



República Federativa do Brasil  
Ministério do Desenvolvimento, Indústria  
e do Comércio Exterior  
Instituto Nacional da Propriedade Industrial

**(21) PI 0807237-0 A2**



(22) Data de Depósito: 01/02/2008  
(43) Data da Publicação: 29/04/2014  
(RPI 2260)

**(51) Int.Cl.:**  
C12N 15/09  
C12N 1/19  
C12N 1/21  
C12N 9/02  
C12P 7/06

**(54) Título:** DNA QUE CODIFICA XILITOL  
DESIDROGENASE, CONSTRUÇÃO DE ÁCIDO  
NUCLÉICO, VETORES, MICROORGANISMOS E  
MÉTODOS PARA PRODUZIR XILITOL  
DESIDROGENASE E ETANOL

**(57) Resumo:**

**(30) Prioridade Unionista:** 02/02/2007 JP 2007-024666

**(73) Titular(es):** Kirin Holdings Kabushiki Kaisha

**(72) Inventor(es):** Hideyuki Tamakawa, Osamu Kobayashi

**(74) Procurador(es):** Dannemann, Siemsen, Bigler &  
Ipanema Moreira

**(86) Pedido Internacional:** PCT JP2008051676 de 01/02/2008

**(87) Publicação Internacional:** WO 2008/093847 de  
07/08/2008

Relatório Descritivo da Patente de Invenção para **"DNA QUE CODIFICA XILITOL DESIDROGENASE, CONSTRUÇÃO DE ÁCIDO NUCLEÍCO, VETORES, MICROORGANISMOS E MÉTODOS PARA PRODUZIR XILITOL DESIDROGENASE E ETANOL".**

5 CAMPO TÉCNICO

A presente invenção refere-se ao DNA que codifica um polipeptídeo que tem uma atividade de xilitol desidrogenase, uma construção de ácido nucleico e um vetor que compreende o DNA, um micro-organismo transformado com o vetor para expressar o DNA, e um método para produzir xilitol desidrogenase. O micro-organismo, etc., da presente invenção podem ser usados em etapas industrialmente úteis tais como etapas de produção de etanol, etapas de produção de biomassa, e etapas de reciclagem de  $\text{NADP}^+$  a partir de NADPH.

ANTECEDENTES DA INVENÇÃO

A xilose é um carboidrato mais existente em biomassa vegetal e madeira, e constitui aproximadamente 40% de substâncias lignocelulósicas. A xilose, em etapas da produção de celulose, é formada como produtos residuais a partir de hidrolisados de xilano, um componente principal de hemicelulose. Deseja-se para uso eficaz de fontes de carbono que a xilose deve ser convertida em etanol ou biomassa para fermentação. Particularmente, o etanol está sendo usado em grandes quantidades como combustível líquido.

*Candida* (Gong, C.S., Chen, L.F., Flickinger, M.C. e Tsao, G.T., "Conversion of hemicelluloses carbohydrate", *Adv. Biochem. Eng.* 20:93-118 (1981); e Jeffries, T.W., "Utilization of xilose by bacteria, yeast and fungi", *Adv. Biochem. Biotech.* 27:1-32 (1983)), *Debaryomyces*, *Hansenula*, *Kluyveromyces*, *Metschnikowia*, *Pachysolen*, *Paecilomyces* (Wu, J.F., Lastick, S.M., Updegraff, D.M., "Ethanol production from sugars derived from plant biomass by a novel fungus", *Nature* 321:887-888 (1986)), *Pichia* (Maleszka, R. e Schneider, H., "Fermentations of D-xylose, xylitol and D-xylulose by yeasts", *Can. J. Microbiol.* 28:360-363 (1982)), e similares são conhecidas como leveduras que podem utilizar pentose (por exemplo, xilose ou D-ribose) ou similares.

Em geral, a conversão de pentose (por exemplo, xilose) para etanol em organismos é mediada pela fosforilação de pentose, e a pentose

fosforilada resultante é convertida em etanol através de uma via de pentose fosfato. A fosforilação de pentose requer primeiramente a redução de pentose acompanhada por conversão de NADPH em NADP<sup>+</sup>, e esta reação é catalisada por redutase. Pentitol, que é formado pela redução de pentose, é subsequentemente submetido à oxidação acompanhada de conversão de NAD<sup>+</sup> em NADH. Esta reação é catalisada por desidrogenase. A D-pentulose é formada através de reação em duas etapas e fosforilada por cinase para formar pentose fosfato (Barnett, J.A., "The utilization of sugars by yeasts. In: "Advances in carbohydrate chemistry and biochemistry"; editado por Tipson, R.S. e Horton, D.; New York: Academic Press, 1976, páginas 125-235).

*S. cerevisiae*, uma levedura usada predominantemente na produção de bioetanol, pode utilizar xilulose (pentulose) convertida a partir de xilose, embora esta levedura não possa fermentar pentoses (Jeffries, T.W., "Emerging technology for fermenting D-xylose", *Trends in Biotechnology* 3:208-212 (1985)). O que é importante é que *S. cerevisiae* contém genes que codificam proteínas que fermentam pentose, mas não expressa estes genes.

Acredita-se que a fermentação de pentose por *S. cerevisiae* seja possivelmente realizada disponibilizando uma via de utilização de xilose a partir de um micro-organismo que metaboliza xilose. Muitas tentativas foram feitas em *S. cerevisiae* para expressar genes de xilose isomerase bacteriana, e a fermentação de xilose, entretanto, terminou de forma mal sucedida, presumivelmente devido à expressão insuficiente dos genes estranhos (Sartny, A.V., McConaugh, B.L., Lodo, Z., Sundstrom, J.A., Furlong, C.E. e Hall, B.D., "Expression of *Escherichia coli* xylose isomerase gene in *S. cerevisiae*", *Appl. Eur. Microbiol.* 53:1996-2000 (1987); Amoer, R., Wilhelm, M. e Hollenberg, C.P., "The fermentation of xylose – an analysis of the expression of *Bacillus* and *Actinoplanes* isomerase genes in yeast", *Appl. Microbiol. Biotechnol.* 30:351-357 (1989); Chan, E.-C., Ueng, P.P. e Chen, L.F., "D-xylose fermentation to ethanol by *Schizosaccharomyces pombe* cloned with xilose isomerase gene", *Biotech. Lett.* 8:231-234 (1986); Chan, E.-C., Ueng, P.P. e Chen, L.F., "Environmental effects on D-xylose fermentation by *Schizosac-*

*charomyces cerevisiae*", *Appl. Biotechnol.* 20:221-231 (1989)).

Por outro lado, foram feitas tentativas como métodos que usam genes de levedura em outra levedura para isolar genes que codificam redutase e xilitol desidrogenase a partir de *Pichia stipitis* que tem a capacidade de utilizar xilose, e para expressar os genes em *S. cerevisiae* (patentes nºs 5 JP 3122153 e 3193917). Além disso, foram feitas tentativas para isolar o gene de xilose redutase (Govinden, R., Pillay, B.van Zyl, W.H. e Pillay, D., "Candida shehatae xylose reductase gene, complete Eds.", GenBank direct submission, Acesso AF278715 (2000)) e purificação de xilitol desidrogenase 10 (Yang, V.W. e Jeffries, T.W., "Purification and properties of xylitol dehydrogenase from the xylose-fermenting yeast *Candida shehatae*", *Appl. Biochem. Biotechnol.* 26:97-206 (1990)) a partir de *Candida shehatae*, outra levedura que tem a capacidade de utilizar xilose. Entretanto, nenhum dos casos foi sucedido em isolar genes de xilitol desidrogenase a partir de *Candida shehatae*. 15

#### DESCRIÇÃO DA INVENÇÃO

Um objeto da presente invenção é fornecer DNA que codifica xilitol desidrogenase inusitada, que pode ser usado na tecnologia da conversão eficiente de xilose em etanol ou biomassa, e um método que usa o mesmo. 20

#### SUMÁRIO DA INVENÇÃO

A presente invenção tem as características abaixo.

(1) Um DNA que codifica um polipeptídeo selecionado entre:

(a) um polipeptídeo que compreende a sequência de aminoácidos de SEQ ID NO: 1, e tem uma atividade de xilitol desidrogenase; 25

(b) um polipeptídeo que compreende uma sequência de aminoácidos que tem 90% ou mais de identidade com a sequência de aminoácidos de SEQ ID NO: 1 e tem uma atividade de xilitol desidrogenase; e

(c) um polipeptídeo que compreende uma sequência de aminoácidos que compreende substituição, deleção, inserção ou adição de um ou 30 vários aminoácidos na sequência de aminoácidos de SEQ ID NO: 1 e tem uma atividade de xilitol desidrogenase.

(2) O DNA de acordo com (1), onde o DNA compreende a sequência de nucleotídeos de SEQ ID NO:2.

(3) O DNA de acordo com (1), onde o DNA compreende uma sequência que difere da sequência de nucleotídeos de SEQ ID NO: 2 devido à degeneração do código genético.

(4) O DNA que codifica qualquer um de (1) a (3), onde o DNA é de uma levedura.

(5) O DNA de acordo com (4), onde a levedura é *Candida shehatae*.

(6) O DNA de acordo com (5), onde a levedura é *Candida shehatae* CBS5813 (NBRC1983).

(7) Uma construção de ácido nucléico que compreende um DNA de acordo com qualquer um entre (1) a (6) e uma sequência reguladora capaz de regular a expressão do DNA em uma célula hospedeira.

(8) A construção de ácido nucléico de acordo com (7), onde o DNA codifica a sequência de aminoácidos de SEQ ID NO: 1.

(9) A construção de ácido nucléico de acordo com (8), onde o DNA compreende a sequência de aminoácidos de SEQ ID NO: 2.

(10) A construção de ácido nucléico de acordo com qualquer um de (7) a (9), onde a sequência reguladora é da célula hospedeira.

(11) A construção de ácido nucléico de acordo com qualquer um de (7) a (9), onde a sequência reguladora é um promotor.

(12) A construção de ácido nucléico de acordo com (11), onde o promotor é um promotor constitutivo ou um promotor induzível.

(13) A construção de ácido nucléico de acordo com (11) ou (12), onde o promotor é selecionado no grupo que consiste nos promotores ADH1, ADH2, PDC, GAL1/10, TDH3, e PGK1.

(14) Um vetor que compreende um DNA de acordo com qualquer um de (1) a (6) ou uma construção de ácido nucléico de acordo com qualquer um de (7) a (13).

(15) O vetor de acordo com (14), onde o vetor é um plasmídeo.

(16) Um micro-organismo que compreende um DNA de acordo

com qualquer um de (1) a (6), uma construção de ácido nucléico de acordo com qualquer um de (7) a (13), ou um vetor de acordo com (14) ou (15), de tal modo que o micro-organismo expresse um polipeptídeo que tem uma atividade de xilitol desidrogenase.

5 (17) O micro-organismo de acordo com (16), onde o micro-organismo é uma levedura ou bactéria.

(18) O micro-organismo de acordo com (17), onde a levedura ou bactéria é selecionada no grupo que consiste em leveduras do gênero *Saccharomyces*, *Schizosaccharomyces*, *Schwanniomyces*, *Kluyveromyces*, *Pichia*, *Hansenula*, *Candida*, *Debaryomyces*, *Metschnikowia*, *Pachulosen*, ou *Paecilomyces* e bactérias do gênero *Zymomonas*.

(19) O micro-organismo de acordo com (18), onde o micro-organismo é *Saccharomyces cerevisiae*.

15 (20) O micro-organismo de acordo com (18), onde o micro-organismo é *Saccharomyces pombe*.

(21) O micro-organismo de acordo com qualquer um de (16) a (20), onde o DNA ou a construção de ácido nucléico é incorporada no genoma do micro-organismo hospedeiro.

20 (22) Um método para produzir xilitol desidrogenase, compreendendo as etapas de:

(a) cultivar um micro-organismo de acordo com qualquer um de (16) a (21) em um meio; e

(b) coletar um produto que tem uma atividade de xilitol desidrogenase a partir do micro-organismo ou do meio.

25 (23) O método de acordo com (22), compreendendo ainda a etapa de selecionar um micro-organismo apropriado para a fermentação de xilulose.

(24) Um método para produzir etanol usando um micro-organismo de acordo com qualquer um de (16) a (21).

30 (25) Um vetor que compreende um cassete de expressão do gene de xilitol desidrogenase, que compreende um DNA de acordo com (1), um cassete de expressão do gene de xilose redutase, e um cassete de ex-

pressão do gene de xilulocinase.

A presente invenção fornece o DNA que codifica a enzima xilitol desidrogenase referente à fermentação de xilulose, um micro-organismo (por exemplo, uma levedura) que expressa de forma recombinante o DNA, etc. O DNA e o micro-organismo são úteis na produção de bioetanol, produção de biomassa, indústria alimentícia, e similares.

#### BREVE DESCRIÇÃO DO DESENHO

A figura 1 ilustra uma delineação de procedimentos para construir um vetor de expressão que compreende um gene de xilitol desidrogenase a partir de *Candida shehatae*.

#### MELHOR MODO PARA CONDUZIR A INVENÇÃO

A presente invenção fornece o DNA que codifica um polipeptídeo que tem uma atividade de xilitol desidrogenase.

A xilitol desidrogenase (EC1.1.1.9) é uma enzima que catalisa uma parte de um processo de fermentação de xilulose, usando xilose como uma matéria-prima. A primeira metade do processo é uma reação através da qual xilitol é formado de forma redutora a partir de xilose. Esta reação redutora é catalisada por uma enzima xilose redutase. A última metade do processo é uma reação através da qual a xilulose é formada de forma oxidante a partir de xilitol. A reação oxidante é catalisada por uma enzima xilitol desidrogenase. A atividade de xilitol desidrogenase é dependente de NAD(P)<sup>+</sup>. Foram feitas tentativas para introduzir os genes de xilose redutase e também xilitol desidrogenase em células de levedura (por exemplo, *Saccharomyces cerevisiae*), melhorando desta forma a capacidade de produzir etanol (vide patente US nº 6.582.944).

O DNA da presente invenção é um DNA que codifica um polipeptídeo selecionado entre:

(a) um polipeptídeo que compreende a sequência de aminoácidos de SEQ ID NO: 1 e tem uma atividade de xilitol desidrogenase;

(b) um polipeptídeo que compreende uma sequência de aminoácidos que tem 90% ou mais identidade com a sequência de aminoácidos de SEQ ID NO: 1, e tem uma atividade de xilitol desidrogenase;

(c) um polipeptídeo que compreende uma sequência de aminoácidos que compreende substituição, deleção, inserção, ou adição de um ou vários aminoácidos na sequência de aminoácidos de SEQ ID NO: 1, e tem uma atividade de xilitol desidrogenase.

5 O DNA que codifica o polipeptídeo de SEQ ID NO: 1 é de uma levedura e especificamente de *Candida shehatae*, e particularmente *Candida shehatae* CBS5813 (NBRC1983).

Na presente invenção, o polipeptídeo pode consistir em apenas a sequência de aminoácidos de SEQ ID NO: 1 ou pode compreender a dita  
10 sequência de aminoácidos bem como uma sequência adicional, por exemplo, no terminal amino ou carboxi do polipeptídeo. Tal sequência de aminoácidos adicional é, por exemplo, uma sequência sinalizadora de peptídeo que proporciona secreção extracelular de proteínas maduras.

Alternativamente, o polipeptídeo engloba também uma variante,  
15 homólogo, análogo, ou similar, que compreende uma sequência de aminoácidos que tem 90% ou mais de identidade com a sequência de aminoácidos de SEQ ID NO: 1. Neste contexto, o termo "identidade" significa a proporção (%) do número de aminoácidos completamente equivalentes ao número total de aminoácidos na sequência de aminoácidos alinhada com ou sem lacunas  
20 introduzidas nela. Tal polipeptídeo compreende genericamente uma sequência de aminoácidos que compreende substituição, deleção, inserção, ou adição de um ou vários aminoácidos na sequência de aminoácidos de SEQ ID NO: 1, e tem de preferência 92% ou mais, ou 93% ou mais, mais preferivelmente 95% ou mais, ou 97% ou mais, ainda mais preferivelmente 98% ou  
25 mais, ou 99% ou mais de identidade de sequência à dita sequência de aminoácidos.

A proteína que tem uma sequência que tem 90% ou mais de identidade com ela pode ser buscada, por exemplo, acessando uma base de dados conhecida nessas técnicas (GenBank, EMBL, etc.). Por exemplo, um  
30 algoritmo conhecido nessas técnicas, tal como BLAST ou FASTA, pode ser usado como um sistema de busca. Os procedimentos específicos da operação para essa busca estão descritos, por exemplo, "Introduction to Geno-

meNet Databases”, 2ª edição, editado por Toshiba Takagi e Minoru Kanehisa (1998), Kyoritsu Shuppan Co. Ltd. (Tóquio, Japão).

Alternativamente, o polipeptídeo engloba também uma variante que compreende uma sequência de aminoácidos que compreende substituição, deleção, inserção, ou adição de um ou vários aminoácidos na sequência de aminoácidos de SEQ ID NO: 1. Neste contexto, o termo “vários” significa um número inteiro de aproximadamente 10 ou menos, por exemplo, um número inteiro de 2 a 10, 2 a 9, 2 a 8, 2 a 7, 2 a 6, 2 a 5, 2 a 4, ou 2 a 3.

A variante que compreende substituição, deleção, inserção, ou adição de aminoácidos está englobada dentro do âmbito da presente invenção, desde que ela seja inusitada e tenha uma atividade de xilitol desidrogenase.

A variação de aminoácidos no polipeptídeo codificado pelo DNA da presente invenção é, de preferência, uma mutação em um sítio que não afeta adversamente, na sequência de aminoácidos de SEQ ID NO: 1, a estrutura tridimensional (ou conformação) de um domínio catalítico associado à atividade de xilitol desidrogenase ou um domínio ligante de coenzima sobre o qual as coenzimas atuam. Entretanto, uma mutação pode ser introduzida mesmo no domínio catalítico ou no domínio ligante de coenzima, a menos que ela reduza significativamente a atividade da enzima. Os exemplos de aminoácidos que constituem o domínio ligante de coenzima incluem Asp<sup>207</sup>, Ile<sup>208</sup>, Phe<sup>209</sup>, Asn<sup>21</sup>, e Lys<sup>212</sup> na sequência de aminoácidos de SEQ ID NO: 1. Os exemplos de aminoácidos que constituem o domínio catalítico incluem Cys<sup>41</sup>, His<sup>66</sup>, Glu<sup>67</sup>, Ser<sup>96</sup>, Ser<sup>99</sup>, Tyr<sup>102</sup>, e Cys<sup>110</sup> na sequência de aminoácidos de SEQ ID NO:1. A introdução de mutação ou mutações fornece, por exemplo, uma variante que tem atividade mais alta, por exemplo, 1,5 vez ou mais alta, 2 vezes ou mais alta, 2,5 vezes ou mais alta, 3 vezes ou mais alta, 5 vezes ou mais alta, 7 vezes ou mais alta, 10 vezes ou mais altas, 15 vezes ou mais alta, 20 vezes ou mais alta, 30 vezes ou mais alta, 40 vezes ou mais alta, ou 50 vezes ou mais alta, do que aquela do tipo selvagem. Alternativamente, a introdução de mutação ou mutações pode melhorar também a estabilidade térmica da enzima. Os exemplos de variações incluem variações

como aquelas descritas em Watanabe, S. *et al.*, *J. Biol. Chem.* 280(11):10340-10349 (2005).

Particularmente, a substituição entre aminoácidos pode ser uma substituição conservativa entre aminoácidos similares em propriedade estrutural ou química ou pode ser uma substituição não-conservativa entre aminoácidos que diferem nessa propriedade. Os aminoácidos similares em propriedade estrutural ou química podem ser classificados da seguinte maneira:

um grupo de aminoácido hidrofóbico inclui alanina, leucina, isoleucina, valina, metionina, e prolina.

Um grupo aminoácido polar inclui serina, treonina, glicina, glutamina, asparagina, e cisteína.

Um grupo aminoácido aromático inclui fenilalanina, tirosina, e triptofano.

Um grupo aminoácido ácido inclui ácido glutâmico e ácido aspártico.

Um grupo aminoácido básico inclui lisina, arginina, e histidina.

O DNA da presente invenção compreende um DNA de filamento único ou filamento duplo, que tem uma sequência de nucleotídeos que codifica qualquer um dos polipeptídeos mencionados acima. O DNA da presente invenção engloba, por exemplo, DNA natural, cDNA, DNA quimicamente sintetizado, DNA obtido por reação de polimerase em cadeia (PCR), e DNA que compreende dois ou mais destes DNA's em combinação.

Especificamente, o DNA da presente invenção é um DNA que compreende a sequência de nucleotídeos de SEQ ID NO: 2. Além disso, outro DNA da presente invenção é um DNA que compreende uma sequência variante da sequência de nucleotídeos de SEQ ID NO: 2 pela degeneração do código genético. Por exemplo, um códon ótimo para uma espécie de organismo diferente daquela da origem do tipo selvagem pode ser selecionado baseado na degeneração do código genético e incluído no DNA. A sequência variante significa uma sequência que compreende variações pelas quais a sequência de nucleotídeos resultante difere da sequência de nucleotídeos originais, embora elas codifiquem as sequências de aminoácidos idênticas.

O DNA da presente invenção pode ser preparado usando técnicas conhecidas nessa área científica, tais como técnicas de recombinação de DNA, PCR, hibridização, clonagem, e mutagênese direcionada a sítios. Tais técnicas estão descritas detalhadamente, por exemplo, em Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1989, e Ausubel *et al.*, "Current Protocols in Molecular Biology", John Wiley & Sons, 1998, e podem ser usadas para a preparação do DNA da presente invenção.

Primeiramente, *Candida shehatae* é desenvolvida, por exemplo, vascolejando a cultura a aproximadamente 25°C, em um meio que contém peptona, glicose, extratos de levedura, e similares, tais como meio YPD ou YM, e em seguida, extração do DNA genômico. Iniciadores diretos e reversos com 19 a 25 bases de comprimento são desenhados e sintetizados, baseado, por exemplo, na sequência de nucleotídeos de SEQ ID NO: 2. Estes iniciadores são usados para realizar ampliação por PCR na presença de polimerase termoestável e  $Mg^{2+}$ , usando o DNA genômico de interesse como um modelo. As condições da PCR envolvem a realização, por exemplo, do tratamento a 94°C por 1 minuto, e subsequentemente, cerca de 35 ciclos de ampliação, cada um envolvendo 94°C por 30 segundos (desnaturação), 55°C por 30 segundos (anelação), e 72°C por 3 minutos (reação de alongamento), e em seguida, tratamento final a 72°C por 7 minutos. A ampliação exitosa do fragmento do DNA de interesse pode ser confirmada por eletroforese em gel de agarose.

O fragmento de DNA de interesse ampliado é inserido, depois da introdução de sítios de restrição nele, caso necessário, em um vetor (por exemplo, um plasmídeo), que é por sua vez introduzido em uma célula hospedeira apropriada para clonar o fragmento de DNA. O vetor pode conter adequadamente um promotor, uma origem de replicação, um sítio ligante de ribossomas ou sequência de Shine-Dalgarno, um terminador, um sítio de multiclonagem, um sinal de poliadenilação, e similares. Os vetores para procariontes tais como bactérias, ou vetores para eucariontes tais como leveduras ou fungos, podem ser usados como vetor. Como vários vetores estão co-

mercialmente disponíveis, um apropriado pode ser selecionado para uso. Além disso, os exemplos de célula hospedeira incluem bactérias (por exemplo, *E. coli*) e leveduras.

A técnica de mutagênese direcionada à sítio envolve sintetizar um iniciador que contém um sítio de variação e realizar a PCR usando este iniciador e, como modelo, um vetor que contém o DNA de interesse. A variante do DNA que compreende a sequência de nucleotídeos de SEQ ID NO: 2 pode ser preparada usando esta técnica.

A presente invenção fornece ainda uma construção de ácido nucleico que compreende o DNA descrito acima e uma sequência reguladora capaz de regular a expressão do DNA.

Na presente invenção, a sequência reguladora é uma sequência de nucleotídeos capaz de regular a expressão do DNA da presente invenção e inclui, por exemplo, um promotor, um intensificador, um sinal de poliadenilação, uma origem de replicação, um sítio ligante de ribossomas ou sequência de Shine-Dalgarno, e um terminador. A sequência reguladora é, de preferência, uma sequência que contém um promotor ou uma sequência que contém um terminador.

De acordo com uma modalidade da presente invenção, a sequência reguladora é da célula hospedeira. Particularmente, quando a sequência genômica que contém a sequência reguladora de 5' e 3' de um gene que codifica xilitol desidrogenase originalmente portada por uma levedura como hospedeira é deixada flanquear o DNA heterólogo da presente invenção para preparar um vetor, a sequência de DNA heteróloga da presente invenção pode ser incorporada no genoma da célula hospedeira por recombinação homóloga.

O promotor pode ser um promotor constitutivo ou um promotor induzível. O promotor não está particularmente limitado, e pode ser selecionado, por exemplo, no grupo que consiste em ADH1, ADH2, PDC, GAL1/10, TDH3, e PGK1. Dependendo dos promotores selecionados (por exemplo, promotores fortes selecionados), o DNA da presente invenção pode ser expressado em um nível de expressão que excede o nível de expressão natu-

ral em seus micro-organismos hospedeiros originais.

A presente invenção fornece ainda um vetor que compreende qualquer um dos DNA's descritos acima ou qualquer uma das construções de ácidos nucléico descritas acima.

5 Os exemplos do vetor incluem plasmídeos, cosmídeos, fagos, fogomídeos, BAC, YAC, e vírus. Em uma modalidade preferível, o vetor é um plasmídeo. O vetor é apropriado para replicação no micro-organismo hospedeiro desejável, e assim sendo, pode conter, por exemplo, um promotor, um intensificador, um sinal de poliadenilação, uma origem de replicação, 10 um sítio ligante de ribossomas ou sequência de Shine-Dalgarno, e um terminador, selecionado adequadamente. Além disso, um marcador seletivo, tal como um gene de resistência a fármaco (por exemplo, um gene de resistência à ampicilina), pode ser introduzido no vetor.

A presente invenção engloba também um vetor que compreende 15 não apenas um cassete de expressão de xilitol desidrogenase, mas também opcionalmente um cassete de expressão do gene de xilulocinase e/ou um cassete de expressão do gene de xilose redutase. O gene de xilitol desidrogenase pode ser, por exemplo, uma sequência de DNA que codifica o polipeptídeo de SEQ ID NO: 1 ou uma sequência de DNA que codifica uma variante, homólogo, ou análogo do polipeptídeo como descrito acima. O gene 20 de xilulocinase pode ser, por exemplo, uma sequência de DNA que codifica o polipeptídeo de SEQ ID NO: 26 ou uma sequência de DNA que codifica uma variante, homólogo, ou análogo do polipeptídeo como descrito acima. O gene de xilose redutase pode ser, por exemplo, uma sequência de DNA que 25 codifica o polipeptídeo de SEQ ID NO: 24 ou uma sequência de DNA que codifica uma variante, homólogo, ou análogo do polipeptídeo como descrito acima.

Na presente invenção, o "cassete de expressão" é uma construção de ácido nucléico que é preparada ligando operacionalmente a sequência 30 gênica de interesse a um promotor e um terminador e outras sequências reguladoras necessárias para a expressão gênica em uma hospedeira. A construção é introduzida em uma hospedeira, permitindo desta forma que a

hospedeira expresse o gene de interesse. Na presente invenção, o termo “ligado operacionalmente” significa que as sequências reguladoras são ligadas ao gene de interesse de tal modo que elas sejam capazes de atuar sobre a expressão do gene de interesse da maneira desejada. A expressão de cada gene pode ser continuamente ou simultaneamente. Uma construção exemplificativa de tal vetor da presente invenção está ilustrada na figura 1. O vetor pode ser usado eficazmente para a produção de etanol usando xilose ou similares como matéria-prima.

A presente invenção fornece ainda um micro-organismo que compreende qualquer um dos DNA's descritos acima, qualquer uma das construções de ácidos nucléicos descritas acima, ou qualquer um dos vetores descritos acima, de tal modo que o micro-organismo expresse um polipeptídeo que tem uma atividade de xilitol desidrogenase.

Uma abordagem rotineira, tal como transformação ou transdução, é usada para introduzir o vetor na célula hospedeira. Tal abordagem engloba, por exemplo, métodos com fosfato de cálcio, eletroporação, protoplastos, e esferoplastos.

Os exemplos de micro-organismos incluem, porém sem limitações, leveduras, fungos, e bactérias. De preferência, o micro-organismo é selecionado no grupo que consiste em leveduras dos gêneros *Saccharomyces*, *Schizosaccharomyces*, *Schwanniomyces*, *Kluyveromyces*, *Pichia*, *Hansenula*, *Candida*, *Debaryomyces*, *Metschnikowia*, *Pachulosen*, ou *Paecilomyces* e bactérias do gênero *Zymomonas*. De preferência, o micro-organismo é *Saccharomyces cerevisiae*, ou *Schizosaccharomyces pombe*.

Para o micro-organismo, o DNA ou a construção de ácido nucléico pode ser incorporada no genoma do micro-organismo hospedeiro. A incorporação no genoma pode ser realizada ligando as sequências do genoma do micro-organismo hospedeiro a montante e a jusante do DNA ou da construção de ácido nucléico. Um método apropriado para a incorporação da sequência de nucleotídeos no genoma da célula hospedeira é genericamente conhecido pelos versados nessas técnicas. O vetor pode ficar localizado também separadamente do genoma da célula hospedeira na célula. Neste

caso, o vetor deve ser capaz de replicar de forma autônoma.

O micro-organismo da presente invenção pode conter ainda um DNA que codifica outra enzima que medeia o processo de conversão de xilose em etanol. A outra enzima é, por exemplo, xilose redutase, ou xilulocinase, de preferência, xilose redutase. Tais enzimas podem ser de um micro-organismo da mesma origem que aquela da enzima xilitol desidrogenase de acordo com a presente invenção ou pode ser de um micro-organismo de uma origem diferente dela. Além disso, o micro-organismo pode ser submetido a um tratamento de mutação para intensificar a atividade da enzima ou a estabilidade térmica (por exemplo, o micro-organismo é irradiado com raios de alta energia tais como raios UV, raios-gama, ou feixes de íons de elétrons, ou tratado com um mutágeno químico tal como nitrosoguanidina, nitrosouréia, ou ácido nitroso).

A presente invenção fornece ainda um método para produzir xilitol desidrogenase. Este método compreende as etapas de: (a) cultivar qualquer um dos micro-organismos descritos acima em um meio; e (b) coletar, a partir do micro-organismo ou do meio, um produto que tem uma atividade de xilitol desidrogenase. A coleta do produto a partir do micro-organismo pode ser realizada, por exemplo, rompendo com ultrassom a célula do micro-organismo. A coleta do produto a partir do meio requer, depois da tradução no polipeptídeo dentro da célula do micro-organismo, secretar de forma extracelular este polipeptídeo. Para este propósito, por exemplo, o polipeptídeo pode ser expressado com um sinal de secreção do terminal amino ligado a ele.

As condições da cultura do micro-organismo podem ser selecionadas arbitrariamente dentre as condições apropriadas para o tipo do micro-organismo. O meio contém genericamente uma fonte de carbono, uma fonte de nitrogênio, e um sal inorgânico. Os exemplos de fonte de carbono incluem glicose, maltose, sacarose, lactose, amido, e outros açúcares. Os exemplos de fontes de nitrogênio incluem sais inorgânicos que contêm nitrogênio (por exemplo, sulfato e nitrato de amônio), peptona, extratos de malte, e extratos de levedura. Os exemplos de sal inorgânico incluem sais de metais alcalinos, sais de metais alcalino-terrosos, e sais de metais do grupo do fer-

ro. As maneiras de cultura englobam cultura estática, cultura com vascolejamento, cultura com aeração, cultura contínua, e similares. A temperatura da cultura é aproximadamente 15 a 60°C, usualmente aproximadamente 25 a 40°C, embora não limitada a esta faixa.

5                   A atividade da enzima formada pode ser medida, por exemplo, medindo a redução de  $\text{NAD(P)}^+$  a 340 nm a 35°C (Rizzi, M. *et al.*, *J. Ferment. Bioeng.* 67:20-24 (1989)). A mistura do ensaio padrão é uma solução que contém  $\text{MgCl}_2$  50 mM e xilitol 300 mM dissolvidos em Tris-HCl 50 mM (pH 9,0). A reação é iniciada pela adição de 100  $\mu\text{L}$  de de  $\text{NAD(P)}^+$  20 mM  
10 em um volume final de 1,0 mL. A atividade da enzima pode ser indicada, por exemplo, usando, como 1 unidade, uma atividade da enzima que produz 1  $\mu\text{mol}$  de  $\text{NAD(P)H}$  por minuto.

                  A enzima xilitol desidrogenase pode ser purificada usando abordagens genéricas de química de proteínas em combinação. Os exemplos  
15 dessas abordagens incluem, porém sem limitações, filtração em gel, cromatografia de troca iônica, cromatografia de afinidade, cromatografia de interação hidrofóbica, HPLC, eletroforese, focagem isoelétrica, precipitação de sais, fracionamento com sulfato de amônio, e ultrafiltração.

                  De acordo com uma modalidade da presente invenção, o método  
20 do pode compreender ainda a etapa de selecionar um micro-organismo apropriado para fermentação eficiente de xilulose. Como resultado, pode ser obtido um polipeptídeo que catalisa a fermentação de xilulose mais eficientemente. A eficiência da fermentação de xilulose é avaliada, por exemplo, cultivando o micro-organismo em um meio que contém xilose ou xilitol por  
25 um dado período e depois medindo a concentração de xilose contida no meio.

                  A presente invenção fornece ainda um método para produzir etanol usando qualquer um dos micro-organismos descritos acima. O micro-organismo pode ser imobilizado sobre uma fase sólida que permite o crescimento do micro-organismo. Os exemplos de tal fase sólida incluem, porém  
30 sem limitações, polímeros tais como poliuretano e derivados dos mesmos, e polissacarídeos tais como alginato e carragenina.

Um carboidrato mais preferível para a produção de etanol é xilose. Assim sendo, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, e/ou uma cepa de *Zymomonas* capaz de fermentar xilose são muito vantajosas para a produção de etanol. As cepas de leveduras da presente invenção

5 têm a capacidade de fermentar uma solução concentrada de carboidrato, um alto nível de resistência ao etanol, e a capacidade de produzir etanol em uma concentração elevada. Estas cepas de leveduras têm alta viabilidade celular para reciclagem repetitiva e apresentam resistências ao pH e temperatura. Em etapas da produção de xilose, a xilose é formada como produto

10 residual a partir de hidrolisados de xilano. O xilano é um componente principal de hemicelulose. Assim sendo, o uso de xilose para a produção de etanol e/ou biomassa é muito vantajoso. Além disso, a introdução de um gene de xilose redutase dependente de NAD(P)H no micro-organismo da presente invenção facilita a conversão de xilose em xilulose. Neste caso, para a reação

15 redutora, esta enzima xilose redutase é particularmente apropriada para, por exemplo, carregar ou reciclar (de NADPH para NADP<sup>+</sup>) as coenzimas correspondentes em biorreatores para a produção de aminoácidos.

#### EXEMPLOS

A presente invenção será descrita adicionalmente fazendo referência aos exemplos abaixo. Entretanto, o âmbito da presente invenção não está limitado a eles.

20

##### Exemplo 1

Isolamento e sequenciamento do gene de xilitol desidrogenase a partir de *Candida shehatae* isolamento do DNA cromossomal

25 *Candida shehatae* NBRC1983 foi adquirida do Departamento de Biotecnologia, Instituto de Tecnologia e Avaliação. *Saccharomyces cerevisiae* ATCC60715 foi adquirida da American Type Culture Collection. Estas leveduras foram desenvolvidas separadamente em meio YPD (2% de peptona, 1% de extratos de leveduras, e 2% de glicose) a 25°C até um estágio

30 quiescente. Os DNA's cromossômicos foram isolados a partir delas usando Dr. GenTLE (Takara Bio Inc., Otsu-shi, Siga, Japão), de acordo com o protocolo aqui incluído.

Determinação da sequência de nucleotídeos do gene de xilitol desidrogenase de *Candida shehatae*

As sequências das xilitol desidrogenases dos DNA's de *Pichia stipitis* e *Candida tropicalis* (nºs de acesso DD278642 e DQ201637, respectivamente) foram obtidas a partir da base de dados GenBank de sequências de ácidos nucleicos, e submetidas a uma comparação de sequências usando o software de análise de genes Genetyx (Genetyx Corp, Tóquio, Japão) para obter as sequências de SE I NOS: 3 e 4 a partir de regiões altamente similares. Estas duas sequências foram usadas como iniciadores para realizar a PCR usando o DNA cromossômico de *Candida shehatae* NBRC1983 como modelo, TaKaRa Ex Taq (Takara Bio Inc.) e GeneAmp PCR System 9700 (Applied Biosystems Japan). O programa da reação foi o seguinte: 94°C por 1 minuto (94°C por 30 segundos, 55°C por 30 segundos, e 72°C por 3 minutos) x 35 ciclos, e 72°C por 7 minutos. O produto da PCR obtido passou por eletroforese sobre um gel de agarose a 1%. Como resultado, foi observado um fragmento de DNA ampliado com um tamanho de aproximadamente 900 pb.

O fragmento assim obtido foi clonado usando o *Kit* de Clonagem TOPO TA (Invitrogen Corp., Tóquio, Japão). Os plasmídeos foram coletados a partir do transformante obtido e submetidos a uma reação de sequenciamento usando o *Kit* de Sequenciamento em Ciclos Big Dye Terminator v3.1 (Applied Biosystems, Tóquio, Japão). A sequência de nucleotídeos foi determinada usando o Analisador Genético de DNA 3100 (Applied Biosystems, Tóquio, Japão). As sequências de SEQ ID NOS: 5 e 6 foram obtidas a partir da sequência de nucleotídeos resultante.

O DNA cromossômico de *Candida shehatae* NBRC1983 foi clivado com várias enzimas de restrição e autoligado, e os produtos da autoligação foram coletados por precipitação com etanol. A PCR foi realizada usando este produto como modelo, as sequências de SEQ ID NOS; 5 e 6 como iniciadores, e TaKaRa Ex Taq (Takara Bio Inc.). O programa da reação foi o seguinte: 94°C por 1 minuto (98°C por 20 segundos e 68°C por 10 minutos) x 30 ciclos, e 72°C por 7 minutos. O produto da PCR obtido passou

por eletroforese sobre um gel da agarose a 1%. Como resultado, foi observado um fragmento de DNA ampliado com um tamanho de aproximadamente 2.300 pb.

O fragmento assim obtido foi clonado usando o *Kit* de Clonagem TOPO TA (Invitrogen Corp., Tóquio, Japão). Os plasmídeos foram coletados a partir do transformante obtido e submetidos a uma reação de sequenciamento usando o *Kit* de Sequenciamento em Ciclos Big Dye Terminator v3.1 (Applied Biosystems, Tóquio, Japão). A sequência de nucleotídeos foi determinada usando o Analisador Genético de DNA 3100 (Applied Biosystems, Tóquio, Japão).

A sequência de nucleotídeos obtida foi montada com a sequência de nucleotídeos obtida anteriormente de aproximadamente 900 pb correspondente à sequência parcial do gene de xilitol desidrogenase para obter a sequência de nucleotídeos de uma região codificadora no gene de xilitol desidrogenase de *Candida shehatae*. O resultado está ilustrado em SEQ ID NO: 2.

#### Exemplo 2

Construção do vetor de expressão de xilitol desidrogenase a introdução na levedura do mesmo

Um vetor de expressão de xilitol desidrogenase foi construído de acordo com os procedimentos ilustrados na figura 1. Aqui doravante a construção será descrita detalhadamente. Neste contexto, as condições da PCR descritas abaixo são as seguintes: 94°C por 1 minuto (94°C por 30 segundos, 55°C por 30 segundos, e 72°C por 10 minutos) x 35 ciclos, e 72°C por 7 minutos.

pBluescript II KS+ (Stratagene) foi digerido com KpnI e SacI, e o fragmento mais curto foi removido usando Chroma Spin TE100 (Invitrogen Corp.). Por outro lado, os oligo DNA's de SEQ ID NOS: 7 e 8 foram desnaturados aquecendo a 95°C por 5 minutos, e depois deixados à temperatura ambiente para anelação espontânea. Este produto de anelação foi clonado no sítio KpnI-SacI no pBluescript II KS+ para construir um vetor pBSNS que compreende os sítios NheI e SpeI inseridos entre os sítios KpnI e SacI do pBluescript II KS+.

Os iniciadores (SEQ ID NOS: 17 e 18) para ampliar o gene de xilitol desidrogenase de *Candida shehatae* foram preparados baseado na sequência de SEQ ID NO: 2. Os oligo DNA's de SEQ ID NOS: 6 a 16, 19, e 20 foram preparados baseado na sequência genômica de *Pichis stipitis*.

5 Um fragmento do promotor TDH3 (PsTDH3p) foi obtido da seguinte maneira: uma combinação dos oligo DNA's de SEQ ID NOS: 9 e 10 e uma combinação dos oligo DNA's de SEQ ID NOS: 11 e 12 foram usados separadamente como iniciadores para realizar a PCR, usando o DNA cromossômico de *Pichis stipitis* como modelo. Os iniciadores foram removidos  
10 dos respectivos produtos obtidos usando colunas de filtração com gel. Estes dois produtos foram misturados e usados como modelo na PCR, usando os oligo DNA's de SEQ ID NOS: 9 e 12 como iniciadores para obter um fragmento do promotor TDH3. O fragmento de promotor obtido tem NheI na extremidade 5' e os sítios BamHI e Apal nesta ordem na extremidade 3' e é ainda isento de NheI dentro da sequência do promotor.  
15

Um fragmento terminador de TDH3 (PsTDH3t) foi obtido por PCR usando uma combinação dos oligo DNA's de SEQ ID NOS: 13 e 14 como iniciadores e o DNA cromossômico de *Pichis stipitis* como modelo. O fragmento terminador obtido tem os sítios BamHI e Apal nesta ordem na extremidade 5' e um sítio SpeI na extremidade 3'.  
20

Estes fragmentos são desenhados de tal modo que a extremidade 3' do fragmento do promotor TDH3 se equipare com a extremidade 5' do terminador TDH3. Estes dois fragmentos foram misturados e usados como modelo na PCR usando os oligo DNA's de SEQ ID NOS: 9 e 14 como iniciadores para preparar um fragmento que compreende o promotor TDH3 e o terminador de TDH3 ligados por intermédio dos sítios BamHI e Apal. O fragmento obtido foi inserido por intermédio de um sítio NheI na extremidade 5' e um sítio SpeI na extremidade 3' até o sítio NheI-SpeI do pBSNS para construir pPsExpBS.  
25

30 A PCR foi realizada usando o DNA cromossômico de *Candida shehatae* como modelo e os oligo DNA's de SEQ ID NOS: 17 e 18 como iniciadores para obter o fragmento do gene de xilitol desidrogenase de *Can-*

*dida shehatae* (CsXYL2). A PCR foi realizada usando o DNA cromossômico de *Pichis stipitis* como modelo e os oligo DNA's de SEQ ID NOS: 15 e 16 como iniciadores para obter um fragmento do gene de xilose desidrogenase de *Pichis stipitis* (PsXYL1; SEQ ID NO: 23). A sequência de aminoácidos da proteína de xilose redutase traduzida a partir da sequência de nucleotídeos do fragmento de DNA obtido está ilustrada em SEQ ID NO: 24.

Um fragmento do gene de xilulocinase de *Pichis stipitis* (PsXYL3; SEQ ID NO: 25) isento de um sítio XbaI originalmente presente na sequência foi obtido da seguinte maneira: uma combinação dos oligo DNA's de SEQ ID NOS: 19 e 20 e uma combinação dos oligo DNA's de SEQ ID NOS: 21 e 22 foram usados separadamente como iniciadores para realizar a PCR, usando o DNA cromossômico de *Pichis stipitis* como modelo. Os iniciadores foram removidos dos respectivos produtos obtidos usando colunas de filtração com gel. Estes dois produtos foram misturados e usados como modelo na PCR, usando os oligo DNA's de SEQ ID NOS: 19 e 22 como iniciadores para obter um fragmento do gene de xilulocinase de *Pichis stipitis* (PsXYL3; SEQ ID NO: 25). A sequência de aminoácidos de uma proteína de xilulocinase traduzida a partir da sequência de nucleotídeos do fragmento de DNA obtido está ilustrada em SEQ ID NO: 26.

Cada um dos fragmentos gênicos obtidos CsXYL2, PsXYL1, e PsXYL3 tem um sítio BamHI na extremidade 5' e um sítio ApaI na extremidade 3'. Cada fragmento gênico foi inserido por intermédio destes sítios de restrição em um sítio BamHI-ApaI do vetor pPsExpBS. Os vetores obtidos foram designados como CsXYL2ExpBS, PsXYL1ExpBS, e PsXYL3ExpBS, respectivamente.

O CsXYL2ExpBS foi digerido com SpeI e subsequentemente submetido a uma reação de desfosforilação. Um fragmento NheI-SpeI (contendo o cassete de expressão de PsXYL3) preparado a partir de CsXYL3ExpBS foi ligado ao sítio SpeI do CsXYL2ExpBS digerido, para obter um plasmídeo CPxBS que compreende o cassete de expressão de PsXYL3 inserido na extremidade 3' do cassete de expressão de CsXYL2 na mesma orientação. A seguir, CPxBS foi digerido com SpeI e subsequentemente

submetido a uma reação de desfosforilação. Um fragmento NheI-SpeI (contendo o cassete de expressão de PsXYL1) preparado a partir de PsXYL1ExpBS foi ligado ao sítio SpeI do CPxBS digerido para obter um plasmídeo CPPxBS que compreende o cassete de expressão de CsXYL2, o cassete de expressão de PsXYL3, e o cassete de expressão de PsXYL1 inseridos nesta ordem na mesma orientação.

CPPxBS foi digerido com NheI e SpeI, e o fragmento NheI-SpeI que contém o cassete de expressão de CsXYL2, o cassete de expressão de PsXYL3, e o cassete de expressão de PsXYL1 foi purificado e inserido no sítio SpeI de pYES2 (Invitrogen Corp.) para obter CPPxYES.

Uma cepa variante de *ura3*, KY1094, foi obtida de acordo com um método-padrão usando Wine Yeast Kyokai N° 3 (The Brewing Society of Japan). CPPxYES ou pYES2 foi introduzido em KY1094 por um método de lítio. Como resultado, um transformante de *Saccharomyces cerevisiae* que compreende o vetor (CPPxYES) que compreende o cassete de expressão de CsXYL2, o cassete de expressão de PsXYL3, e o cassete de expressão de PsXYL1 nesta ordem na mesma orientação, bem como um transformante de *Saccharomyces cerevisiae* (controle) que compreende o vetor pYES2 isento destas inserções de cassetes de expressão, foi obtido.

### Exemplo 3

#### Medição da atividade de xilitol desidrogenase

Os transformantes de *Saccharomyces cerevisiae* obtidos foram submetidos à medição da atividade de xilitol desidrogenase como indicado abaixo.

Para a pré-cultura, as cepas foram inoculadas em 2 mL de SD-ura (2% de glicose) por tubo de ensaio e cultivadas sob vascoejamento a 25°C por 24 horas. 200 µL da solução de pré-cultura contendo as cepas foram inoculados em 5 mL de SD-ura (2% de glicose) por tubo de ensaio e cultivados à mão (*main-cultured*) a 120 rpm a 25°C por 24 horas. Depois da cultura, as leveduras foram coletadas por centrifugação, lavas com uma solução de tampão de fosfato de sódio 100 mM (pH 7,0), EDTA 1 mM, e mercaptoetanol 5 mM, e depois colocadas em suspensão em 1 mL da mesma

solução de tampão. A esta suspensão de leveduras adicionou-se 1 g de Pérolas de Vidro de 212-300 microns (Sigma), e as leveduras foram rompidas a 4°C por 15 min, usando MIX-TOWER A14 (Taiyo). Depois de centrifugação a 120 rpm 4°C por 5 min, o sobrenadante foi coletado e usado como uma solução de enzima bruta. Solução de enzima bruta foi medida quanto à concentração de proteína pelo método de Bradford. Albumina de soro bovino foi usada como proteína-padrão.

Uma reação foi realizada a 30°C por 10 min em uma sistema de reação que contém Rris-HCl 100 mM (pH 8,0), cloreto de magnésio 5 mM, NAD<sup>+</sup> 3 mM, xilitol 100 mM, e 5% (v/v) de solução de enzima bruta. A absorvância foi medida em um comprimento de onda de 340 nm. Uma curva de calibração foi preparada usando NADH, e 1 unidade da atividade de enzima foi calculada como um nível de enzima (atividade específica) que produz 1 µmol de NADH por 1 minuto por miligrama de proteína. O resultado está indicado abaixo. Neste contexto, o valor numérico é indicado em média ± desvio-padrão a partir de 9 lotes.

Tabela 1

| Cepa    | Atividade Específica<br>(µmol/min·mg de proteína) |
|---------|---------------------------------------------------|
| CPPxYES | 1,878 ± 0,166                                     |
| pYES    | 0,000 ± 0,008                                     |

A atividade da enzima foi detectada apenas na cepa transformada com o vetor portador de cassetes de expressão que incluem o gene de xilitol desidrogenase de *Candida shehatae*. Isto demonstrou que a proteína expressada a partir do gene de xilitol desidrogenase de *Candida shehatae* tem uma atividade desejada.

#### Aplicabilidade Industrial

O DNA e o micro-organismo da presente invenção são úteis na produção de bioetanol, produção de biomassa, indústria alimentícia, e similares.

#### Texto Livre na Listagem de Sequências

SEQ ID NOS: 3, 4, e 7 a 22: iniciadores.

## Sequência de Listagem

<110> Kirin Holdings Kabushiki Kaisha  
 <120> DNA que codifica Xilitol Desidrogenase  
 <130> PH-3469-PCT  
 5 <150> JP2007-024666  
 <151> 2007-0202  
 <160> 26  
 <170> PatentIn versão 3.4  
 <210> 1  
 10 <211> 364  
 <212> PRT  
 <213> Candida shehatae  
 <400> 1  
 Met Thr Ala Asn Pro Ser Leu Val Leu Asn Lys Ile Asp Asp Ile Thr  
 15 1 5 10 15  
 Phe Glu Ser Tyr Asp Ala Pro Glu Ile Thr Glu Pro Thr Asp Val Leu  
 20 25 30  
 Val Glu Val Lys Lys Thr Gly Ile Cys Gly Ser Asp Ile His Tyr Tyr  
 35 40 45  
 20 Ala His Gly Lys Ile Gly Asn Phe Val Leu Thr Lys Pro Met Val Leu  
 50 55 60  
 Gly His Glu Ser Ser Gly Val Val Thr Lys Val Gly Thr Gly Val Thr  
 65 70 75 80  
 Ser Leu Lys Val Gly Asp Lys Val Ala Ile Glu Pro Gly Ile Pro Ser  
 25 85 90 95  
 Arg Phe Ser Asp Ala Tyr Lys Ser Gly His Tyr Asn Leu Cys Pro His  
 100 105 110  
 Met Cys Phe Ala Ala Thr Pro Asn Ser Thr Glu Gly Glu Pro Asn Pro  
 115 120 125  
 30 Pro Gly Thr Leu Cys Lys Tyr Phe Lys Ser Pro Glu Asp Phe Leu Val  
 130 135 140  
 Lys Leu Pro Glu His Val Ser Leu Glu Met Gly Ala Leu Val Glu Pro

145                      150                      155                      160  
 Leu Ser Val Gly Val His Ala Ser Lys Leu Ala Ser Val Arg Phe Gly  
                          165                      170                      175  
 Asp Tyr Val Ala Val Phe Gly Ala Gly Pro Val Gly Leu Leu Ala Ala  
 5                      180                      185                      190  
 Ala Val Ala Lys Thr Phe Gly Ala Lys Gly Val Ile Val Ile Asp Ile  
                          195                      200                      205  
 Phe Asp Asn Lys Leu Gln Met Ala Lys Asp Ile Gly Ala Ala Thr His  
                          210                      215                      220  
 10 Ile Phe Asn Ser Lys Thr Gly Gly Asp Ala Ala Ala Leu Val Lys Ala  
 225                      230                      235                      240  
 Phe Asp Gly Arg Glu Pro Thr Val Val Leu Glu Cys Thr Gly Ala Glu  
                          245                      250                      255  
 Pro Cys Ile Asn Gln Gly Val Ala Ile Leu Ala Gln Gly Gly Arg Phe  
 15                      260                      265                      270  
 Val Gln Val Gly Asn Ala Pro Gly Pro Val Lys Phe Pro Ile Thr Glu  
                          275                      280                      285  
 Phe Ala Thr Lys Glu Leu Thr Leu Phe Gly Ser Phe Arg Tyr Gly Phe  
                          290                      295                      300  
 20 Asn Asp Tyr Lys Thr Ser Val Asp Ile Met Asp Thr Asn Tyr Lys Asn  
 305                      310                      315                      320  
 Gly Lys Glu Lys Ala Pro Ile Asp Phe Glu Gln Leu Ile Thr His Arg  
                          325                      330                      335  
 Phe Lys Phe Asp Asp Ala Ile Lys Ala Tyr Asp Leu Val Arg Ala Gly  
 25                      340                      345                      350  
 Ser Gly Ala Val Lys Cys Ile Ile Asp Gly Pro Glu  
                          355                      360  
 <210> 2  
 <211> 1095  
 30 <212> DNA  
 <213> Candida shehatae  
 <400> 2

|    |                                                                     |      |
|----|---------------------------------------------------------------------|------|
|    | atgactgcta acccatcgct cgtgcttaac aagatcgacg acatcacctt cgagtcgtac   | 60   |
|    | gatgccccag aaatcaccga gccaacagac gttctcgtgg aagtcaagaa aaccgggtatc  | 120  |
|    | tgtggttccg atatccacta ctacgcccac ggtaagatcg gtaacttcgt gttgaccaag   | 180  |
|    | ccaatggttc ttggccacga atcctcaggt gtcgtcacca aggtcggtag cggcgtcacc   | 240  |
| 5  | tcgctcaagg taggtgacaa ggtcgccatt gagccaggta ttccatccag attcagtgac   | 300  |
|    | gcctacaaga gcggtcacta caacttatgt ccacacatgt gcttcgccgc cactccaaac   | 360  |
|    | tccaccgagg gcgagccaaa cccaccaggt accttatgta agtacttcaa gtccccagag   | 420  |
|    | gacttcttgg tcaagttgcc agaacacggt tccttggaag tgggtgctct tgcgagcca    | 480  |
|    | ttgtccgtcg gtgtccacgc ctccaagttg gcttcgcgtca gattcggtag ctacgtcgct  | 540  |
| 10 | gttttcggtg cgggtccagt cgggtctctta gctgctgccg tcgccaagac cttcgggtgcc | 600  |
|    | aagggtgtca ttgtcattga cattttcgac aacaagttgc aaatggccaa ggacattggg   | 660  |
|    | gctgtacccc acatcttcaa ctccaagacc ggtggtgacg ccgctgcctt gtcaaggct    | 720  |
|    | ttcgacggcc gcgagccaac cgtcgtcttg gaatgtactg gtgctgagcc atgtatcaac   | 780  |
|    | caaggtgtcg ctatcttggc ccaaggtggg cgtttcgtcc aagtcggtaa cgtcccaggt   | 840  |
| 15 | ccagttaagt tcccaatcac tgaattcgct accaaggaa tcaccttggt cggctctttc    | 900  |
|    | agatacgggt tcaacgatta caagacctct gtcgacatca tggacaccaa ctacaagaac   | 960  |
|    | ggtaaggaaa aggccccaat tgacttcgag caattgatca cccacagatt caagttcgac   | 1020 |
|    | gacgccatca aggcctacga cttggtcaga gctggtagtg gtgctgtcaa gtgtatcatt   | 1080 |
|    | gacggtcctg agtaa                                                    | 1095 |
| 20 | <210> 3                                                             |      |
|    | <211> 25                                                            |      |
|    | <212> DNA                                                           |      |
|    | <213> Artificial                                                    |      |
|    | <220>                                                               |      |
| 25 | <223> iniciador                                                     |      |
|    | <400> 3                                                             |      |
|    | agaaaaccgg tatctgtggt tccga                                         | 25   |
|    | <210> 4                                                             |      |
|    | <211> 25                                                            |      |
| 30 | <212> DNA                                                           |      |
|    | <213> Artificial                                                    |      |
|    | <220>                                                               |      |

<223> iniciador  
 <400> 4  
 tgaacttgta tctgtgggtg atcaa 25  
 <210> 5  
**5** <211> 30  
 <212> DNA  
 <213> Candida shehatae  
 <400> 5  
 tagtggatat cggaaccaca gataccggtt 30  
**10** <210> 6  
 <211> 30  
 <212> DNA  
 <213> Candida shehatae  
 <400> 6  
**15** caattgatca cccacagata caagttcaaa 30  
 <210> 7  
 <211> 34  
 <212> DNA  
 <213> Artificial  
**20** <220>  
 <223> iniciador  
 <400> 7  
 aaaaaaaaaa cgctagcccc gggactagtg agct 34  
 <210> 8  
**25** <211> 24  
 <212> DNA  
 <213> Artificial  
 <220>  
 <223> iniciador  
**30** <400> 8  
 cactagtccc ggggctagcg gtac 24  
 <210> 9

<211> 38

<212> DNA

<213> Artificial

<220>

5 <223> iniciador

<400> 9

cggctagcga tgactctgta gaaagttgag tcaaatgc 38

<210> 10

<211> 25

10 <212> DNA

<213> Artificial

<220>

<223> iniciador

<400> 10

15 ctggaaagtg ctcgcaaaaa agcct 25

<210> 11

<211> 25

<212> DNA

<213> Artificial

20 <220>

<223> iniciador

<400> 11

aggctttttt gcgagcactt tccag 25

<210> 12

25 <211> 45

<212> DNA

<213> Artificial

<220>

<223> iniciador

30 <400> 12

tgttggggccc ttttggatcc gccgaaaccg ttaataccaa tcttg 45

<210> 13

<211> 44  
 <212> DNA  
 <213> Artificial  
 <220>  
**5** <223> iniciador  
 <400> 13  
 ggcggatcca aaagggccca acacgttgcc aaggtttctg ctta 44  
 <210> 14  
 <211> 33  
**10** <212> DNA  
 <213> Artificial  
 <220>  
 <223> iniciador  
 <400> 14  
**15** ggactagtag gggatatccac ctattagtaa ttc 33  
 <210> 15  
 <211> 37  
 <212> DNA  
 <213> Artificial  
**20** <220>  
 <223> iniciador  
 <400> 15  
 cgggatccga ggtgaccgat ggggtgcaa ttaatgt 37  
 <210> 16  
**25** <211> 36  
 <212> DNA  
 <213> Artificial  
 <220>  
 <223> iniciador  
**30** <400> 16  
 cggggccccc ttcttagacg aagataggaa tcttgt 36  
 <210> 17

<211> 35  
 <212> DNA  
 <213> Artificial  
 <220>  
**5** <223> iniciador  
 <400> 17  
 cgggataccta gtccgctcta gttataccct acaaa 35  
 <210> 18  
 <211> 35  
**10** <212> DNA  
 <213> Artificial  
 <220>  
 <223> iniciador  
 <400> 18  
**15** cggggcccggt ccaattattt actcaggacc gtcaa 35  
 <210> 19  
 <211> 35  
 <212> DNA  
 <213> Artificial  
**20** <220>  
 <223> iniciador  
 <400> 19  
 cgggatccaa atactcacgt agttgacact cacaa 35  
 <210> 20  
**25** <211> 25  
 <212> DNA  
 <213> Artificial  
 <220>  
 <223> iniciador  
**30** <400> 20  
 ctcggccggtt ctggaccagt atacc 25  
 <210> 21

<211> 24  
 <212> DNA  
 <213> Artificial  
 <220>  
**5** <223> iniciador  
 <400> 21  
 ggtatactgg tccagaacgg ccga 24  
 <210> 22  
 <211> 33  
**10** <212> DNA  
 <213> Artificial  
 <220>  
 <223> iniciador  
 <400> 22  
**15** cggggcccta acgtctatcg tgatattcgc aca 33  
 <210> 23  
 <211> 957  
 <212> DNA  
 <213> Pichia stipitis  
**20** <220>  
 <221> CDS  
 <222> (1)..(957)  
 <220>  
 <223> PsXYL1  
**25** <400> 23  
 atg cct tct att aag ttg aac tct ggt tac gac atg cca gcc gtc ggt 48  
 Met Pro Ser Ile Lys Leu Asn Ser Gly Tyr Asp Met Pro Ala Val Gly  
 1 5 10 15  
 ttc ggc tgt tgg aaa gtc gac gtc gac acc tgt tct gaa cag atc tac 96  
**30** Phe Gly Cys Trp Lys Val Asp Val Asp Thr Cys Ser Glu Gln Ile Tyr  
 20 25 30  
 cgt gct atc aag acc ggt tac aga ttg ttc gac ggt gcc gaa gat tac 144

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    | Arg | Ala | Ile | Lys | Thr | Gly | Tyr | Arg | Leu | Phe | Asp | Gly | Ala | Glu | Asp | Tyr |     |
|    |     |     |     | 35  |     |     |     | 40  |     |     |     |     | 45  |     |     |     |     |
|    | gcc | aac | gaa | aag | tta | ggt | ggt | gcc | ggt | gtc | aag | aag | gcc | att | gac | gaa | 192 |
|    | Ala | Asn | Glu | Lys | Leu | Val | Gly | Ala | Gly | Val | Lys | Lys | Ala | Ile | Asp | Glu |     |
| 5  |     |     |     | 50  |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
|    | ggt | atc | gtc | aag | cgt | gaa | gac | ttg | ttc | ctt | acc | tcc | aag | ttg | tgg | aac | 240 |
|    | Gly | Ile | Val | Lys | Arg | Glu | Asp | Leu | Phe | Leu | Thr | Ser | Lys | Leu | Trp | Asn |     |
|    | 65  |     |     |     |     | 70  |     |     |     | 75  |     |     |     |     | 80  |     |     |
|    | aac | tac | cac | cac | cca | gac | aac | gtc | gaa | aag | gcc | ttg | aac | aga | acc | ctt | 288 |
| 10 | Asn | Tyr | His | His | Pro | Asp | Asn | Val | Glu | Lys | Ala | Leu | Asn | Arg | Thr | Leu |     |
|    |     |     |     |     | 85  |     |     |     |     | 90  |     |     |     | 95  |     |     |     |
|    | tct | gac | ttg | caa | ggt | gac | tac | ggt | gac | ttg | ttc | ttg | atc | cac | ttc | cca | 336 |
|    | Ser | Asp | Leu | Gln | Val | Asp | Tyr | Val | Asp | Leu | Phe | Leu | Ile | His | Phe | Pro |     |
|    |     |     |     | 100 |     |     |     |     | 105 |     |     |     | 110 |     |     |     |     |
| 15 | gtc | acc | ttc | aag | ttc | ggt | cca | tta | gaa | gaa | aag | tac | cca | cca | gga | ttc | 384 |
|    | Val | Thr | Phe | Lys | Phe | Val | Pro | Leu | Glu | Glu | Lys | Tyr | Pro | Pro | Gly | Phe |     |
|    |     |     |     | 115 |     |     |     | 120 |     |     |     |     | 125 |     |     |     |     |
|    | tac | tgt | ggt | aag | ggt | gac | aac | ttc | gac | tac | gaa | gat | ggt | cca | att | tta | 432 |
|    | Tyr | Cys | Gly | Lys | Gly | Asp | Asn | Phe | Asp | Tyr | Glu | Asp | Val | Pro | Ile | Leu |     |
| 20 |     |     |     | 130 |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
|    | gag | acc | tgg | aag | gct | ctt | gaa | aag | ttg | gtc | aag | gcc | ggt | aag | atc | aga | 480 |
|    | Glu | Thr | Trp | Lys | Ala | Leu | Glu | Lys | Leu | Val | Lys | Ala | Gly | Lys | Ile | Arg |     |
|    | 145 |     |     |     | 150 |     |     |     | 155 |     |     |     | 160 |     |     |     |     |
|    | tct | atc | ggt | ggt | tct | aac | ttc | cca | ggt | gct | ttg | ctc | ttg | gac | ttg | ttg | 528 |
| 25 | Ser | Ile | Gly | Val | Ser | Asn | Phe | Pro | Gly | Ala | Leu | Leu | Leu | Asp | Leu | Leu |     |
|    |     |     |     | 165 |     |     |     |     | 170 |     |     |     | 175 |     |     |     |     |
|    | aga | ggt | gct | acc | atc | aag | cca | tct | gtc | ttg | caa | ggt | gaa | cac | cac | cca | 576 |
|    | Arg | Gly | Ala | Thr | Ile | Lys | Pro | Ser | Val | Leu | Gln | Val | Glu | His | His | Pro |     |
|    |     |     |     | 180 |     |     |     | 185 |     |     |     |     | 190 |     |     |     |     |
| 30 | tac | ttg | caa | caa | cca | aga | ttg | atc | gaa | ttc | gct | caa | tcc | cgt | ggt | att | 624 |
|    | Tyr | Leu | Gln | Gln | Pro | Arg | Leu | Ile | Glu | Phe | Ala | Gln | Ser | Arg | Gly | Ile |     |
|    |     |     |     | 195 |     |     |     | 200 |     |     |     |     | 205 |     |     |     |     |

gct gtc acc gct tac tct tcg ttc ggt cct caa tct ttc gtt gaa ttg 672  
 Ala Val Thr Ala Tyr Ser Ser Phe Gly Pro Gln Ser Phe Val Glu Leu  
 210 215 220  
 aac caa ggt aga gct ttg aac act tct cca ttg ttc gag aac gaa act 720  
 5 Asn Gln Gly Arg Ala Leu Asn Thr Ser Pro Leu Phe Glu Asn Glu Thr  
 225 230 235 240  
 atc aag gct atc gct gct aag cac ggt aag tct cca gct caa gtc ttg 768  
 Ile Lys Ala Ile Ala Ala Lys His Gly Lys Ser Pro Ala Gln Val Leu  
 245 250 255  
 10 ttg aga tgg tct tcc caa aga ggc att gcc atc att cca aag tcc aac 816  
 Leu Arg Trp Ser Ser Gln Arg Gly Ile Ala Ile Ile Pro Lys Ser Asn  
 260 265 270  
 act gtc cca aga ttg ttg gaa aac aag gac gtc aac agc ttc gac ttg 864  
 Thr Val Pro Arg Leu Leu Glu Asn Lys Asp Val Asn Ser Phe Asp Leu  
 15 275 280 285  
 gac gaa caa gat ttc gct gac att gcc aag ttg gac atc aac ttg aga 912  
 Asp Glu Gln Asp Phe Ala Asp Ile Ala Lys Leu Asp Ile Asn Leu Arg  
 290 295 300  
 ttc aac gac cca tgg gac tgg gac aag att cct atc ttc gtc taa 957  
 20 Phe Asn Asp Pro Trp Asp Trp Asp Lys Ile Pro Ile Phe Val  
 305 310 315  
 <210> 24  
 <211> 318  
 <212> PRT  
 25 <213> Pichia stipitis  
 <220>  
 <223> PsXYL1  
 <400> 24  
 Met Pro Ser Ile Lys Leu Asn Ser Gly Tyr Asp Met Pro Ala Val Gly  
 30 1 5 10 15  
 Phe Gly Cys Trp Lys Val Asp Val Asp Thr Cys Ser Glu Gln Ile Tyr  
 20 25 30

Arg Ala Ile Lys Thr Gly Tyr Arg Leu Phe Asp Gly Ala Glu Asp Tyr  
 35 40 45  
 Ala Asn Glu Lys Leu Val Gly Ala Gly Val Lys Lys Ala Ile Asp Glu  
 50 55 60  
 5 Gly Ile Val Lys Arg Glu Asp Leu Phe Leu Thr Ser Lys Leu Trp Asn  
 65 70 75 80  
 Asn Tyr His His Pro Asp Asn Val Glu Lys Ala Leu Asn Arg Thr Leu  
 85 90 95  
 Ser Asp Leu Gln Val Asp Tyr Val Asp Leu Phe Leu Ile His Phe Pro  
 10 100 105 110  
 Val Thr Phe Lys Phe Val Pro Leu Glu Glu Lys Tyr Pro Pro Gly Phe  
 115 120 125  
 Tyr Cys Gly Lys Gly Asp Asn Phe Asp Tyr Glu Asp Val Pro Ile Leu  
 130 135 140  
 15 Glu Thr Trp Lys Ala Leu Glu Lys Leu Val Lys Ala Gly Lys Ile Arg  
 145 150 155 160  
 Ser Ile Gly Val Ser Asn Phe Pro Gly Ala Leu Leu Leu Asp Leu Leu  
 165 170 175  
 Arg Gly Ala Thr Ile Lys Pro Ser Val Leu Gln Val Glu His His Pro  
 20 180 185 190  
 Tyr Leu Gln Gln Pro Arg Leu Ile Glu Phe Ala Gln Ser Arg Gly Ile  
 195 200 205  
 Ala Val Thr Ala Tyr Ser Ser Phe Gly Pro Gln Ser Phe Val Glu Leu  
 210 215 220  
 25 Asn Gln Gly Arg Ala Leu Asn Thr Ser Pro Leu Phe Glu Asn Glu Thr  
 225 230 235 240  
 Ile Lys Ala Ile Ala Ala Lys His Gly Lys Ser Pro Ala Gln Val Leu  
 245 250 255  
 Leu Arg Trp Ser Ser Gln Arg Gly Ile Ala Ile Ile Pro Lys Ser Asn  
 30 260 265 270  
 Thr Val Pro Arg Leu Leu Glu Asn Lys Asp Val Asn Ser Phe Asp Leu  
 275 280 285

```

Asp Glu Gln Asp Phe Ala Asp Ile Ala Lys Leu Asp Ile Asn Leu Arg
      290              295              300
Phe Asn Asp Pro Trp Asp Trp Asp Lys Ile Pro Ile Phe Val
305              310              315

5  <210> 25
    <211> 1872
    <212> DNA
    <213> Artificial
    <220>
10 <221> CDS
    <222> (1)..(1872)
    <223> PsXYL3_CDS_XbaI(-)
    <400> 25

15 atg acc act acc cca ttt gat gct cca gat aag ctc ttc ctc ggg ttc      48
    Met Thr Thr Thr Pro Phe Asp Ala Pro Asp Lys Leu Phe Leu Gly Phe
      1              5              10              15
    gat ctt tcg act cag cag ttg aag atc atc gtc acc gat gaa aac ctc      96
    Asp Leu Ser Thr Gln Gln Leu Lys Ile Ile Val Thr Asp Glu Asn Leu
20              20              25              30
    gct gct ctc aaa acc tac aat gtc gag ttc gat agc atc aac agc tct      144
    Ala Ala Leu Lys Thr Tyr Asn Val Glu Phe Asp Ser Ile Asn Ser Ser
      35              40              45
    gtc cag aag ggt gtc att gct atc aac gac gaa atc agc aag ggt gcc      192
25 Val Gln Lys Gly Val Ile Ala Ile Asn Asp Glu Ile Ser Lys Gly Ala
      50              55              60
    att att tcc ccc gtt tac atg tgg ttg gat gcc ctt gac cat gtt ttt      240
    Ile Ile Ser Pro Val Tyr Met Trp Leu Asp Ala Leu Asp His Val Phe
      65              70              75              80
30 gaa gac atg aag aag gac gga ttc ccc ttc aac aag gtt gtt ggt att      288
    Glu Asp Met Lys Lys Asp Gly Phe Pro Phe Asn Lys Val Val Gly Ile
      85              90              95

```

|    |                                                                 |     |
|----|-----------------------------------------------------------------|-----|
|    | tcc ggt tct tgt caa cag cac ggt tcg gta tac tgg tcc aga acg gcc | 336 |
|    | Ser Gly Ser Cys Gln Gln His Gly Ser Val Tyr Trp Ser Arg Thr Ala |     |
|    | 100 105 110                                                     |     |
|    | gag aag gtc ttg tcc gaa ttg gac gct gaa tct tcg tta tcg agc cag | 384 |
| 5  | Glu Lys Val Leu Ser Glu Leu Asp Ala Glu Ser Ser Leu Ser Ser Gln |     |
|    | 115 120 125                                                     |     |
|    | atg aga tct gct ttc acc ttc aag cac gct cca aac tgg cag gat cac | 432 |
|    | Met Arg Ser Ala Phe Thr Phe Lys His Ala Pro Asn Trp Gln Asp His |     |
|    | 130 135 140                                                     |     |
| 10 | tct acc ggt aaa gag ctt gaa gag ttc gaa aga gtg att ggt gct gat | 480 |
|    | Ser Thr Gly Lys Glu Leu Glu Glu Phe Glu Arg Val Ile Gly Ala Asp |     |
|    | 145 150 155 160                                                 |     |
|    | gcc ttg gct gat atc tct ggt tcc aga gcc cat tac aga ttc aca ggg | 528 |
|    | Ala Leu Ala Asp Ile Ser Gly Ser Arg Ala His Tyr Arg Phe Thr Gly |     |
| 15 | 165 170 175                                                     |     |
|    | ctc cag att aga aag ttg tct acc aga ttc aag ccc gaa aag tac aac | 576 |
|    | Leu Gln Ile Arg Lys Leu Ser Thr Arg Phe Lys Pro Glu Lys Tyr Asn |     |
|    | 180 185 190                                                     |     |
|    | aga act gct cgt atc tct tta gtt tcg tca ttt gtt gcc agt gtg ttg | 624 |
| 20 | Arg Thr Ala Arg Ile Ser Leu Val Ser Ser Phe Val Ala Ser Val Leu |     |
|    | 195 200 205                                                     |     |
|    | ctt ggt aga atc acc tcc att gaa gaa gcc gat gct tgt gga atg aac | 672 |
|    | Leu Gly Arg Ile Thr Ser Ile Glu Glu Ala Asp Ala Cys Gly Met Asn |     |
|    | 210 215 220                                                     |     |
| 25 | ttg tac gat atc gaa aag cgc gag ttc aac gaa gag ctc ttg gcc atc | 720 |
|    | Leu Tyr Asp Ile Glu Lys Arg Glu Phe Asn Glu Glu Leu Leu Ala Ile |     |
|    | 225 230 235 240                                                 |     |
|    | gct gct ggt gtc cac cct gag ttg gat ggt gta gaa caa gac ggt gaa | 768 |
|    | Ala Ala Gly Val His Pro Glu Leu Asp Gly Val Glu Gln Asp Gly Glu |     |
| 30 | 245 250 255                                                     |     |
|    | att tac aga gct ggt atc aat gag ttg aag aga aag ttg ggt cct gtc | 816 |
|    | Ile Tyr Arg Ala Gly Ile Asn Glu Leu Lys Arg Lys Leu Gly Pro Val |     |

|    |                                                                 |     |     |      |
|----|-----------------------------------------------------------------|-----|-----|------|
|    | 260                                                             | 265 | 270 |      |
|    | aaa cct ata aca tac gaa agc gaa ggt gac att gcc tct tac ttt gtc |     |     | 864  |
|    | Lys Pro Ile Thr Tyr Glu Ser Glu Gly Asp Ile Ala Ser Tyr Phe Val |     |     |      |
|    | 275                                                             | 280 | 285 |      |
| 5  | acc aga tac ggc ttc aac ccc gac tgt aaa atc tac tcg ttc acc gga |     |     | 912  |
|    | Thr Arg Tyr Gly Phe Asn Pro Asp Cys Lys Ile Tyr Ser Phe Thr Gly |     |     |      |
|    | 290                                                             | 295 | 300 |      |
|    | gac aat ttg gcc acg att atc tcg ttg cct ttg gct cca aat gat gct |     |     | 960  |
|    | Asp Asn Leu Ala Thr Ile Ile Ser Leu Pro Leu Ala Pro Asn Asp Ala |     |     |      |
| 10 | 305                                                             | 310 | 315 | 320  |
|    | ttg atc tca ttg ggt act tct act aca gtt tta att atc acc aag aac |     |     | 1008 |
|    | Leu Ile Ser Leu Gly Thr Ser Thr Thr Val Leu Ile Ile Thr Lys Asn |     |     |      |
|    | 325                                                             | 330 | 335 |      |
|    | tac gct cct tct tct caa tac cat ttg ttt aaa cat cca acc atg cct |     |     | 1056 |
| 15 | Tyr Ala Pro Ser Ser Gln Tyr His Leu Phe Lys His Pro Thr Met Pro |     |     |      |
|    | 340                                                             | 345 | 350 |      |
|    | gac cac tac atg ggc atg atc tgc tac tgt aac ggt tcc ttg gcc aga |     |     | 1104 |
|    | Asp His Tyr Met Gly Met Ile Cys Tyr Cys Asn Gly Ser Leu Ala Arg |     |     |      |
|    | 355                                                             | 360 | 365 |      |
| 20 | gaa aag gtt aga gac gaa gtc aac gaa aag ttc aat gta gaa gac aag |     |     | 1152 |
|    | Glu Lys Val Arg Asp Glu Val Asn Glu Lys Phe Asn Val Glu Asp Lys |     |     |      |
|    | 370                                                             | 375 | 380 |      |
|    | aag tcg tgg gac aag ttc aat gaa atc ttg gac aaa tcc aca gac ttc |     |     | 1200 |
|    | Lys Ser Trp Asp Lys Phe Asn Glu Ile Leu Asp Lys Ser Thr Asp Phe |     |     |      |
| 25 | 385                                                             | 390 | 395 | 400  |
|    | aac aac aag ttg ggt att tac ttc cca ctt ggc gaa att gtc cct aat |     |     | 1248 |
|    | Asn Asn Lys Leu Gly Ile Tyr Phe Pro Leu Gly Glu Ile Val Pro Asn |     |     |      |
|    | 405                                                             | 410 | 415 |      |
|    | gcc gct gct cag atc aag aga tcg gtg ttg aac agc aag aac gaa att |     |     | 1296 |
| 30 | Ala Ala Ala Gln Ile Lys Arg Ser Val Leu Asn Ser Lys Asn Glu Ile |     |     |      |
|    | 420                                                             | 425 | 430 |      |
|    | gta gac gtt gag ttg ggc gac aag aac tgg caa cct gaa gat gat gtt |     |     | 1344 |

|    |                                                                 |      |
|----|-----------------------------------------------------------------|------|
|    | Val Asp Val Glu Leu Gly Asp Lys Asn Trp Gln Pro Glu Asp Asp Val |      |
|    | 435 440 445                                                     |      |
|    | tct tca att gta gaa tca cag act ttg tct tgt aga ttg aga act ggt | 1392 |
|    | Ser Ser Ile Val Glu Ser Gln Thr Leu Ser Cys Arg Leu Arg Thr Gly |      |
| 5  | 450 455 460                                                     |      |
|    | cca atg ttg agc aag agt gga gat tct tct gct tcc agc tct gcc tca | 1440 |
|    | Pro Met Leu Ser Lys Ser Gly Asp Ser Ser Ala Ser Ser Ser Ala Ser |      |
|    | 465 470 475 480                                                 |      |
|    | cct caa cca gaa ggt gat ggt aca gat ttg cac aag gtc tac caa gac | 1488 |
| 10 | Pro Gln Pro Glu Gly Asp Gly Thr Asp Leu His Lys Val Tyr Gln Asp |      |
|    | 485 490 495                                                     |      |
|    | ttg gtt aaa aag ttt ggt gac ttg ttc act gat gga aag aag caa acc | 1536 |
|    | Leu Val Lys Lys Phe Gly Asp Leu Phe Thr Asp Gly Lys Lys Gln Thr |      |
|    | 500 505 510                                                     |      |
| 15 | ttt gag tct ttg acc gcc aga cct aac cgt tgt tac tac gtc ggt ggt | 1584 |
|    | Phe Glu Ser Leu Thr Ala Arg Pro Asn Arg Cys Tyr Tyr Val Gly Gly |      |
|    | 515 520 525                                                     |      |
|    | gct tcc aac aac ggc agc att atc cgc aag atg ggt tcc atc ttg gct | 1632 |
|    | Ala Ser Asn Asn Gly Ser Ile Ile Arg Lys Met Gly Ser Ile Leu Ala |      |
| 20 | 530 535 540                                                     |      |
|    | ccc gtc aac gga aac tac aag gtt gac att cct aac gcc tgt gca ttg | 1680 |
|    | Pro Val Asn Gly Asn Tyr Lys Val Asp Ile Pro Asn Ala Cys Ala Leu |      |
|    | 545 550 555 560                                                 |      |
|    | ggt ggt gct tac aag gcc agt tgg agt tac gag tgt gaa gcc aag aag | 1728 |
| 25 | Gly Gly Ala Tyr Lys Ala Ser Trp Ser Tyr Glu Cys Glu Ala Lys Lys |      |
|    | 565 570 575                                                     |      |
|    | gaa tgg atc gga tac gat cag tat atc aac aga ttg ttt gaa gta agt | 1776 |
|    | Glu Trp Ile Gly Tyr Asp Gln Tyr Ile Asn Arg Leu Phe Glu Val Ser |      |
|    | 580 585 590                                                     |      |
| 30 | gac gag atg aat ctg ttc gaa gtc aag gat aaa tgg ctc gaa tat gcc | 1824 |
|    | Asp Glu Met Asn Leu Phe Glu Val Lys Asp Lys Trp Leu Glu Tyr Ala |      |
|    | 595 600 605                                                     |      |

aac ggg gtt gga atg ttg gcc aag atg gaa agt gaa ttg aaa cac taa 1872  
 Asn Gly Val Gly Met Leu Ala Lys Met Glu Ser Glu Leu Lys His  
 610 615 620  
 <210> 26  
 5 <211> 623  
 <212> PRT  
 <213> Artificial  
 <220>  
 <223> PsXYL3\_CDS\_XbaI(-)  
 10 <400> 26  
 Met Thr Thr Thr Pro Phe Asp Ala Pro Asp Lys Leu Phe Leu Gly Phe  
 1 5 10 15  
 Asp Leu Ser Thr Gln Gln Leu Lys Ile Ile Val Thr Asp Glu Asn Leu  
 20 25 30  
 15 Ala Ala Leu Lys Thr Tyr Asn Val Glu Phe Asp Ser Ile Asn Ser Ser  
 35 40 45  
 Val Gln Lys Gly Val Ile Ala Ile Asn Asp Glu Ile Ser Lys Gly Ala  
 50 55 60  
 Ile Ile Ser Pro Val Tyr Met Trp Leu Asp Ala Leu Asp His Val Phe  
 20 65 70 75 80  
 Glu Asp Met Lys Lys Asp Gly Phe Pro Phe Asn Lys Val Val Gly Ile  
 85 90 95  
 Ser Gly Ser Cys Gln Gln His Gly Ser Val Tyr Trp Ser Arg Thr Ala  
 100 105 110  
 25 Glu Lys Val Leu Ser Glu Leu Asp Ala Glu Ser Ser Leu Ser Ser Gln  
 115 120 125  
 Met Arg Ser Ala Phe Thr Phe Lys His Ala Pro Asn Trp Gln Asp His  
 130 135 140  
 Ser Thr Gly Lys Glu Leu Glu Glu Phe Glu Arg Val Ile Gly Ala Asp  
 30 145 150 155 160  
 Ala Leu Ala Asp Ile Ser Gly Ser Arg Ala His Tyr Arg Phe Thr Gly  
 165 170 175

Leu Gln Ile Arg Lys Leu Ser Thr Arg Phe Lys Pro Glu Lys Tyr Asn  
 180 185 190  
 Arg Thr Ala Arg Ile Ser Leu Val Ser Ser Phe Val Ala Ser Val Leu  
 195 200 205  
**5** Leu Gly Arg Ile Thr Ser Ile Glu Glu Ala Asp Ala Cys Gly Met Asn  
 210 215 220  
 Leu Tyr Asp Ile Glu Lys Arg Glu Phe Asn Glu Glu Leu Leu Ala Ile  
 225 230 235 240  
 Ala Ala Gly Val His Pro Glu Leu Asp Gly Val Glu Gln Asp Gly Glu  
**10** 245 250 255  
 Ile Tyr Arg Ala Gly Ile Asn Glu Leu Lys Arg Lys Leu Gly Pro Val  
 260 265 270  
 Lys Pro Ile Thr Tyr Glu Ser Glu Gly Asp Ile Ala Ser Tyr Phe Val  
 275 280 285  
**15** Thr Arg Tyr Gly Phe Asn Pro Asp Cys Lys Ile Tyr Ser Phe Thr Gly  
 290 295 300  
 Asp Asn Leu Ala Thr Ile Ile Ser Leu Pro Leu Ala Pro Asn Asp Ala  
 305 310 315 320  
 Leu Ile Ser Leu Gly Thr Ser Thr Thr Val Leu Ile Ile Thr Lys Asn  
**20** 325 330 335  
 Tyr Ala Pro Ser Ser Gln Tyr His Leu Phe Lys His Pro Thr Met Pro  
 340 345 350  
 Asp His Tyr Met Gly Met Ile Cys Tyr Cys Asn Gly Ser Leu Ala Arg  
 355 360 365  
**25** Glu Lys Val Arg Asp Glu Val Asn Glu Lys Phe Asn Val Glu Asp Lys  
 370 375 380  
 Lys Ser Trp Asp Lys Phe Asn Glu Ile Leu Asp Lys Ser Thr Asp Phe  
 385 390 395 400  
 Asn Asn Lys Leu Gly Ile Tyr Phe Pro Leu Gly Glu Ile Val Pro Asn  
**30** 405 410 415  
 Ala Ala Ala Gln Ile Lys Arg Ser Val Leu Asn Ser Lys Asn Glu Ile  
 420 425 430

Val Asp Val Glu Leu Gly Asp Lys Asn Trp Gln Pro Glu Asp Asp Val  
435 440 445

Ser Ser Ile Val Glu Ser Gln Thr Leu Ser Cys Arg Leu Arg Thr Gly  
450 455 460

5 Pro Met Leu Ser Lys Ser Gly Asp Ser Ser Ala Ser Ser Ser Ala Ser  
465 470 475 480

Pro Gln Pro Glu Gly Asp Gly Thr Asp Leu His Lys Val Tyr Gln Asp  
485 490 495

Leu Val Lys Lys Phe Gly Asp Leu Phe Thr Asp Gly Lys Lys Gln Thr  
10 500 505 510

Phe Glu Ser Leu Thr Ala Arg Pro Asn Arg Cys Tyr Tyr Val Gly Gly  
515 520 525

Ala Ser Asn Asn Gly Ser Ile Ile Arg Lys Met Gly Ser Ile Leu Ala  
530 535 540

15 Pro Val Asn Gly Asn Tyr Lys Val Asp Ile Pro Asn Ala Cys Ala Leu  
545 550 555 560

Gly Gly Ala Tyr Lys Ala Ser Trp Ser Tyr Glu Cys Glu Ala Lys Lys  
565 570 575

Glu Trp Ile Gly Tyr Asp Gln Tyr Ile Asn Arg Leu Phe Glu Val Ser  
20 580 585 590

Asp Glu Met Asn Leu Phe Glu Val Lys Asp Lys Trp Leu Glu Tyr Ala  
595 600 605

Asn Gly Val Gly Met Leu Ala Lys Met Glu Ser Glu Leu Lys His  
610 615 620

## REIVINDICAÇÕES

1. DNA, caracterizado pelo fato de que codifica um polipeptídeo selecionado entre:

5 (a) um polipeptídeo que compreende a sequência de aminoácidos de SEQ ID NO: 1, e tem uma atividade de xilitol desidrogenase;

(b) um polipeptídeo que compreende uma sequência de aminoácidos que tem 90% ou mais de identidade com a sequência de aminoácidos de SEQ ID NO: 1 e tem uma atividade de xilitol desidrogenase; e

10 (c) um polipeptídeo que compreende uma sequência de aminoácidos que compreende substituição, deleção, inserção ou adição de um ou vários aminoácidos na sequência de aminoácidos de SEQ ID NO: 1 e tem uma atividade de xilitol desidrogenase.

2. DNA de acordo com a reivindicação 1, caracterizado pelo fato de que o DNA compreende a sequência de nucleotídeos de SEQ ID NO: 2.

15 3. DNA de acordo com a reivindicação 1, caracterizado pelo fato de que o DNA compreende uma sequência que difere da sequência de nucleotídeos de SEQ ID NO: 2 devido à degeneração do código genético.

4. DNA de acordo com qualquer uma das reivindicações 1 a 3, caracterizado pelo fato de que o DNA é de uma levedura.

20 5. DNA de acordo com a reivindicação 4, caracterizado pelo fato de que a levedura é *Candida shehatae*.

6. DNA de acordo com a reivindicação 5, caracterizado pelo fato de que a levedura é *Candida shehatae* CBS5813 (NBRC1983).

25 7. Construção de ácido nucléico, caracterizada pelo fato de que compreende um DNA como definido em qualquer uma das reivindicações 1 a 6, e uma sequência reguladora capaz de regular a expressão do DNA em uma célula hospedeira.

30 8. Construção de ácido nucléico de acordo com a reivindicação 7, caracterizada pelo fato de que o DNA codifica a sequência de aminoácidos de SEQ ID NO: 1.

9. Construção de ácido nucléico de acordo com a reivindicação 8, caracterizada pelo fato de que o DNA compreende a sequência de nucleo-

tídeos de SEQ ID NO: 2.

10. Construção de ácido nucléico de acordo com qualquer uma das reivindicações 7 a 9, caracterizada pelo fato de que a sequência reguladora é da célula hospedeira.

5 11. Construção de ácido nucléico de acordo com qualquer uma das reivindicações 7 a 9, caracterizada pelo fato de que a sequência reguladora é um promotor.

10 12. Construção de ácido nucléico de acordo com a reivindicação 11, caracterizada pelo fato de que o promotor é um promotor constitutivo ou um promotor induzível.

13. Construção de ácido nucléico de acordo com a reivindicação 11 ou 12, caracterizada pelo fato de que o promotor é selecionado no grupo que consiste nos promotores ADH1, ADH2, PDC, GAL1/10, TDH3, e PGK1.

15 14. Vetor, caracterizado pelo fato de que compreende um DNA como definido em qualquer uma das reivindicações 1 a 6, ou uma construção de ácido nucléico como definida em qualquer uma das reivindicações 7 a 13.

15. Vetor de acordo com a reivindicação 14, caracterizado pelo fato de que o vetor é um plasmídeo.

20 16. Micro-organismo, caracterizado pelo fato de que compreende um DNA como definido em qualquer uma das reivindicações 1 a 6, uma construção de ácido nucléico como definida em qualquer uma das reivindicações 7 a 13, ou um vetor como definido na reivindicação 14 ou 15, de tal modo que o micro-organismo expresse um polipeptídeo que tem uma atividade de xilitol desidrogenase.

17. Micro-organismo de acordo com a reivindicação 16, caracterizado pelo fato de que o micro-organismo é uma levedura ou bactéria.

30 18. Micro-organismo de acordo com a reivindicação 17, caracterizado pelo fato de que a levedura ou bactéria é selecionada dos grupos que consistem em leveduras do gênero *Saccharomyces*, *Schizosaccharomyces*, *Schwanniomyces*, *Kluyveromyces*, *Pichia*, *Hansenula*, *Candida*, *Debaryomyces*, *Metschnikowia*, *Pachulosen*, ou *Paecilomyces* e bactérias do gênero

*Zymomonas*.

19. Micro-organismo de acordo com a reivindicação 18, caracterizado pelo fato de que o micro-organismo é *Saccharomyces cerevisiae*.

5 20. Micro-organismo de acordo com a reivindicação 18, caracterizado pelo fato de que o micro-organismo é *Schizosaccharomyces pombe*.

21. Micro-organismo de acordo com qualquer uma das reivindicações 16 a 20, caracterizado pelo fato de que o DNA ou a construção de ácido nucléico é incorporada no genoma do micro-organismo hospedeiro.

10 22. Método para produzir xilitol desidrogenase, caracterizado pelo fato de que compreende as etapas de:

(a) cultivar um micro-organismo como definido em qualquer uma das reivindicações 16 a 21, em um meio; e

(b) coletar um produto que tem uma atividade de xilitol desidrogenase a partir do micro-organismo ou do meio.

15 23. Método de acordo com a reivindicação 22, caracterizado pelo fato de que compreende ainda a etapa de selecionar um micro-organismo apropriado para fermentação de xilulose.

20 24. Método para produzir etanol, caracterizado pelo fato de que usa um micro-organismo como definido em qualquer uma das reivindicações 16 a 21.

25. Vetor, caracterizado pelo fato de que compreende um cassete de expressão do gene de xilitol desidrogenase, que compreende um DNA de acordo com a reivindicação 1, um cassete de expressão do gene de xilulose redutase, e um cassete de expressão do gene de xilulocinase.

FIG.1-1

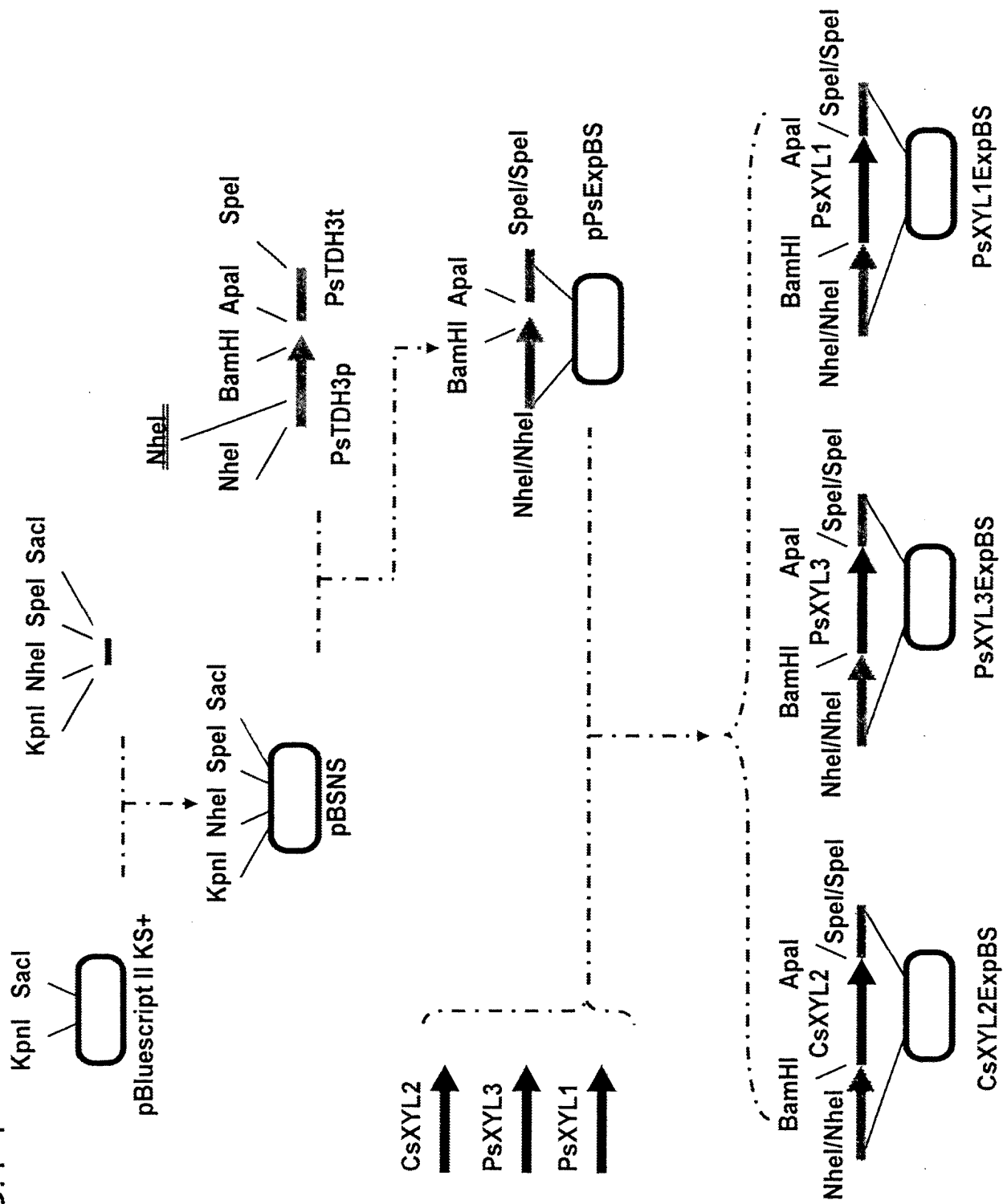


FIG.1-2

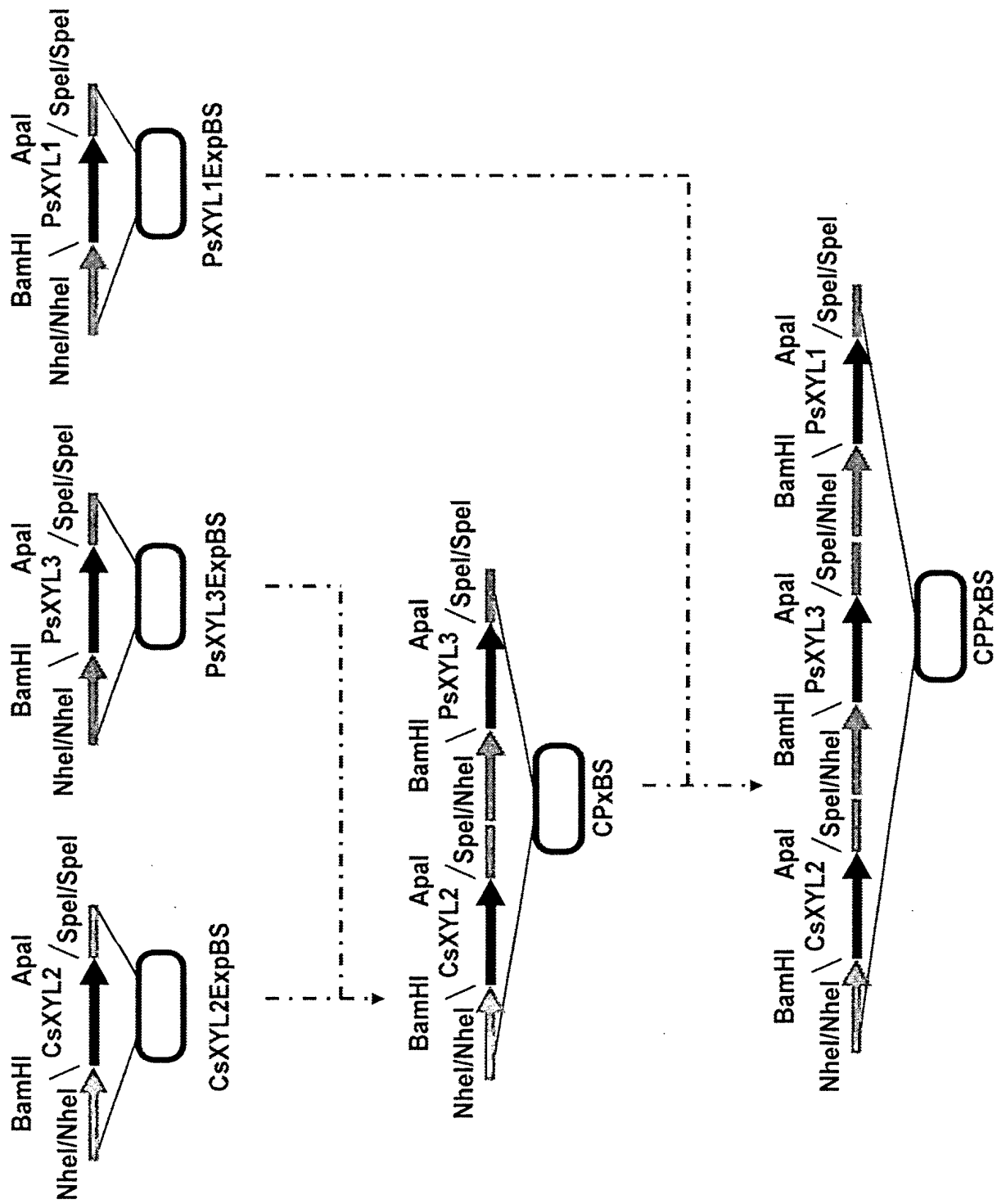
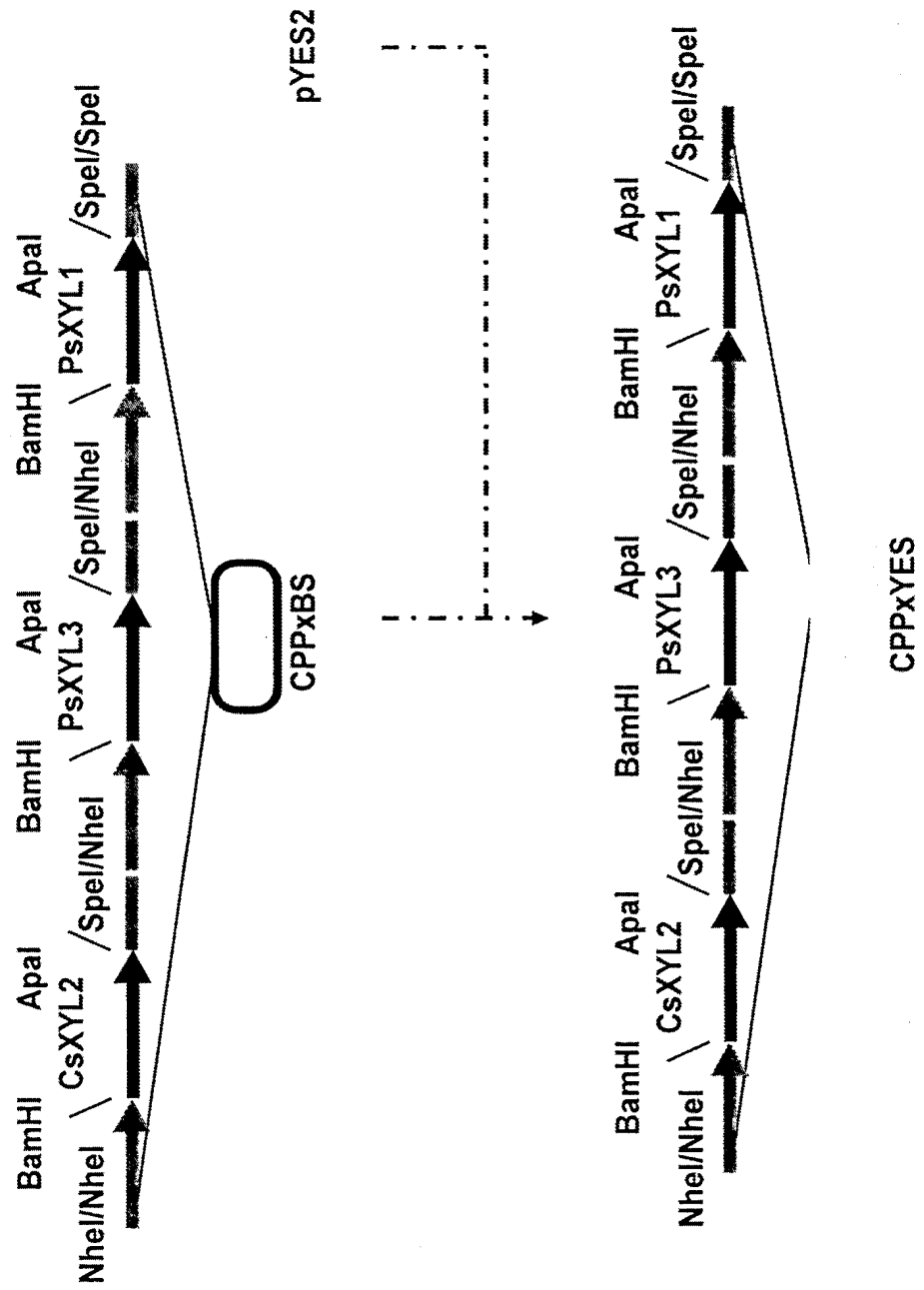


FIG.1-3



## RESUMO

Patente de Invenção: **"DNA QUE CODIFICA XILITOL DESIDROGENASE, CONSTRUÇÃO DE ÁCIDO NUCLEÍCO, VETORES, MICROORGANISMOS E MÉTODOS PARA PRODUZIR XILITOL DESIDROGENASE E ETANOL".**

5                   A presente invenção refere-se ao DNA que codifica xilitol desidrogenase inusitada e um método para usar o mesmo. Especificamente, a presente invenção compreende fornecer a sequência de ácidos nucleicos de um gene de xilitol desidrogenase a partir de *Candida shehatae* e introduzir este gene em um micro-organismo hospedeiro, produzindo desta forma um  
10 micro-organismo que tem a capacidade de utilizar xilose.