 70% de identidad de secuencia respecto de una secuencia de aminoácidos expuesta en SEQ ID N°: 2 o 4, en la que dicho polipéptido es capaz de catalizar la conversión de etanol hasta acetaldehído.
- 30 7. Un método para producir una célula de levadura modificada genéticamente de la reivindicación 6, que comprende delecionar o alterar un gen endógeno que codifica un polipéptido que comprende una secuencia de aminoácidos con al menos un 90% de identidad de secuencia respecto de una secuencia de aminoácidos expuesta en SEQ ID N°: 2 o 4.
- 35 8. La célula de levadura modificada genéticamente de cualquiera de las reivindicaciones 1, 2, 3, 4, y 5, o la célula de levadura modificada genéticamente producida mediante el método de la reivindicación 6 o 7, en la que dicha célula de levadura pertenece al clado *Issatchenkia orientalis*/Pichia fermentans.
9. La célula de levadura modificada genéticamente de la reivindicación 8, en la que dicha célula de levadura es *I. orientalis*.
10. Un proceso de fermentación en el que una célula de levadura modificada genéticamente como se mencionó en cualquiera de las reivindicaciones 1-5 o 8-9, o la célula de levadura modificada genéticamente producible mediante el método de la reivindicación 6 o 7, se cultiva en un medio de fermentación que comprende xilosa.
- 40 11. El proceso de fermentación de la reivindicación 10, en el que dicho medio de fermentación comprende al menos 10 g/L de xilosa de un hidrolizado de biomasa vegetal.
12. El proceso de fermentación de la reivindicación 11, en el que la xilosa es el azúcar más abundante en dicho medio de fermentación.
- 45 13. Un método para producir etanol a partir de un medio que contiene xilosa, que comprende cultivar una célula de levadura modificada genéticamente como se enumeró en cualquiera de las reivindicaciones 1-5 o 8-9, o la célula de levadura modificada genéticamente producible mediante el método de la reivindicación 6 o 7, en un medio que contiene xilosa.
14. El método de la reivindicación 13, en el que dicho medio que contiene xilosa comprende al menos 10 g/L de xilosa de un hidrolizado de biomasa vegetal.
- 50 15. El método de la reivindicación 14, en el que la xilosa es el azúcar más abundante en dicho medio.

Figura 1

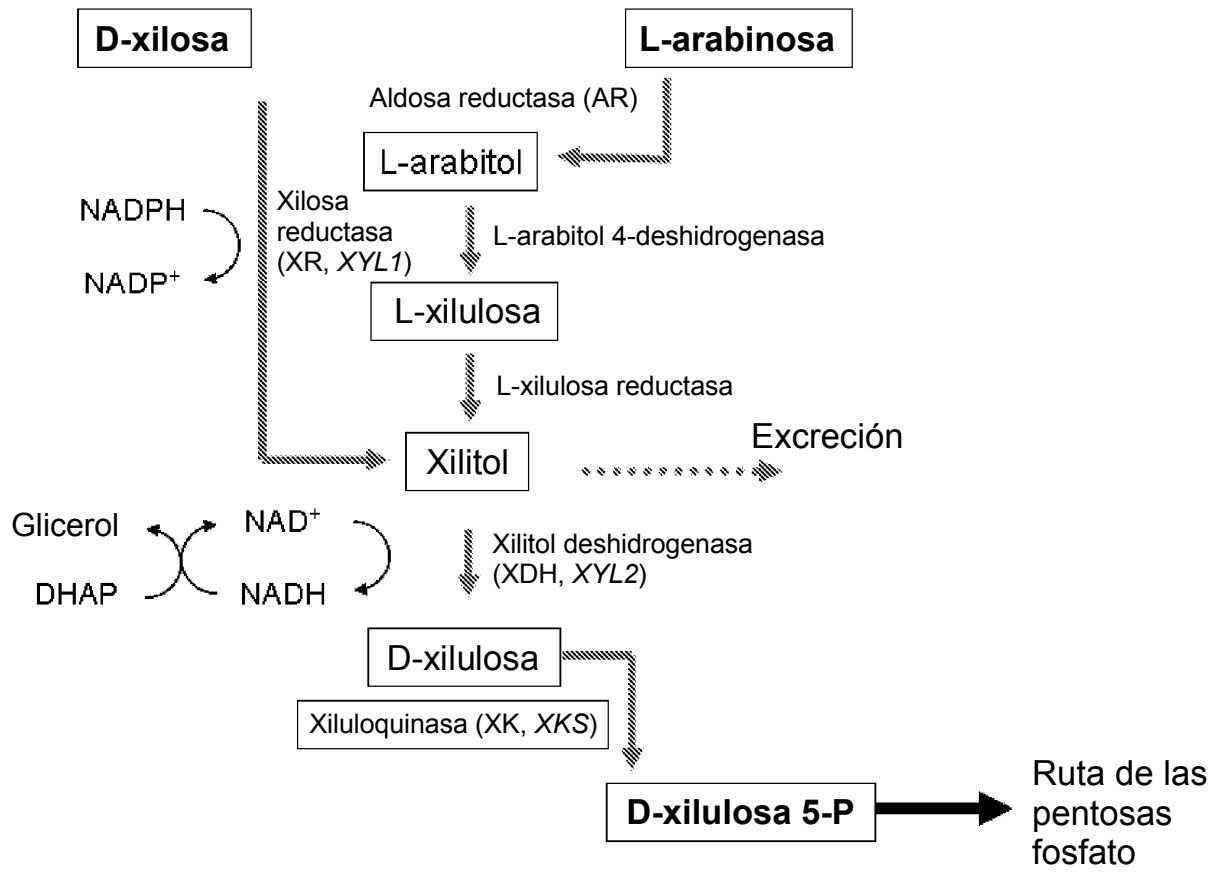


Figura 2

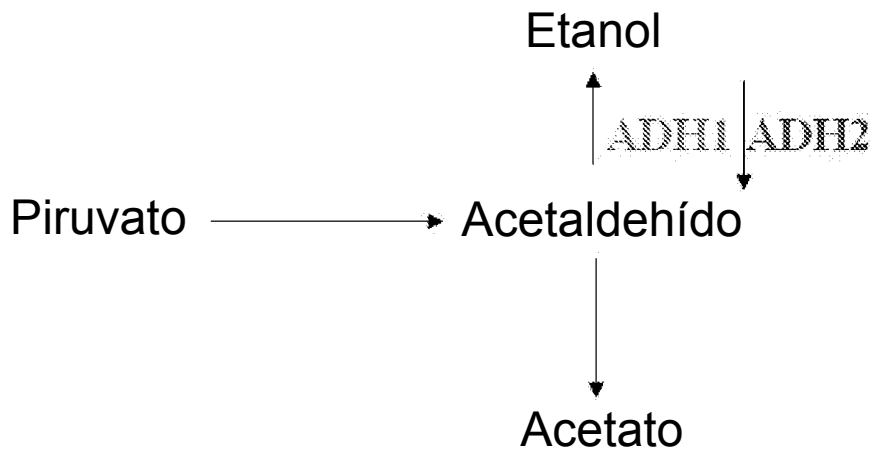


Figura 3

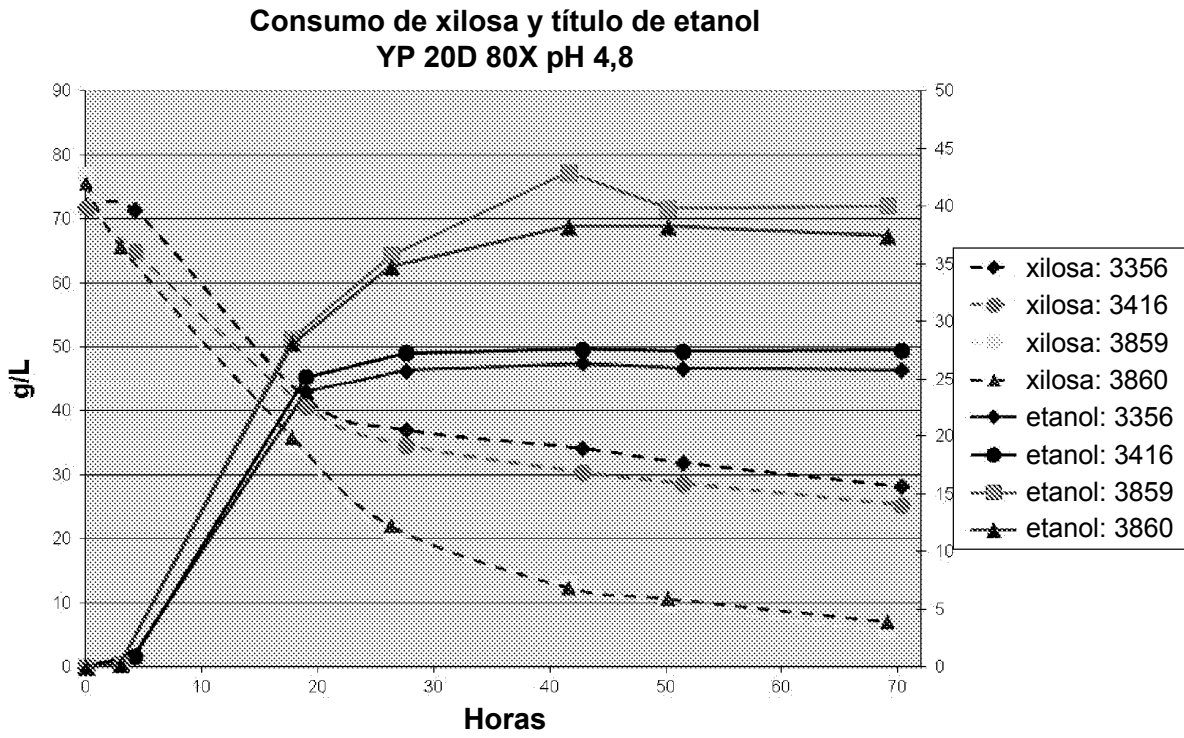


Figura 4

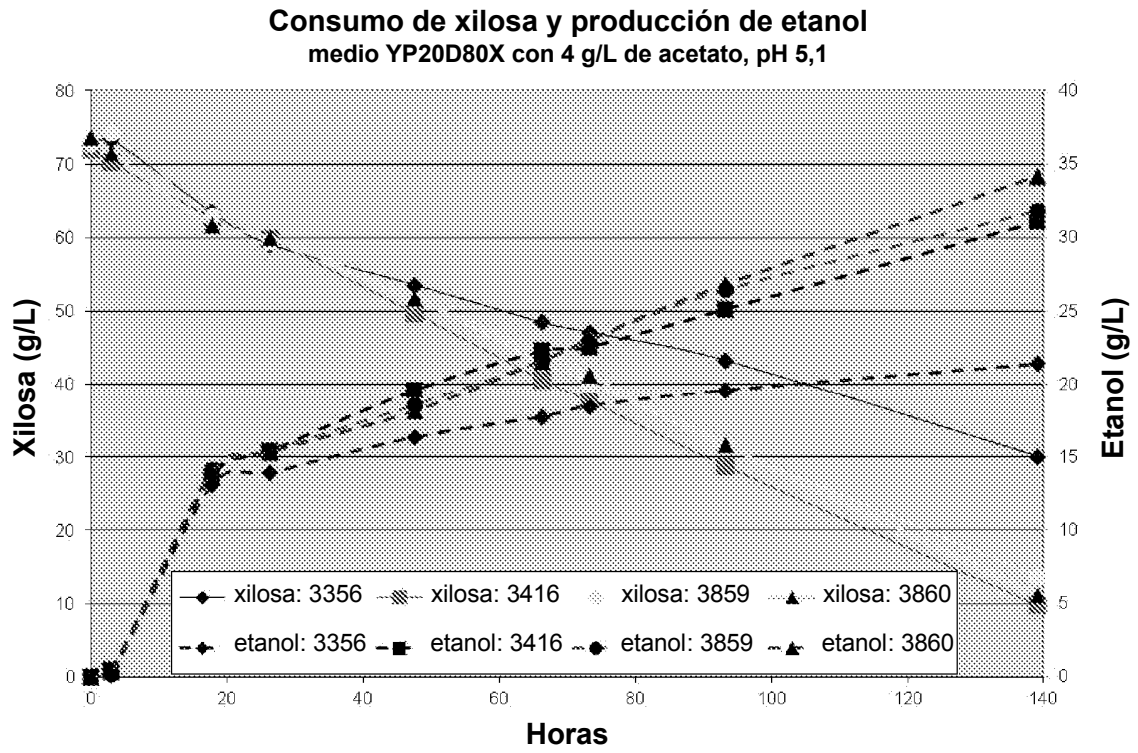




Figura 5

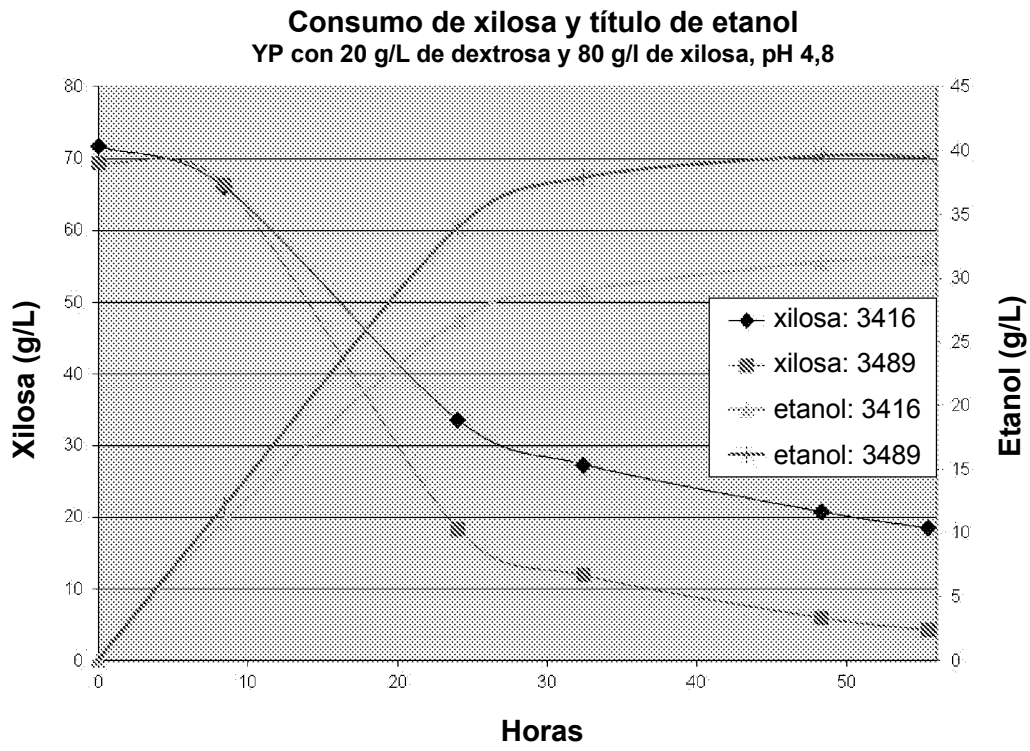


Figura 6

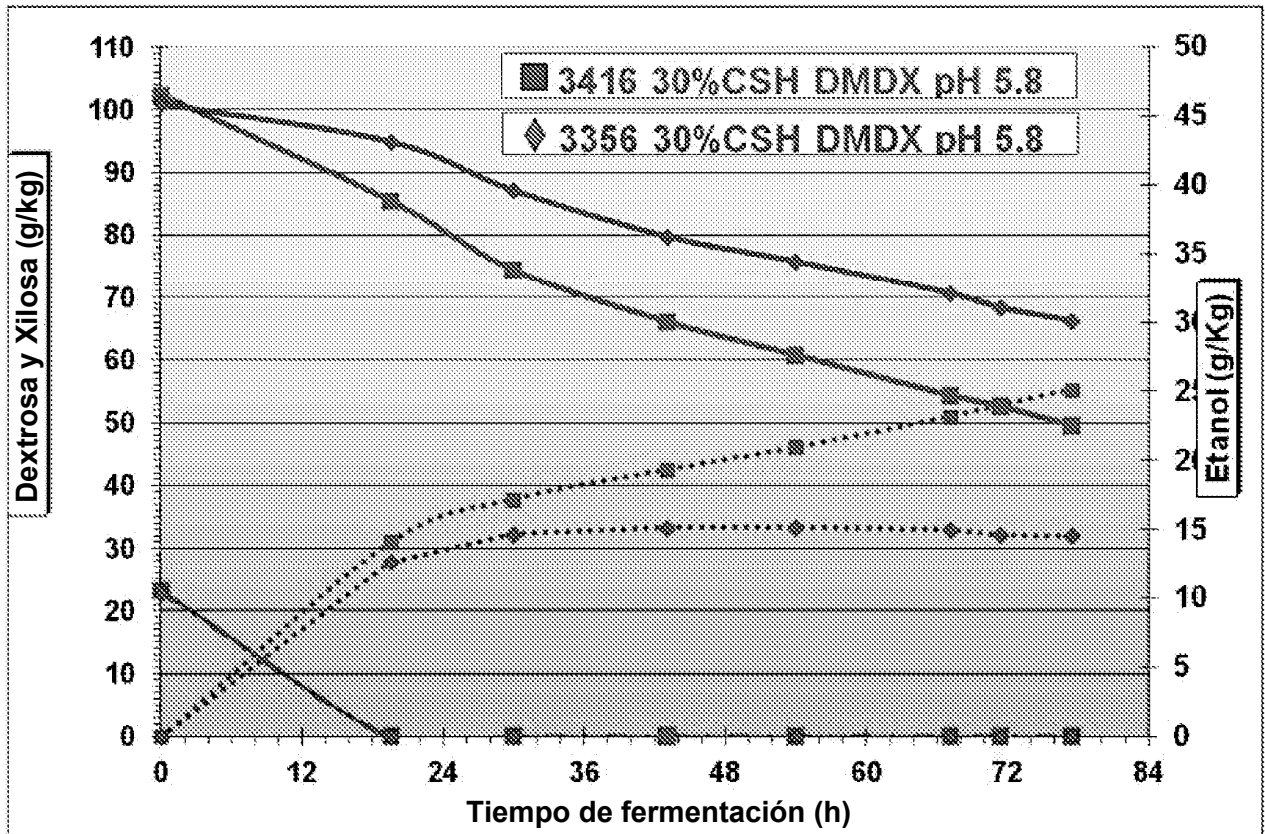


Figura 7

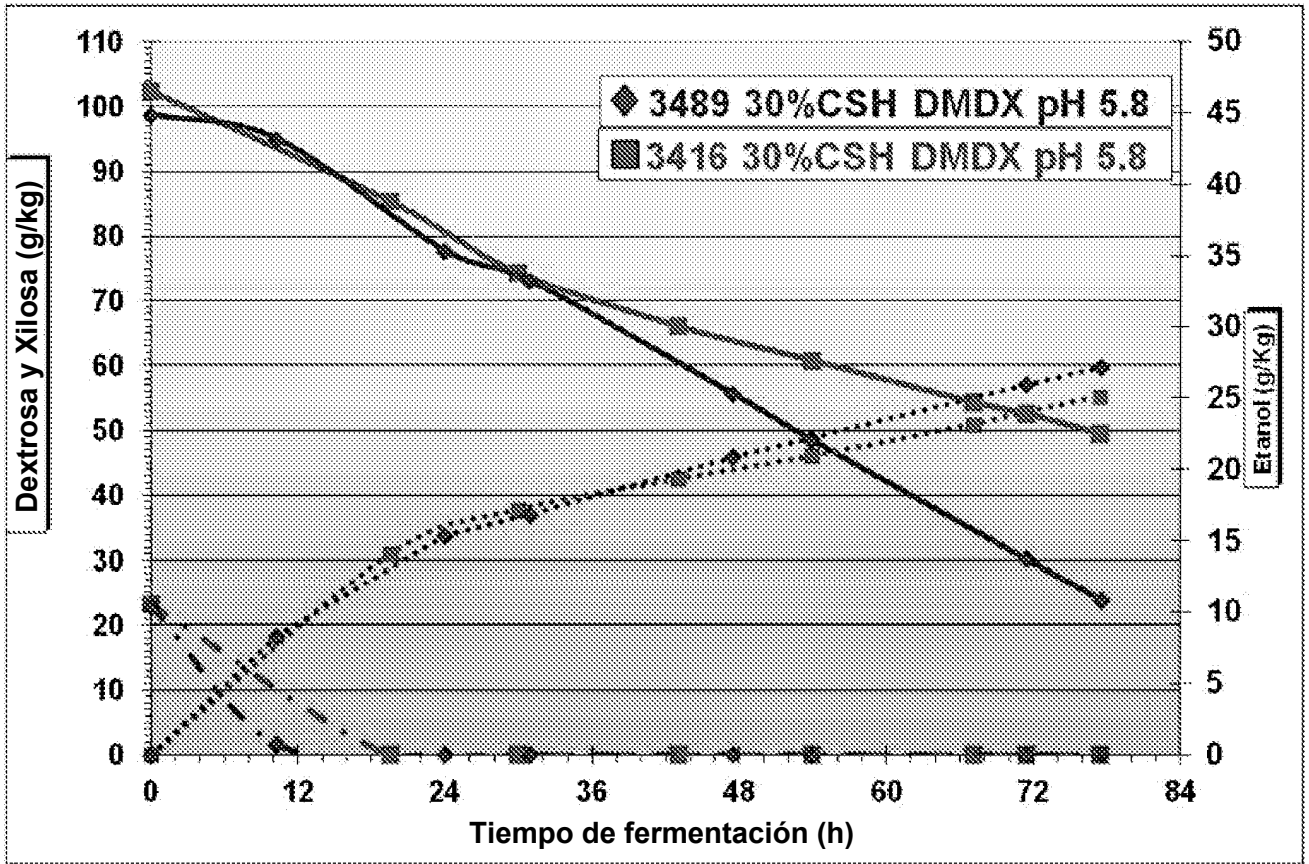


Figura 8a

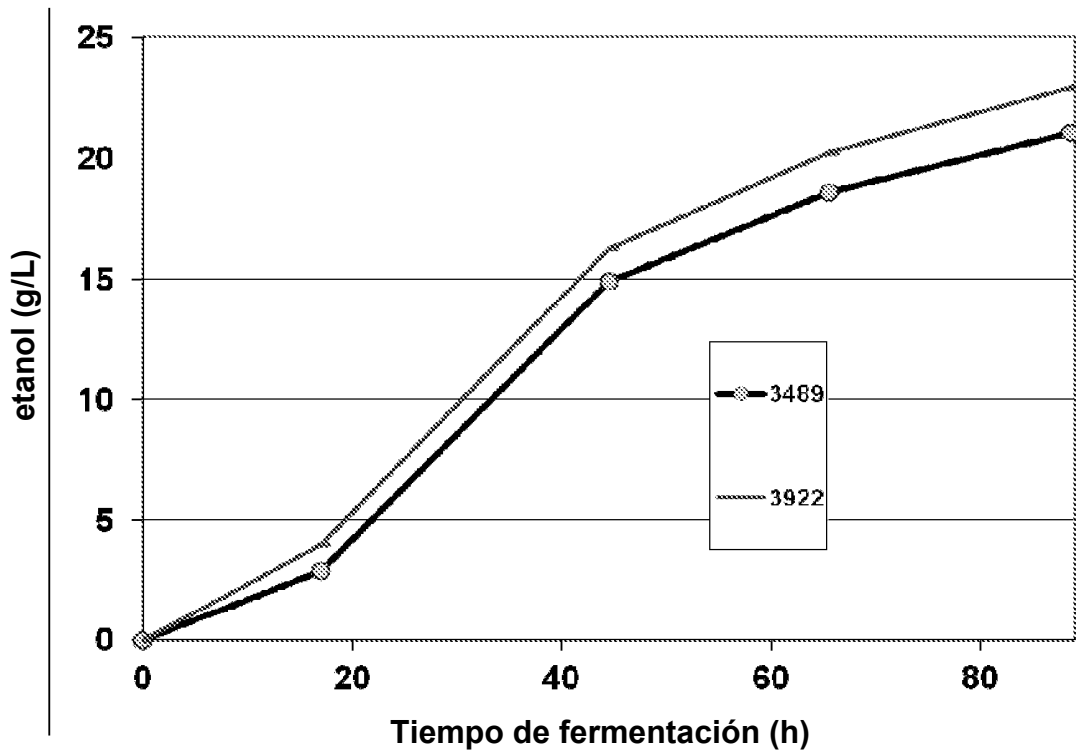


Figura 8b

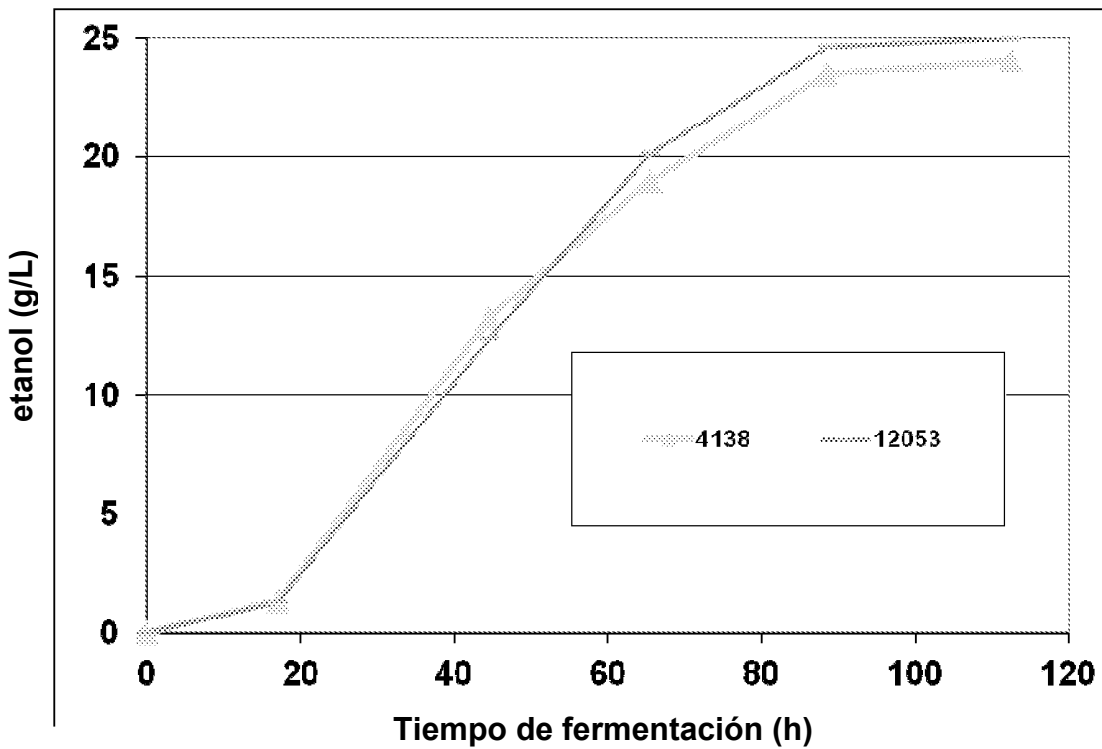


Figura 9a

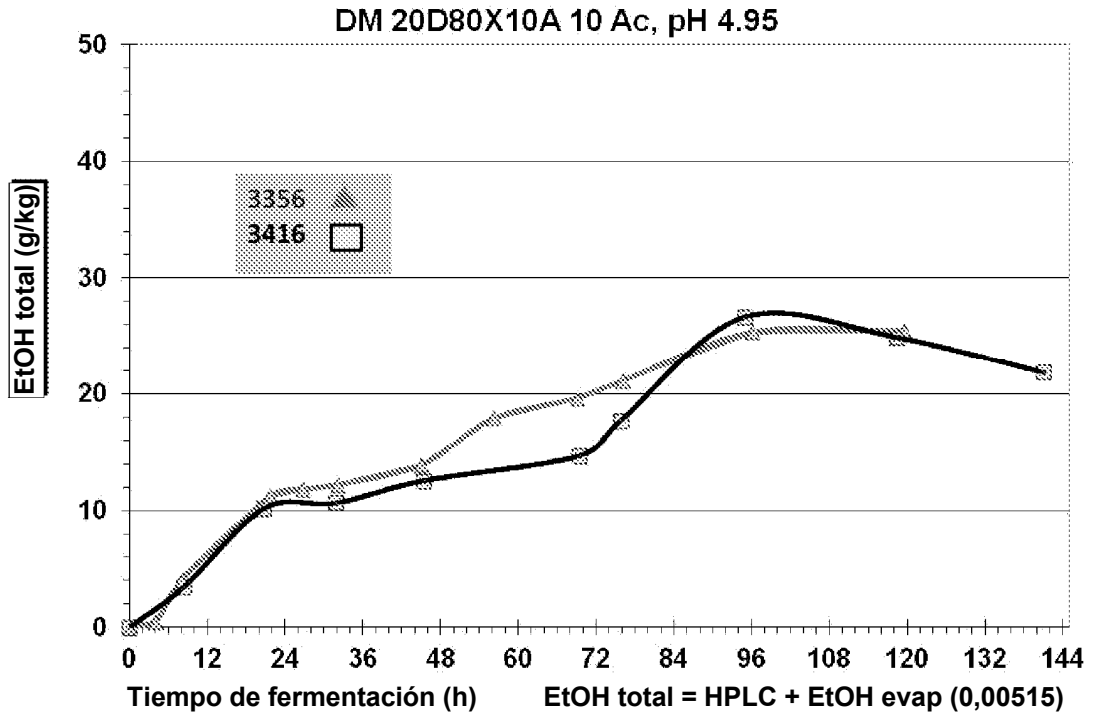


Figura 9b

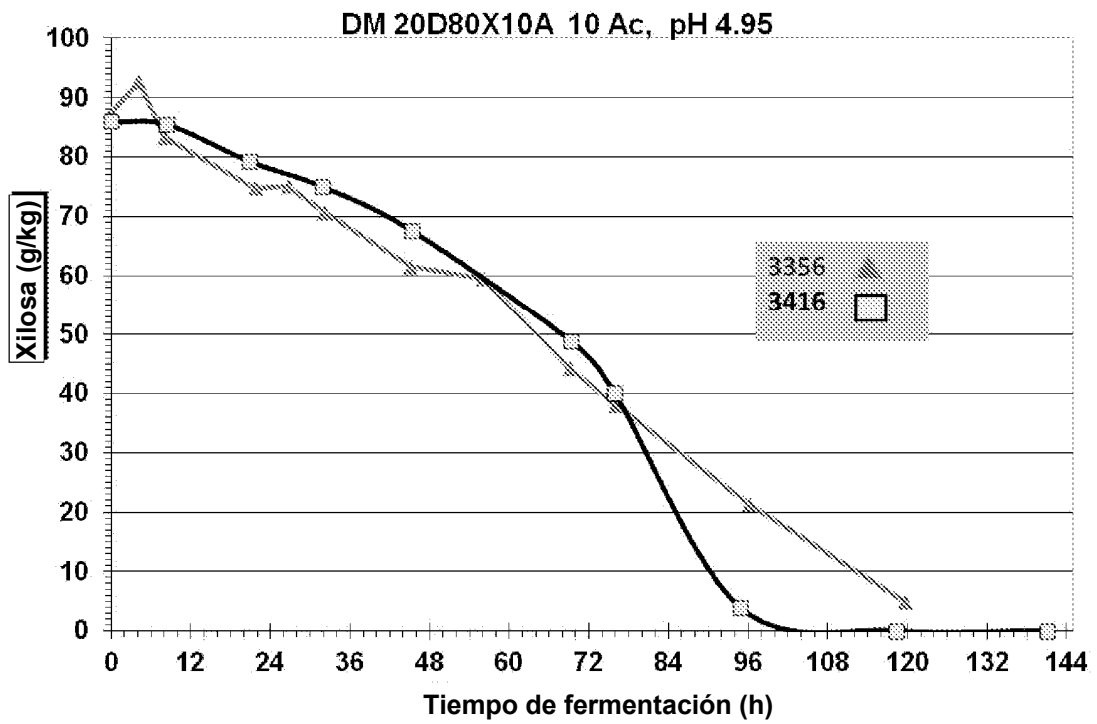


Figura 10a

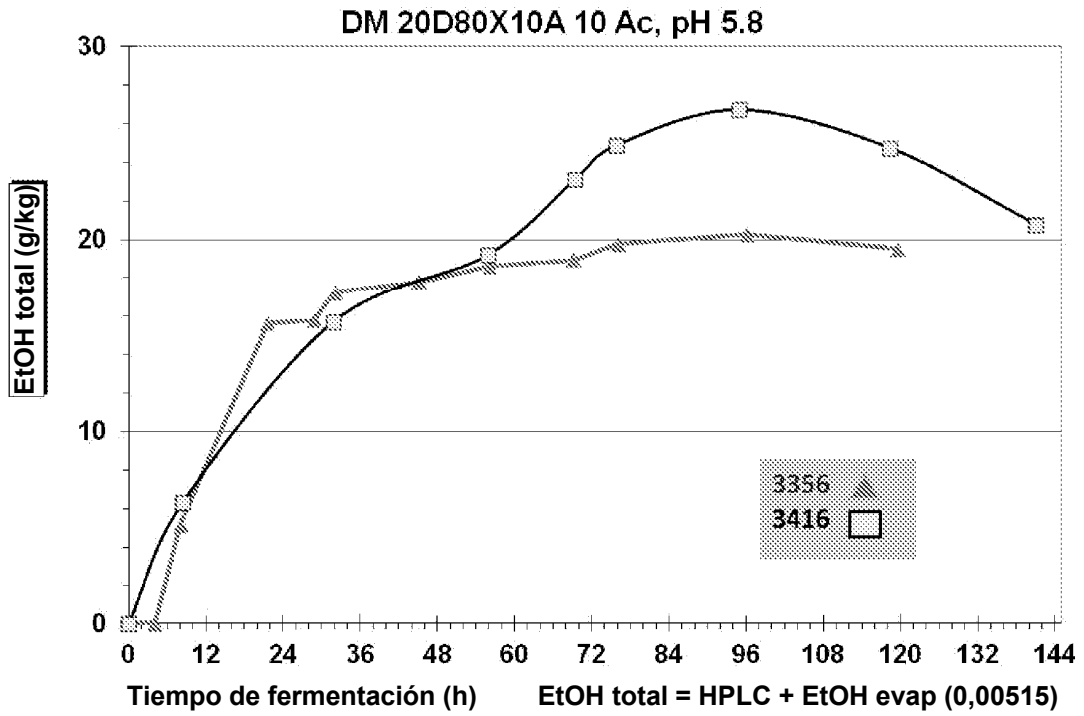


Figura 10b

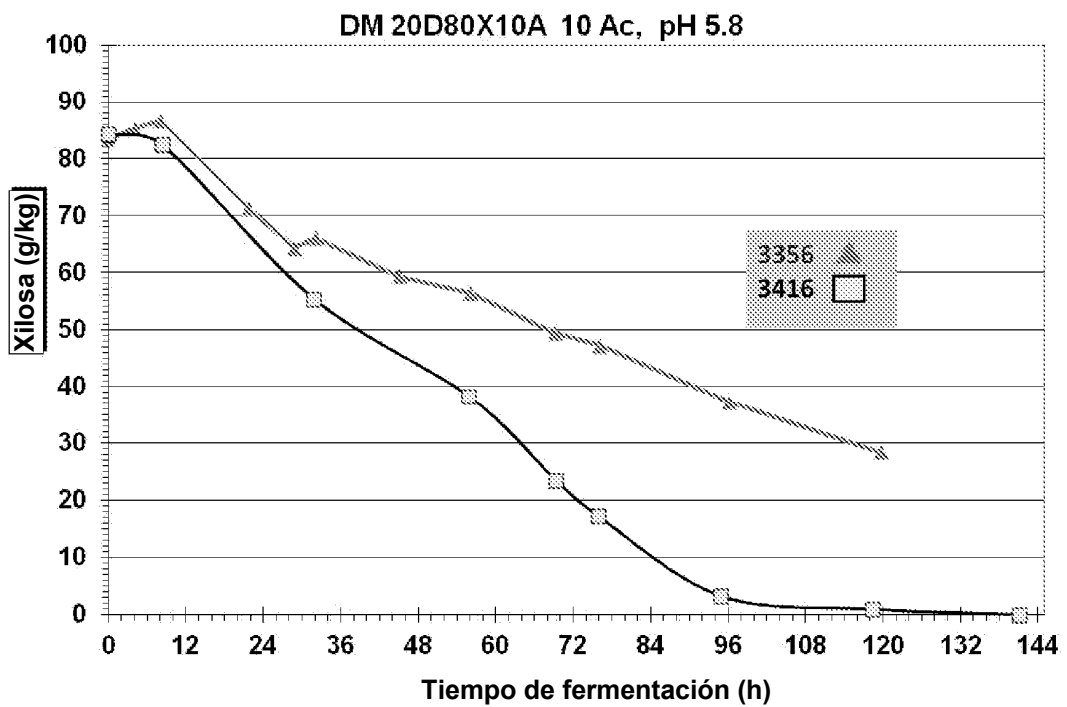


Figura 11a

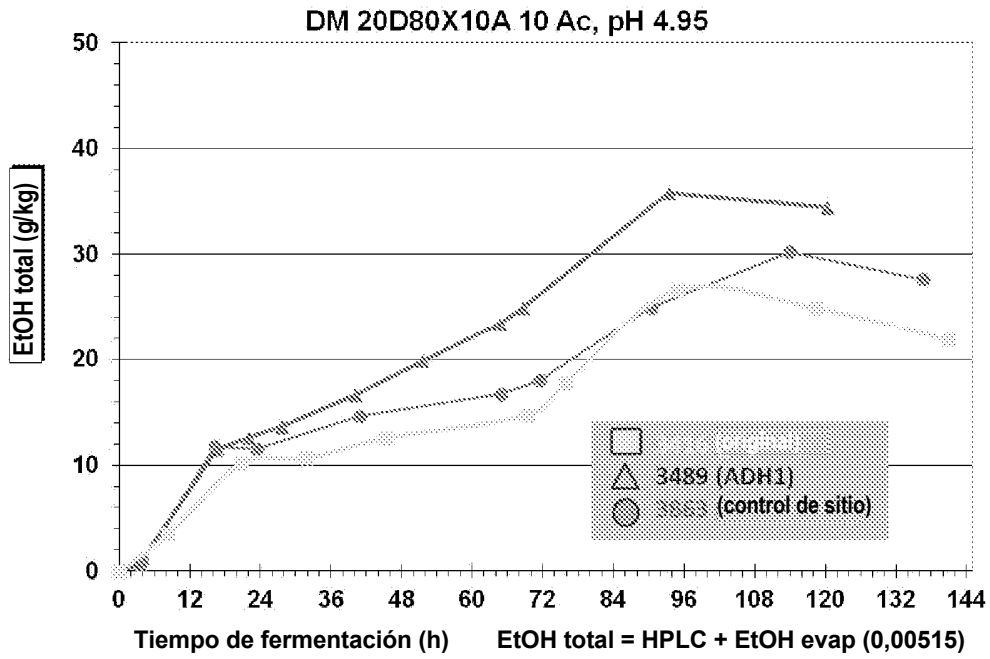


Figura 11b

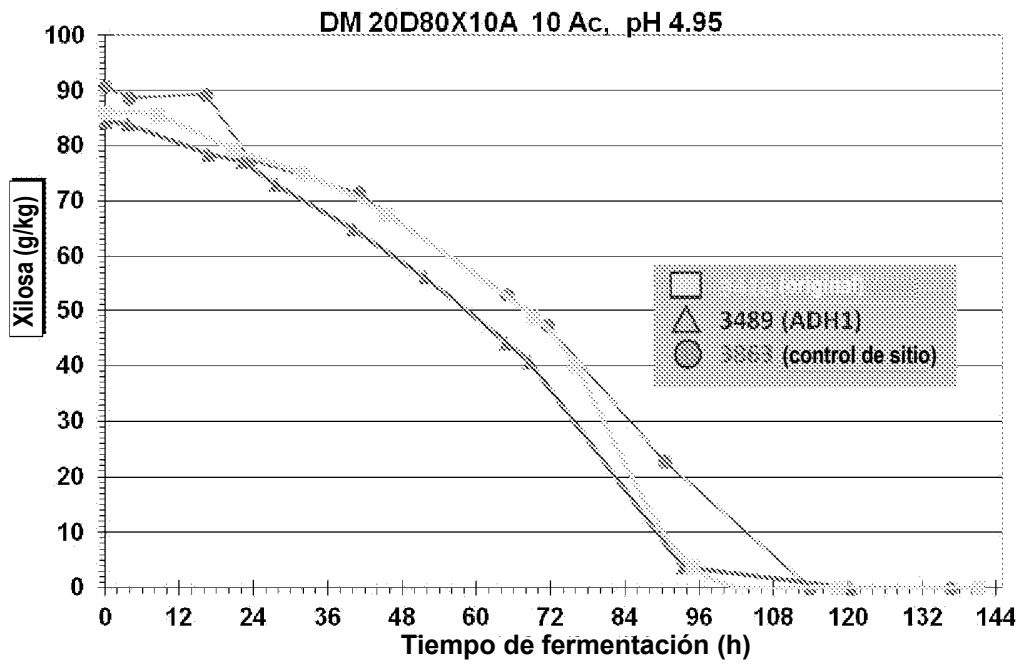


Figura 12a

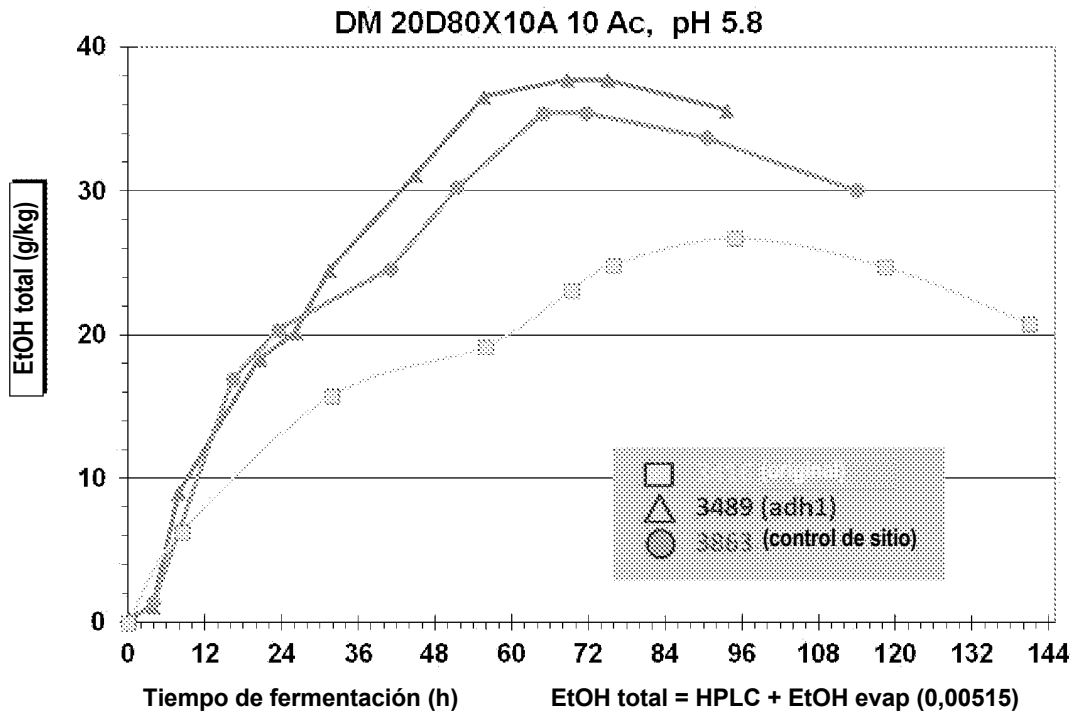


Figura 12b

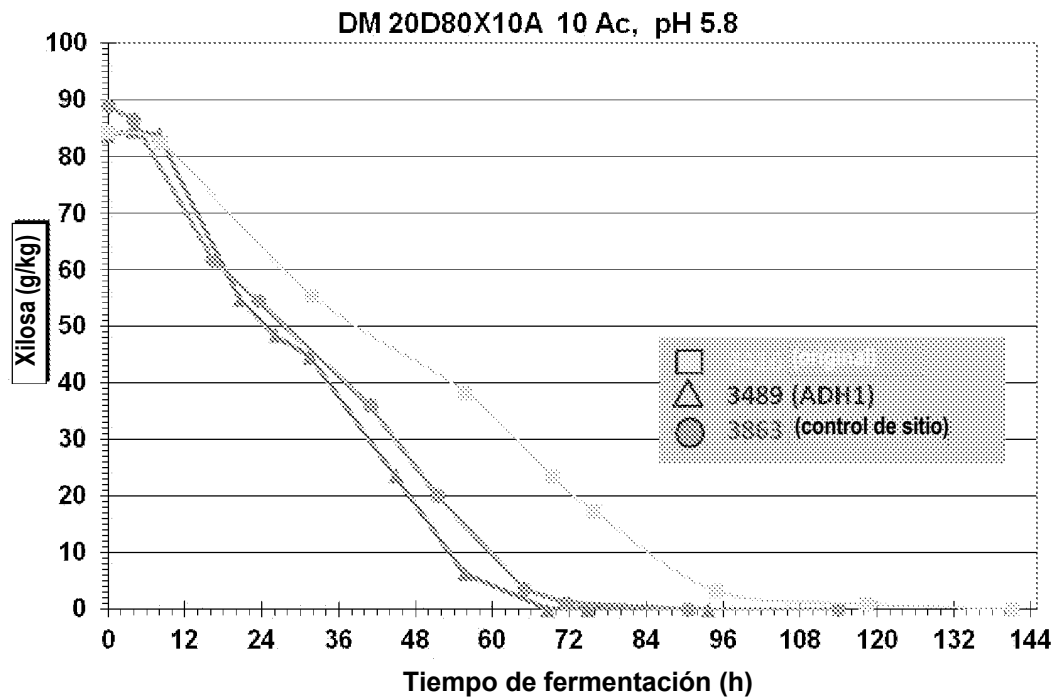




Figura 13a

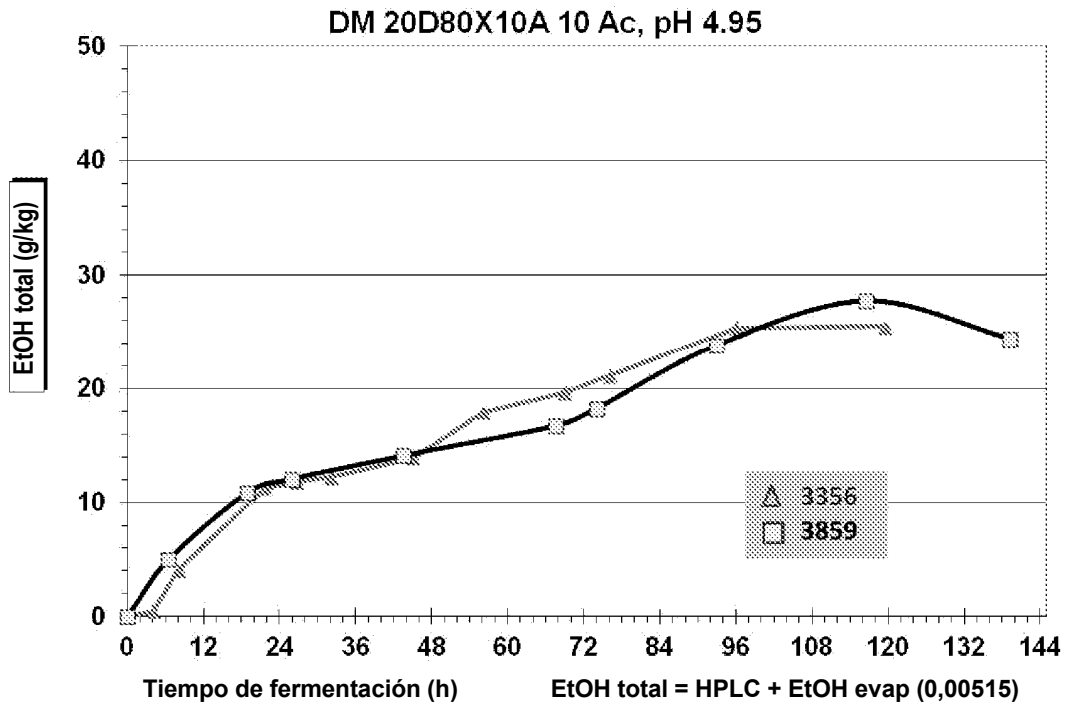


Figura 13b

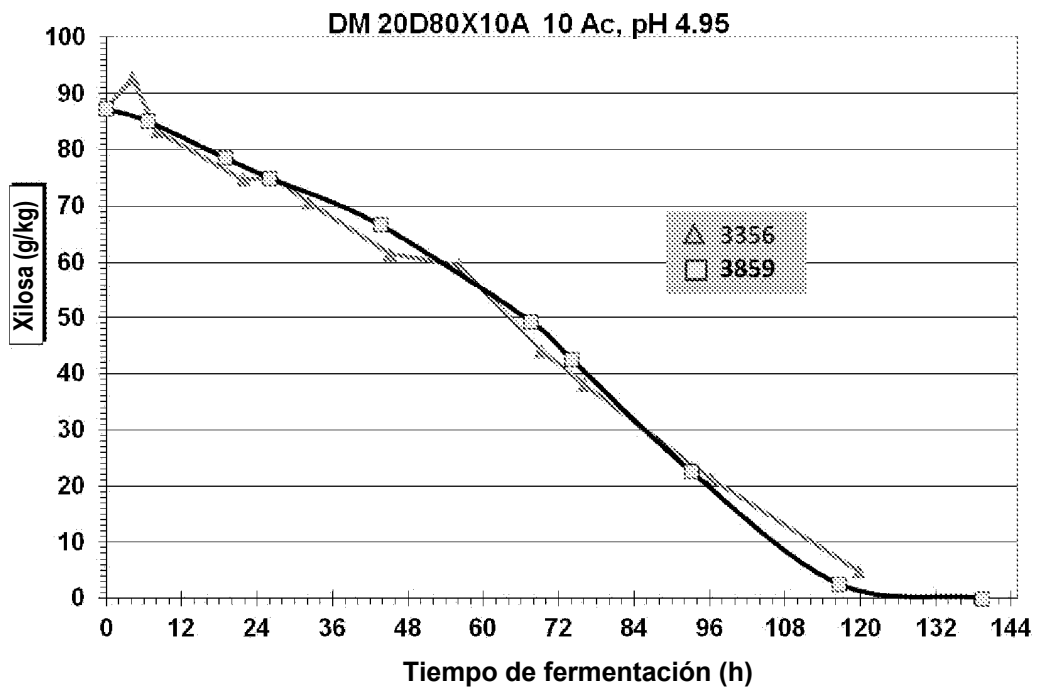


Figura 14a

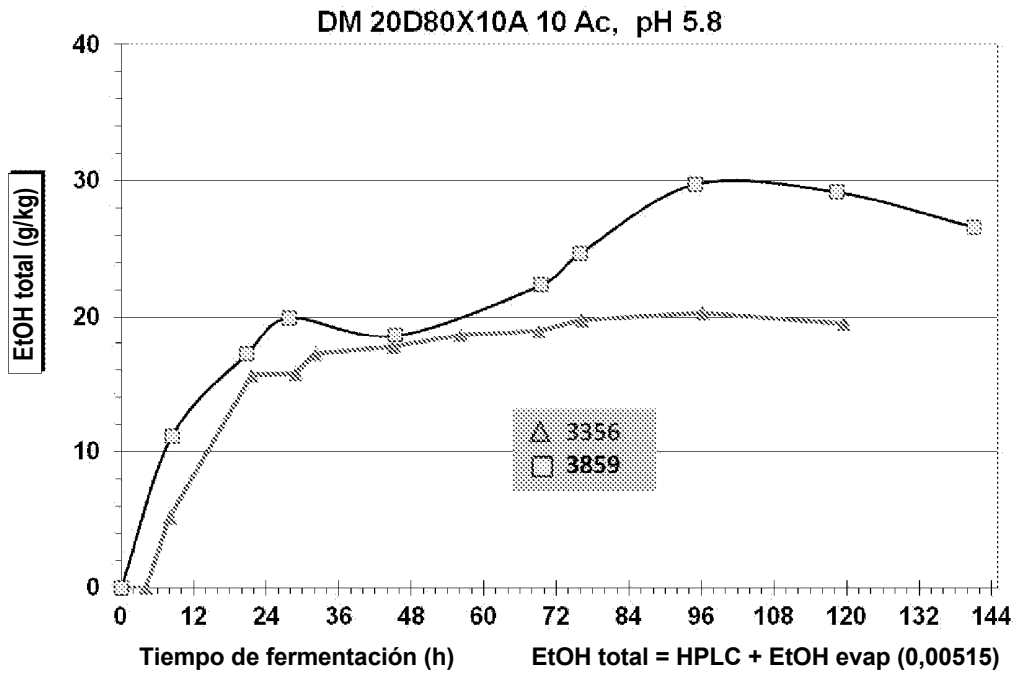
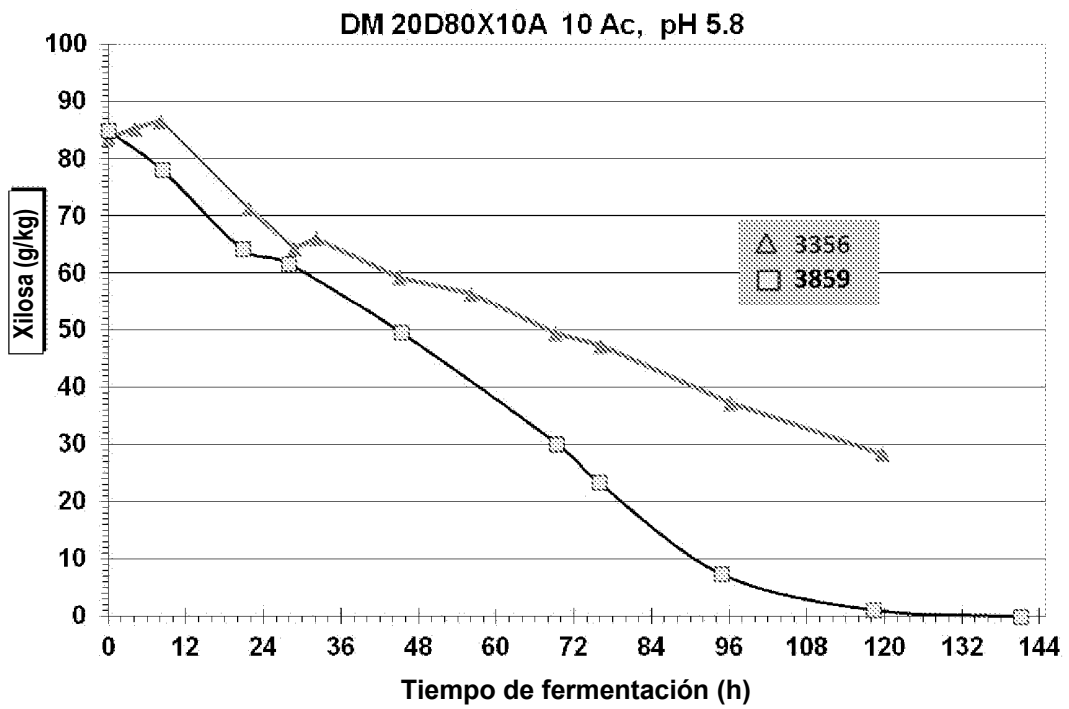


Figura 14b



**Figura 15**

Sc = *S. cerevisiae*

```

ScADH1 (SEQ ID NO:13)      1  -----ms--ipetqkgvifyeshgkleykdipvpkp
ScADH2 (SEQ ID NO:14)      1  -----ms--ipetqkaiifyesngklehkdiipvpkp
ScADH3 (SEQ ID NO:15)      1  mlrtstlfrvrvqpslfsrnilrlqst--aa--ipktqkgvifyenkglhykdipvpep
S141G2556 (SEQ ID NO:6)    1  -----msyeipqtqkacvfyenggpitykdipvpkp
S141G9091 (SEQ ID NO:2)    1  mfastfrsqavraarftrfqst-----fa--ipekqmgvifethggplykeipvpkp
S141G1202 (SEQ ID NO:4)    1  mlsktitaalrgnttrtafrinairslaipa--ipetqkgvifyenggelvykdipvpkp

ScADH1 (SEQ ID NO:13)      30  kanellinvkysgvchtdlhawhgdwplpvklplvgghegagvvvmgenvkgwkigdya
ScADH2 (SEQ ID NO:14)      30  kpnellinvkysgvchtdlhawhgdwplptklplvgghegagvvvmgenvkgwkigdya
ScADH3 (SEQ ID NO:15)      57  kpneilinvkysgvchtdlhawhgdwplpvklplvgghegagvvvklgsnvkgwkvgdla
S141G2556 (SEQ ID NO:6)    32  kpteilvkvlysgvchtdlhawkgdwplatklplvgghegagvvvakgenvtsfeigdya
S141G9091 (SEQ ID NO:2)    52  kpteilinvkysgvchtdlhawkgdwplpaklplvgghegagivvakgsavtnfeigdya
S141G1202 (SEQ ID NO:4)    59  kpneilvkvlysgvchtdlhawkgdwplatklplvgghegagvvvakgdnvtnfeigdya

ScADH1 (SEQ ID NO:13)      90  gikwlngscmaceycelgnesncphadlsgythdgsfqyatadavqaahipqgtdlaqv
ScADH2 (SEQ ID NO:14)      90  gikwlngscmaceycelgnesncphadlsgythdgsfqyatadavqaahipqgtdlaev
ScADH3 (SEQ ID NO:15)     117  gikwlngsamtcefcesghesncpdadlsgythdgsfqyatadaiqaakiqqgtdlaev
S141G2556 (SEQ ID NO:6)    92  gikwlngscmgcefceqgaepncpkadlsgythdgsfqyatadaiqaahisketdlagv
S141G9091 (SEQ ID NO:2)   112  gikwlngscmscefceqgdesncehadlsgythdgsfqyatadaiqaakipkgtdlsev
S141G1202 (SEQ ID NO:4)   119  gikwlngscmgcefceqgaepncpqadlsgythdgsfqyatadavqaakipqgtdlaqv

ScADH1 (SEQ ID NO:13)     150  apilcagitvykalksanlmaghwaaisgaagglgslavqyakamgyrvlgidggegkee
ScADH2 (SEQ ID NO:14)     150  apilcagitvykalksanlraghwaaisgaagglgslavqyakamgyrvlgidggpgkee
ScADH3 (SEQ ID NO:15)     177  apilcagvtvykalkeadlkagdwvwaaisgaagglgslavqyatamgyrvlgidageekek
S141G2556 (SEQ ID NO:6)   152  apilcagvtvykalktadlragewvcisgaagglgslaiqyakamglrvvgidggdekke
S141G9091 (SEQ ID NO:2)   172  apilcagvtvykalktadlragewvwaaisgaagglgslavqyakamglrvlgidggegkke
S141G1202 (SEQ ID NO:4)   179  apilcagitvykalktaelrpgqvwvwaaisgaagglgslavqyakamglrvlgidggegkke

ScADH1 (SEQ ID NO:13)     210  lfrsiggevfidftk-----ekdivgavlkatdg-gahgvinvsvseaaieastryvran
ScADH2 (SEQ ID NO:14)     210  lftslggevfidftk-----ekdivsavvkatng-gahglinvsvseaaieastrycran
ScADH3 (SEQ ID NO:15)     237  lfkklggevfidftk-----tknmvsdiqeatkg-gphgvinvsvseaaaislsteyvrpc
S141G2556 (SEQ ID NO:6)   212  lckslgaeafidftk-----tkdivkavqeatng-gphgvinvsvseaaaisqsceyvrat
S141G9091 (SEQ ID NO:2)   232  lfeqcggdvfidftryprdapekmvadikaatnglgphgvinvsvspaaaisqscedyvrat
S141G1202 (SEQ ID NO:4)   239  fakslgaevfidftk-----skdivkdiqeatng-gphgvinvsvspaaaisqstqyvrtl

ScADH1 (SEQ ID NO:13)     264  gttvvlvgmpagakccsdvfnqvksisivgsyvgnrdradrealdfarglvkspikvvg1
ScADH2 (SEQ ID NO:14)     264  gtvvvlvglpagakccsdvfnhvksisivgsyvgnrdradrealdfarglvkspikvvg1
ScADH3 (SEQ ID NO:15)     291  gtvvvlvglpanayvksevfvshvksinikgsyvgnrdradrealdfsrqlikspikivgl
S141G2556 (SEQ ID NO:6)   266  gkvvvlvglpagaqvktgvfeavvksieikgsyvgnrkdtaealdfytrglvkspfkivgl
S141G9091 (SEQ ID NO:2)   292  gkvvvlvgmpsgavcksdvfvthvkslqikgsyvgnrdradrealdefnegkvrspikvvp1
S141G1202 (SEQ ID NO:4)   293  gkvvvlvglpahavcessvfdhvksiqirgsyvgnrdradrealdeftrglvkspfkivgl

ScADH1 (SEQ ID NO:13)     324  stlpeiyekekmgqivgryvvdtsk
ScADH2 (SEQ ID NO:14)     324  sslpeiyekekmgqiaqryvvdtsk
ScADH3 (SEQ ID NO:15)     351  selpkvydlmekgkilgryvvdtsk
S141G2556 (SEQ ID NO:6)   326  selpkvfelmeqgkilgrmvltdtsk
S141G9091 (SEQ ID NO:2)   352  stlpeiylmekgkilgryvvdtsk
S141G1202 (SEQ ID NO:4)   353  selpkiyelmeqgkilgryvvdtsk

```