



- (51) **International Patent Classification:**
C12N 15/81 (2006.01) *C12N 1/16* (2006.01)
- (21) **International Application Number:**
PCT/US2016/040050
- (22) **International Filing Date:**
29 June 2016 (29.06.2016)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
14/796,125 10 July 2015 (10.07.2015) US
- (71) **Applicant:** THE UNITED STATES OF AMERICA, AS
REPRESENTED BY THE SECRETARY OF AGRICULTURE [US/US]; Washington, District of Columbia
20250 (US).
- (72) **Inventor:** HECTOR, Ronald E.; 1905 Stoneridge, Wash-
ington, Illinois 61571 (US).
- (74) **Agent:** GOLDBERG, Joshua B.; Nath, Goldberg & Mey-
er, 112 S. West Street, Alexandria, Virginia 22314 (US).
- (81) **Designated States** (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,

AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG,
MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM,
PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC,
SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) **Designated States** (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ,
TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU,
TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,
DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,
LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
— with sequence listing part of description (Rule 5.2(a))

(54) **Title:** SYNTHETIC PROMOTER FOR XYLOSE-REGULATED GENE EXPRESSION

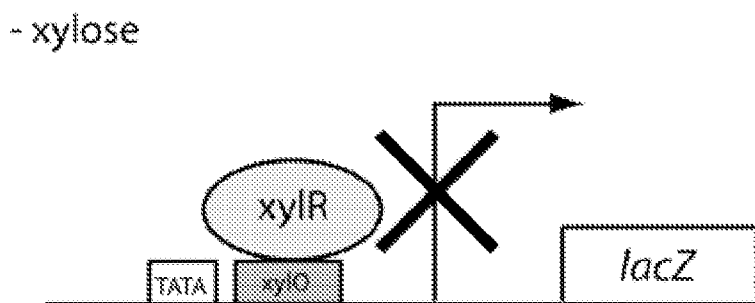


FIG. 1A

(57) **Abstract:** Disclosed are isolated nucleic acid molecules that have promoter activity specific to xylose. The synthetic promoters, SEQ ID NO: 2, SEQ ID NO: 3, and SEQ ID NO: 4, promote the expression of a coding region of interest in transformed yeast cells.

SYNTHETIC PROMOTER FOR XYLOSE-REGULATED GENE EXPRESSION

FIELD OF THE INVENTION

[0001] This invention relates to synthetic promoter regions derived from the *Ashbya gossypii* *TEF* promoter that are useful for gene expression in yeast. More specifically, the synthetic promoter controls gene expression in response to xylose availability.

BACKGROUND OF INVENTION

[0002] Xylose is the second most abundant sugar in nature (Saha BC: Hemicellulose bioconversion. J Ind Microbiol Biotechnol 2003, 30:279-291), however yeast strains such as *Saccharomyces* are unable to metabolize xylose. A technical challenge to enable and enhance yeast capability in utilization of pentose sugars such as xylose and arabinose harbored in biomass is to engineer yeast strains that can metabolize those sugars.

[0003] Genetic engineering efforts have been made to improve xylose utilization by overexpressing genes encoding pentose phosphate pathway (PPP) enzymes to enhance xylose flux into central carbon metabolism. For native *S. cerevisiae*, there are no xylose-specific transporters available and xylose uptake is via certain hexose transporters such as Hxt4, Hxt5, Hxt7, and Gal2. Recently, several heterologous sugar transporter genes possessing xylose transport functions have been expressed in recombinant *S. cerevisiae* such as *SUT1*, *XUT1* or *XUT3* from *S. stipitis*, *At5g59250* and *At5g17010* from *A. thaliana*, *An25* from *N. crassa*, *DEHA0D02167* and *XylHP* from *D. hansenii*, and symporters *GXS1* and *GXF1* genes from *C.*

intermedia. Improvement of xylose utilization by such efforts was observed but a satisfactory level has not been reached.

[0004] While there has been focus on engineering yeast strains that metabolize xylose as a sugar source, the environment yeast operate have both xylose and glucose as part of the batch. For *S. cerevisiae* strains, uncontrolled, high-level expression of many genes required for xylose fermentation can be detrimental to cell growth and fermentation (*Id.*) Additionally, certain genes required for efficient xylose fermentation negatively affect glucose fermentation (Meinander NQ, et al., Fermentation of xylose/glucose mixtures by metabolically engineered *Saccharomyces cerevisiae* strains expressing *XYL1* and *XYL2* from *Pichia stipitis* with and without overexpression of *TAL1*. Bioresource Technol 1999, 68:79-87).

[0005] Recombinant production of any heterologous protein is generally accomplished by constructing an expression cassette in which the DNA coding for the protein of interest is placed under the control of a promoter suitable for the host cell. The expression cassette is then introduced into the host cell (i.e., usually by plasmid-mediated transformation or targeted integration into the host genome) and production of the heterologous protein is achieved by culturing the transformed host cell under conditions necessary for the proper function of the promoter contained within the expression cassette. Thus, the development of new host cells (e.g., transformed yeast) for recombinant production of proteins generally requires the availability of promoters that are suitable for controlling the expression of a protein of interest in the host cell.

[0006] While there are promoters that have been isolated from yeast cells that are useful in heterologous gene expression in yeast, most of these promoters provide constitutive gene expression. There are fewer inducible promoters available from *S. cerevisiae* for regulating

gene expression and none of these are regulated by xylose. Thus, there is a need to develop a promoter that controls gene expression in response to xylose availability. Such an induced promoter would ensure efficient metabolism energy for the yeast cell without the cell growth and fermentation disadvantages stemming from unregulated, high-level expression, of multiple genes.

BRIEF SUMMARY OF THE INVENTION

[0007] Disclosed herein is an isolated nucleic acid molecule that has promoter activity specific to xylose and that comprises a DNA sequence selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4, wherein said isolated nucleic acid molecule is operatively linked to at least one heterologous nucleic acid sequence of interest. In one embodiment of the invention, a vector comprises the DNA sequence of SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4. In another embodiment of the invention, a cell comprises a vector having the DNA sequence of SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4. In yet another embodiment of the invention, the vector is stably integrated into the genome of the cell.

[0008] Disclosed herein is a method for expressing a coding region of interest in a transformed yeast cell comprising: a) providing a transformed yeast cell having a recombinant construct, wherein the recombinant construct comprises: (1) a promoter region comprising SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4, and (2) a coding region of interest which is expressible in the yeast cell; wherein the promoter region is operably linked to the coding region of interest; and b) growing the transformed yeast cell of step (a) under conditions whereby the recombinant construct of step (a) is expressed. In one embodiment of the method, the yeast cell is a member of a genus selected from the group consisting of *Saccharomyces*, *Kluyveromyces*,

Candida, Scheffersomyces, Spathaspora, Yarrowia, Schizosaccharomyces, Zygosaccharomyces, Brettanomyces, Debaryomyces, Schwanniomyces, Pachysolen, Torulaspora, Hansenula, Pichia.

In yet another embodiment of the method, the coding region of interest encodes a polypeptide, wherein the polypeptide is selected from the group consisting of: xylanases, xylose reductases, xylose dehydrogenases, xylitol dehydrogenases, xylulokinases, xylose transporters, glucose transporters, galactose transporters, myo-inositol transporters, xylose isomerases, transhydrogenases, NADH kinases, NADP-dependent d-glyceraldehyde-3-phosphate dehydrogenases, transketolases, transaldolases, glucose-6-phosphated dehydrogenases, ribulose-5-phosphate isomerase, ribulose-5-phosphate epimerases, phosphoglucose isomerases, alcohol dehydrogenases, aldehyde dehydrogenases, 2-pyrone synthases, beta-xylosidases, acetyl-CoA synthases, acetyl-CoA carboxylase, phosphoketolases, acetate kinases, transcription factors, and phosphotransacetylases.

[0009] Also disclosed is a transformed yeast comprising a promoter and a heterologous gene encoding a protein, wherein the promoter comprises a polynucleotide of SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4, wherein the polynucleotide is capable of activating expression of a gene in a yeast cell. In one embodiment of the invention, the transformed yeast having the promoter is operably linked to the nucleic acid sequence. In yet another embodiment of the invention, the transformed yeast is a *Saccharomyces cerevisiae*, *Kluyveromyces*, *Candida*, *Scheffersomyces*, *Spathaspora*, *Yarrowia*, *Schizosaccharomyces*, *Zygosaccharomyces*, *Brettanomyces*, *Debaryomyces*, *Schwanniomyces*, *Pachysolen*, *Torulaspora*, *Hansenula*, or *Pichia*.

BRIEF DESCRIPTION OF THE DRAWING

[0010] The present invention together with the disclosed embodiments may best be understood from the following detailed description of the drawings, wherein:

[0011] **FIG. 1A** depicts XylR binding to a xylO sequence and inhibiting transcription in the absence of xylose. **FIG. 1B** depicts xylose (•) is present and xylR dissociates from the xylO sequence allowing transcription. **FIG. 1C** depicts the location of xylO sequence with respect to the position of the TATA site and upstream activation sites (UAS_{rp1} and UAS_{rp2}) of the *Ashbya gossypii* *TEF* promoter. Base position numbers are relative to the start codon of the β -galactosidase gene (*lacZ*).

[0012] **FIG. 2** is a graph of β -galactosidase activity measured for strains containing all promoter variations shown in Figure 1. β -galactosidase activity is reported in Relative Light Units (RLUs) in millions. All variation shown were induced by xylose. The promoter variation xylO2-2, containing 2 repressor-binding sites 3' of the TATA element, yielded the greatest xylose induction.

[0013] **FIG. 3** is a graph comparing repressor variants and their β -galactosidase activity, measured in Relative Light Units (RLUs in millions). Addition of a nuclear localization signal (NLS) increased repression when xylose was not present. Additional fusion to a chromatin modifying protein, Ssn6p from *Saccharomyces cerevisiae*, resulted in a slight increase in repression.

[0014] **FIG. 4A** depicts the location of xylO sequence with respect to the position of the TATA site and upstream activation sites (UAS_{rp1} and UAS_{rp2}) of the *Ashpya gossypii* *TEF* promoter for variations of the promoter that were not inducible by xylose. Base position numbers are relative to the start codon. **FIG. 4B** is a graph of β -galactosidase (*lacZ*) activity

measured for strains containing all promoter variations shown in **FIG. 4 A**. β -galactosidase activity is reported in Relative Light Units (RLUs) in millions.

BRIEF DESCRIPTION OF THE SEQUENCES

[0015] SEQ ID NO: 1 is a synthetic promoter:

GAGCTCAAGCTTGCCTCGTCCCGCCGGGTCACCCGGCCAGCGACATGGAGGCCCAG
AATACCCTCCTTGACAGTCTTGACGTGCGCAGCTCAGGGGCATGATGTGACTGTCGC
CCGTACATTTAGCCCATACATCCCCATGTATAATCATTTCATCCATACATTTTGATG
GCCGCACGGCGCGAACGAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCAGGGA
AACGCTCCCCTCACAGACGCGTTGAATTGTCCCCACGCCGCGCCCCTGTAGAGAAAT
ATAAAAGGTTAGGATTTGCCACTGAGGTTCTTCTTTACATACTTCCTTTTAAATCT
TGCTAGGATACAGTTCTCACATCACATCCGAACATAAACAACAACTAGT.

[0016] SEQ ID NO: 2 is a synthetic promoter:

GAGCTCAAGCTTGCCTCGTCCCGCCGGGTCACCCGGCCAGCGACATGGAGGCCCAG
AATACCCTCCTTGACAGTCTTGACGTGCGCAGCTCAGGGGCATGATGTGACTGTCGC
CCGTACATTTAGCCCATACATCCCCATGTATAATCATTTCATCCATACATTTTGATG
GCCGCACGGCGCGAACGAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCAGGGA
AACGCTCCCCTCACAGACGCGTTGAATTGTCCCCACGCCGCGCCCCTGTAGAGAAAT
ATAAAACATGTTAGCGCTACCAAGTGGTTCTTCTTTACATACTTCCTTTTAAATCT
TGCTAGGATACAGTTCTCACATCACATCCGAACATAAACAACAACTAGT.

[0017] SEQ ID NO: 3 is a synthetic promoter:

GAGCTCAAGCTTGCCTCGTCCCGCCGGGTCACCCGGCCAGCGACATGGAGGCCCAG
AATACCCTCCTTGACAGTCTTGACGTGCGCAGCTCAGGGGCATGATGTGACTGTCGC
CCGTACATTTAGCCCATACATCCCCATGTATAATCATTTCATCCATACATTTTGATG

GCCGCACGGCGCGAACGAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCAGGGA
AACGCTCCCCTCACAGACGCGTTGAATTGTCCCACATGTTAGCGCTACCAAGTAAAT
ATAAAACATGTTAGCGCTACCAAGTGGTTCTTCTTTCACATACTTCCTTTTAAATCT
TGCTAGGATACAGTTCTCACATCACATCCGAACATAAACAAAACTAGT.

[0018] SEQ ID NO: 4 is a synthetic promoter:

GAGCTCAAGCTTGCCTCGTCCCGCCGGGTCACCCGGCCAGCGACATGGAGGCCAG
AATACCCTCCTTGACAGTCTTGACGTGCGCAGCTCAGGGGCATGATGTGACTGTCGC
CCGTACATTTAGCCCATACATCCCCATGTATAATCATTTCATCCATACATTTTGATG
GCCGCACGGCGCGAACGAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCAGGGA
AACGCTCCCCTCACAGACGCGTTGAATTGTCCCCACGCCGCGCCCCTGTAGAGAAAT
ATAAAACATGTTAGCGCTACCAAGTACATGTTAGCGCTACCAAGTCCTTTTAAATC
TTGCTAGGATACAGTTCTCACATCACATCCGAACATAAACAAAACTAGT.

[0019] SEQ ID NO: 5 is a nucleotide sequence of a nuclear localization signal:

ATGCCCAAGAAGAAAAGGAAAGTT.

[0020] SEQ ID NO: 6 is a gene that encodes a sequence-specific DNA binding protein that is optimized for expression in *Saccharomyces cerevisiae*:

ATGAATCAACCAGTAGAAAGACAGCGTAGGAGAACTACTCAAAGTGCTACAATTCG
TGACGTAGCTGCAAGAGCAGGTGTCTCTCCTATGACAGTCTCACGTGTAATCAATAG
AGAGTCCACAGTTAAAGAGGAACTAGACAGTTGGTTGAAAAGGCAATAGCCGACC
TTAACTATGCTCCTAATCCTGCAGCCAGATCTTTGGCAGGTAGTGCCCCTTTTAGAAT
TGGCTTACTGTACGATAATCCTTCAACTGGCTACCTTTCTGAATTTCTAGTTGGTGCC
TTAGATGAATCAAGTAGAACCGGTGCTCAAATTGTTATCGAGAAATGTGCTGAACCA
GAATTAGCCAGAGCTACACTTGCTAGATTGTTGAAAACCTGGAGTTGATGGACTTATC

TTACCTCCACCATTATGCGAATCTCCAGAAGTTCTGGCCGAGATAAGAGCCGCAGGA
GCTGCCGCTGTCGCAGTGGCACCTGGTACAGCTTCTGCCGACATGGCTACTATTAGA
ATCGACAACGAAGCAGCTGCATTTGAGTTGACCCAGCATTTGATTGGCTTGGGTAC
AAAAGATTCGGATTCATTAAGGGTCATCCAAATCAAACCGTGTCTCAACAAAGGCTT
GATGGGTTTATGACTGCTCTTAAGGCTGCAGGGATCCCACAAGAGAATATCAGAGT
GGAACAAGGTTACTTCACATATCGTTCAGGTCTAGAAGCTGCAGAGAGACTACTAG
CAGCCGAACCTAGGCCAACTGCCATCTTCGCTGCTAACGATGATATGGCAGCTGCAA
CAGCAGGCGTAGCACATAGACTAGGCTTGGATGTACCAGGCGACGTGTCTATAGTG
GGATTTGATGATACTTCCATAGCTGATAACATTTGGCCACCATTAAACAACAGTTCAC
CAACCAATTGCCGCTATGGCCAGAGCCGCTGTTGATCTGGTTCTAGAAGAGATCAGA
AGGCATAGAGATGGTGGTGGCGAACCTAGACAATTGATGCATCCACACACTCTGAT
CGTTAGAGACTCCTCAGGCCCTGCTGGAGTCTAA.

[0021] SEQ ID NO: 7 is a gene that encodes a sequence specific DNA binding protein that is optimized for expression in *Saccharomyces cerevisiae* and connected to a nuclear localization signal (SEQ ID NO 5):

ATGCCCAAGAAGAAAAGGAAAGTTAATCAACCAGTAGAAAGACAGCGTAGGAGAA
CTACTCAAAGTGCTACAATTCGTGACGTAGCTGCAAGAGCAGGTGTCTCTCCTATGA
CAGTCTCACGTGTAATCAATAGAGAGTCCACAGTTAAAGAGGAAACTAGACAGTTG
GTTGAAAAGGCAATAGCCGACCTTAACTATGCTCCTAATCCTGCAGCCAGATCTTTG
GCAGGTAGTGCCCCTTTTAGAATTGGCTTACTGTACGATAATCCTTCAACTGGCTAC
CTTTCTGAATTTCTAGTTGGTGCCTTAGATGAATCAAGTAGAACCGGTGCTCAAATT
GTTATCGAGAAATGTGCTGAACCAGAATTAGCCAGAGCTACACTTGCTAGATTGTTG
AAAAGTGGAGTTGATGGACTTATCTTACCTCCACCATTATGCGAATCTCCAGAAGTT

CTGGCCGAGATAAGAGCCGCAGGAGCTGCCGCTGTCGCAGTGGCACCTGGTACAGC
 TTCTGCCGACATGGCTACTATTAGAATCGACAACGAAGCAGCTGCATTTGAGTTGAC
 CCAGCATTTGATTGGCTTGGGTCACAAAAGATTCGGATTCATTAAGGGTCATCCAAA
 TCAAACCGTGTCTCAACAAAGGCTTGATGGGTTTATGACTGCTCTTAAGGCTGCAGG
 GATCCCACAAGAGAATATCAGAGTGGAACAAGGTTACTTCACATATCGTTCAGGTCT
 AGAAGCTGCAGAGAGACTACTAGCAGCCGAACCTAGGCCAACTGCCATCTTCGCTG
 CTAACGATGATATGGCAGCTGCAACAGCAGGCGTAGCACATAGACTAGGCTTGGAT
 GTACCAGGCGACGTGTCTATAGTGGGATTTGATGATACTTCCATAGCTGATAACATT
 TGGCCACCATTAACAACAGTTCACCAACCAATTGCCGCTATGGCCAGAGCCGCTGTT
 GATCTGGTTCTAGAAGAGATCAGAAGGCATAGAGATGGTGGTGGCGAACCTAGACA
 ATTGATGCATCCACACACTCTGATCGTTAGAGACTCCTCAGGCCCTGCTGGAGTCTA
 A.

[0022] SEQ ID NO: 8 is a gene that encodes a sequence specific DNA binding protein that is optimized for expression in *Saccharomyces cerevisiae* and connected to a nuclear localization signal (SEQ ID NO 5) and the *S. cerevisiae* *SSN6* gene:

ATGCCCAAGAAGAAAAGGAAAGTTAATCAACCAGTAGAAAGACAGCGTAGGAGAA
 CTA CTCAAAGTGCTACAATTCGTGACGTAGCTGCAAGAGCAGGTGTCTCTCCTATGA
 CAGTCTCACGTGTAATCAATAGAGAGTCCACAGTTAAAGAGGAAACTAGACAGTTG
 GTTGAAAAGGCAATAGCCGACCTTAACTATGCTCCTAATCCTGCAGCCAGATCTTTG
 GCAGGTAGTGCCCCTTTTAGAATTGGCTTACTGTACGATAATCCTTCAACTGGCTAC
 CTTTCTGAATTTCTAGTTGGTGCCTTAGATGAATCAAGTAGAACCGGTGCTCAAATT
 GTTATCGAGAAATGTGCTGAACCAGAATTAGCCAGAGCTACACTTGCTAGATTGTTG
 AAAACTGGAGTTGATGGACTTATCTTACCTCCACCATTATGCGAATCTCCAGAAGTT

CTGGCCGAGATAAGAGCCGCAGGAGCTGCCGCTGTCGCAGTGGCACCTGGTACAGC
TTCTGCCGACATGGCTACTATTAGAATCGACAACGAAGCAGCTGCATTTGAGTTGAC
CCAGCATTTGATTGGCTTGGGTCACAAAAGATTCGGATTCATTAAGGGTCATCCAAA
TCAAACCGTGTCTCAACAAAGGCTTGATGGGTTTATGACTGCTCTTAAGGCTGCAGG
GATCCCACAAGAGAATATCAGAGTGGAACAAGGTTACTTCACATATCGTTCAGGTCT
AGAAGCTGCAGAGAGACTACTAGCAGCCGAACCTAGGCCAACTGCCATCTTCGCTG
CTAACGATGATATGGCAGCTGCAACAGCAGGCGTAGCACATAGACTAGGCTTGGAT
GTACCAGGCGACGTGTCTATAGTGGGATTTGATGATACTTCCATAGCTGATAACATT
TGGCCACCATTAACAACAGTTCACCAACCAATTGCCGCTATGGCCAGAGCCGCTGTT
GATCTGGTTCTAGAAGAGATCAGAAGGCATAGAGATGGTGGTGGCGAACCTAGACA
ATTGATGCATCCACACACTCTGATCGTTAGAGACTCCTCAGGCCCTGCTGGAGTCGG
TTCCGGAGGTGGAGGTTCTATGAATCCGGGCGGTGAACAAACAATAATGGAACAAC
CCGCTCAACAGCAACAACAACAGCAACAACAACAGCAGCAACAGCAACAGCAGGC
AGCAGTTCCTCAGCAGCCACTCGACCCATTAACACAATCAACTGCGGAAACTTGGCT
CTCCATTGCTTCTTTGGCAGAAACCCTTGGTGATGGCGACAGGGCCGCAATGGCATA
TGACGCCACTTTACAGTTCAATCCCTCATCTGCAAAGGCTTTAACATCTTTGGCTCAC
TTGTACCGTTCCAGAGACATGTTCCAAAGAGCTGCAGAATTATATGAAAGAGCACTT
TTGGTAAATCCCGAACTATCAGATGTGTGGGCTACTTTAGGTCATTGTTATCTGATGC
TGGATGATCTGCAAAGAGCTTACAATGCCTATCAACAGGCTCTCTACCACCTCAGTA
ATCCCAACGTACCGAAATTATGGCATGGAATCGGCATTCTTTATGACAGATATGGTT
CGCTCGACTATGCCGAAGAAGCTTTTGCCAAAGTTTTGGAATTGGACCCTCATTTTG
AAAAGGCCAAACGAAATTTACTTCAGACTAGGTATTATTTATAAACATCAGGGTAAAT
GGTCTCAAGCTTTGGAATGCTTCAGATACATTCTCCCTCAACCTCCTGCTCCCTTGCA

GGAGTGGGACATATGGTTTCAGTTGGGTAGTGTTTTGGAGAGTATGGGAGAGTGGC
AAGGTGCGAAGGAAGCCTACGAGCATGTCTTGGCTCAAAATCAACATCATGCCAAA
GTATTACAACAATTAGGTTGTCTTTACGGTATGAGTAACGTACAATTTTATGACCCTC
AAAAGGCATTGGATTATCTTCTAAAGTCGTTAGAAGCAGATCCCTCCGATGCCACTA
CATGGTACCATCTCGGTAGAGTGCATATGATTAGAACAGATTATACTGCCGCATATG
ATGCTTTCCAACAAGCTGTTAATAGAGATTCAAGAAACCCTATCTTTTGGTGCTCAA
TCGGTGTTTTATATTACCAAATTTCTCAATACAGAGACGCCTTAGACGCGTACACAA
GAGCCATAAGATTAAATCCTTATATTAGTGAAGTTTGGTACGATCTAGGTACTCTTT
ACGAAACTTGTAACAACCAATTATCTGACGCCCTTGATGCGTATAAGCAAGCTGCAA
GACTGGACGTAAATAATGTTACATAAGAGAAAGATTAGAAGCTTTAACAAAGCAG
TTAGAAAACCCAGGCAATATAAACAAATCGAACGGTGCGCCAACGAATGCCTCTCC
TGCCCCACCTCCTGTGATTTTACAACCTACCTTACAACCTAATGATCAAGGAAATCC
TTTGAACACTAGAAATTTAGCCCAATCTGCCAATGCTACTGCTTCAATGGTACAACA
ACAGCATCCTGCTCAACAAACGCCTATTA ACTCTTCTGCAACAATGTACAGTAATGG
AGCTTCCCCTCAATTACAAGCTCAAGCTCAAGCTCAAGCTCAAGCACAAGCTCAAGC
ACAAGCACAAGCTCAAGCACAAGCACAAGCACAAGCGCAAGCACAAGCACAAGCA
CAGGCGCAAGCACAGGCACAAGCACAAGCACAAGCACATGCACAAGCGCAAGCAC
AAGCACAAGCACAGGCACAAGCACAAGCACAGGCGCAGGCACAACAACAACAACA
ACAACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA
CAACAACAACAACAGCAGCAGCAATTACAGCCCCTACCAAGACAACAGCTGCAGCA
AAAGGGAGTTTCTGTGCAAATGTTAAATCCTCAACAAGGGCAACCATATATCACAC
AGCCAACAGTCATACAAGCTCACCAACTGCAACCATTTTCTACACAAGCTATGGAAC
ATCCGCAAAGCTCTCAACTGCCACCTCAACAGCAACA ACTACAATCTGTTCAACATC

CACAACAACCTTCAAGGCCAGCCTCAAGCCCAAGCTCCCCAACCTTTAATCCAGCATA
ACGTGGAACAGAACGTTTTACCTCAAAAGAGATACATGGAAGGTGCAATCCACACT
TTAGTAGATGCCGCCGTATCCAGTAGCACCCACACAGAGAATAACACAAAGTCTCC
TCGTCAACCAACCCATGCCATTCCAACGCAAGCTCCCGCAACAGGAATAACGAACG
CTGAACCACAGGTAAAGAAGCAAAAGTTGAACTCTCCAAATTCAAACATCAACAAA
TTAGTAAATACTGCTACTTCCATTGAAGAAAATGCAAAATCTGAGGTGAGCAACCA
ATCGCCAGCAGTAGTGGAGTCTAATACCAATAATACTTCACAAGAAGAAAAACCTG
TAAAAGCAAACCTCAATACCTTCAGTAATTGGCGCACAGGAACCTCCACAGGAAGCT
AGTCCTGCTGAAGAAGCTACCAAAGCAGCTTCTGTTTCTCCTTCTACAAAACCGCTT
AATACGGAACCAGAGTCATCTAGTGTCCAACCAACTGTATCATCAGAAAGTTCAAC
AACAAAAGCAAATGACCAAAGCACTGCTGAGACCATAGAACTTTCTACTGCTACTG
TTCCTGCAGAAGCAAGCCCTGTAGAAGACGAAGTAAGACAGCATTCTAAAGAGGAA
AACGGCACAACCTGAAGCATCTGCACCTTCTACTGAAGAGGCGGAGCCAGCAGCTTC
CAGAGATGCTGAAAAACAACAAGATGAAACCGCTGCTACAACGATAACTGTAATCA
AACCTACTTTGGAAACAATGGAAACAGTGAAAGAGGAGGCCAAAATGCGTGAGGA
AGAGCAAACATCTCAAGAAAAATCCCCACAGGAGAACACACTTCCAAGAGAAAAAT
GTAGTAAGGCAAGTGGAAGAAGATGAAAACCTACGACGACTAA.

[0023] SEQ ID NO: 9 is a synthetic promoter:

GAGCTCAAGCTTGCCTCGTCCCGCCGGGTCACCCGGCCAGCGACATGGAGGCCAG
AATACCCTCCTTGACAGTCTTGACGTGCGCAGCTCAGGGGCATGATGTGACTGTTCG
CCGTACATTTAGCCCATACATCACATGTTAGCGCTACCAAGTTGCATCCATACATTTT
GATGGCCGCACGGCGCGAACGAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCA
GGGAAACGCTCCCCTCACAGACGCGTTGAATTGTCCCCACGCCGCGCCCCTGTAGAG

AAATATAAAAGGTTAGGATTTGCCACTGAGGTTCTTCTTTTCACATACTTCCTTTTAAA
ATCTTGCTAGGATACAGTTCTCACATCACATCCGAACATAAAACAAAACTAGT.

[0024] SEQ ID NO: 10 is a synthetic promoter:

GAGCTCAAGCTTGCCTCGTCCCGCCGGGTCACCCGGCCAGCGACATGGAGGCCCAG
AATACCCTCCTTGACAGTCTTGACGTGCGCAGCTCAGGGGCATGATGTGACTGTTCGC
CCGTACATTTAGCCCATACATCACATGTTAGCGCTACCAAGTTGCATCCATACATTTT
ACATGTTAGCGCTACCAAGTGAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCA
GGGAAACGCTCCCCTCACAGACGCGTTGAATTGTCCCCACGCCGCGCCCCTGTAGAG
AAATATAAAAGGTTAGGATTTGCCACTGAGGTTCTTCTTTTCACATACTTCCTTTTAAA
ATCTTGCTAGGATACAGTTCTCACATCACATCCGAACATAAAACAAAACTAGT.

[0025] SEQ ID NO: 11 is a synthetic promoter:

GAGCTCAAGCTTGCCTCGTCCCGCCGGGTCACCCGGCCAGCGACATGGAGGCCCAG
AATACCCTCCTTGACAGTCTTGACGTGCGCAGCTCAGGGGCATGATGTGACTGTTCGC
CCGTACATTTAGCCCATACATCACATGTTAGCGCTACCAAGTTGCATCCATACATTTT
ACATGTTAGCGCTACCAAGTGAAAAATTACGGCTCCTCGCTGCAGACCTGCGAGCA
GGGAAACGCTCCCCTCACAGACGCGTTGAATTGTCCCCACGCCGCGCCCCTGTAGAG
AAATATAAAACATGTTAGCGCTACCAAGTACATGTTAGCGCTACCAAGTCCTTTTAA
AATCTTGCTAGGATACAGTTCTCACATCACATCCGAACATAAAACAAAACTAGT.

[0026] SEQ ID NO: 12 is a synthetic sequence: ACATGTTAGCGCTACCAAGT.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

[0027] As used in the specification and claims, the singular form “a”, “an” and “the” include plural references unless the context clearly dictates otherwise. For example, the term “a cell” includes a plurality of cells, including mixtures thereof.

[0028] As used herein, the term “cloning” refers to the selection and propagation of (a) genetic material from a single individual, (b) a vector containing one gene or gene fragment, or (c) a single organism containing one such gene or gene fragment.

[0029] As used herein, the term “vector” or “plasmid” each refer to a non-chromosomal (episomal) double-stranded DNA sequence comprising an intact "replicon" such that the vector or plasmid is replicated in a host cell. When the plasmid is placed within a unicellular organism, the characteristics of that organism may be changed or transformed as a result of the DNA of the plasmid. A cell transformed by a plasmid is called a "transformant".

[0030] As used herein, the term “cloning vector” refers to a plasmid, virus, retrovirus, bacteriophage, cosmid, artificial chromosome (bacterial or yeast), or nucleic acid sequence which is able to replicate in a host cell, characterized by one or a small number of restriction endonuclease recognition sites at which the sequence may be cut in a predetermined fashion, and which may contain an optional marker suitable for use in the identification of transformed cells, *e.g.*, tetracycline resistance or ampicillin resistance. A cloning vector may or may not possess the features necessary for it to operate as an expression vector.

[0031] As used herein, the term “codon” refers to a DNA sequence of three nucleotides (a triplet) which codes (through mRNA) for an amino acid, a translational start signal, or a translational termination signal. For example, the nucleotide triplets TTA, TTG, CTT, CTC, CTA, and CTG encode for the amino acid leucine, while TAG, TAA, and TGA are translational stop signals, and ATG is a translational start signal.

[0032] As used herein, the term “DNA coding sequence” refers to a DNA sequence which is transcribed and translated into a polypeptide in vivo when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, but is not limited to, prokaryotic sequences and cDNA from eukaryotic mRNA. A polyadenylation signal and transcription termination sequence will usually be located 3' to the coding sequence.

[0033] As used herein, the term “DNA construct” refers to an artificially constructed (*i.e.*, non-naturally occurring) DNA molecules useful for introducing DNA into host cells, including chimeric genes, expression cassettes, and vectors.

[0034] As used herein, the term “DNA sequence” refers a linear series of nucleotides connected one to the other by phosphodiester bonds between the 3' and 5' carbons of adjacent pentoses.

[0035] As used herein, the term “expression” refers to the process undergone by a structural gene to produce a polypeptide. Expression requires transcription of DNA, post-transcriptional modification of the initial RNA transcript, and translation of RNA.

[0036] As used herein, the term “expression cassette” refers to a chimeric nucleic acid construct, typically generated recombinantly or synthetically, which comprises a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a host cell. In an exemplary embodiment, an expression cassette comprises a heterologous nucleic acid to be transcribed, operably linked to a promoter. Typically, an expression cassette is part of an expression vector.

[0037] As used herein, the term “operably linked”, “operably encodes”, or “operably associated” each refer to the functional linkage between a promoter and nucleic acid sequence, wherein the promoter initiates transcription of RNA corresponding to the DNA sequence. A heterologous DNA sequence is "operatively associated" with the promoter in a cell when RNA polymerase which binds the promoter sequence transcribes the coding sequence into mRNA which is then in turn translated into the protein encoded by the coding sequence.

[0038] As used herein, the term “expression control sequence” refers to expression control sequences that are DNA sequences involved in any way in the control of transcription or translation and must include a promoter. Suitable expression control sequences and methods of making and using them are well known in the art.

[0039] As used herein, the term “expression vector” refers a nucleic acid which comprises an expression cassette and which is capable of replicating in a selected host cell or organism. An expression vector may be a plasmid, virus, retrovirus, bacteriophage, cosmid, artificial chromosome (bacterial or yeast), or nucleic acid sequence which is able to replicate in a host cell, characterized by a restriction endonuclease recognition site at which the sequence may be cut in a predetermined fashion for the insertion of a heterologous DNA sequence. An expression vector may include the promoter positioned upstream of the site at which the sequence is cut for the insertion of the heterologous DNA sequence, the recognition site being selected so that the promoter will be operatively associated with the heterologous DNA sequence. A heterologous DNA sequence is "operatively associated" with the promoter in a cell when RNA polymerase which binds the promoter sequence transcribes the coding sequence into mRNA which is then in turn translated into the protein encoded by the coding sequence.

[0040] As used herein, the term “gene” refers to a segment of DNA which encodes a specific protein or polypeptide, or RNA.

[0041] As used herein, the term “genome” refers to the entire DNA of an organism. It includes, among other things, the structural genes encoding for the polypeptides of the substance, as well as operator, promoter and ribosome binding and interaction sequences.

[0042] As used herein, the term “heterologous DNA” refers to a DNA sequence inserted within or connected to another DNA sequence which codes for polypeptides not coded for in nature by the DNA sequence to which it is joined. Allelic variations or naturally occurring mutational events do not give rise to a heterologous DNA sequence as defined herein.

[0043] As used herein, the term “hybridization” refers to the pairing together or annealing of single stranded regions of nucleic acids to form double-stranded molecules.

[0044] As used herein, the term “nucleotide” refers to a monomeric unit of DNA or RNA consisting of a sugar moiety (pentose), a phosphate, and a nitrogenous heterocyclic base. The base is linked to the sugar moiety via the glycosidic carbon (1' carbon of the pentose) and that combination of base and sugar is a nucleoside. The base characterizes the nucleotide. The four DNA bases are adenine ("A"), guanine ("G"), cytosine ("C"), and thymine ("T"). The four RNA bases are A, G, C, and uracil ("U").

[0045] As used herein, the term “promoter” refers to a DNA sequence within a larger DNA sequence defining a site to which RNA polymerase may bind and initiate transcription. A promoter may include optional distal enhancer or repressor elements. The promoter may be either homologous, *i.e.*, occurring naturally to direct the expression of the desired nucleic acid, or heterologous, *i.e.*, occurring naturally to direct the expression of a nucleic acid derived from a gene other than the desired nucleic acid. A promoter may be constitutive or inducible. A

constitutive promoter is a promoter that is active under most environmental and developmental conditions. An inducible promoter is a promoter that is active under environmental or developmental regulation, *e.g.*, upregulation in response to xylose availability. Promoters may be derived in their entirety from a native gene, may comprise a segment or fragment of a native gene, or may be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental or physiological conditions. It is further understood that the same promoter may be differentially expressed in different tissues and/or differentially expressed under different conditions.

[0046] As used herein, the term “promoter activity” will refer to an assessment of the transcriptional efficiency of a promoter. This may, for instance, be determined directly by measurement of the amount of mRNA transcription from the promoter (*e.g.*, by quantitative PCR or Northern blotting or primer extension methods) or indirectly by measuring the amount of gene product expressed from the promoter.

[0047] As used herein, the term “yeast” refers to a phylogenetically diverse grouping of single-celled fungi. Yeast do not form a specific taxonomic or phylogenetic grouping, but instead comprise a diverse assemblage of unicellular organisms that occur in the Ascomycotina and Basidiomycotina. Collectively, about 100 genera of yeast have been identified, comprising approximately 1,500 species (Kurtzman and Fell, *Yeast Systematics And Phylogeny: Implications Of Molecular Identification Methods For Studies In Ecology*. In C. A. Rosa and G. Peter, eds., *The Yeast Handbook*. Germany: Springer-Verlag Berlin Herdelberg, 2006). Yeast reproduce principally by budding (or fission) and derive energy from fermentation, via conversion of

carbohydrates to ethanol and carbon dioxide. Examples of some yeast genera include, but are not limited to: *Saccharomyces*, *Kluyveromyces*, *Candida*, *Scheffersomyces*, *Spathaspora*, *Yarrowia*, *Schizosaccharomyces*, *Zygosaccharomyces*, *Brettanomyces*, *Debaryomyces*, *Schwanniomyces*, *Pachysolen*, *Torulaspora*, *Hansenula*, *Pichia*.

[0048] The disclosure herein teaches partial or complete nucleotide sequences containing one or more particular yeast promoters. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above, are encompassed in the present disclosure.

[0049] Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook, J., Fritsch, E. F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Cold Spring Harbor, N.Y. (1989); by Silhavy, T. J., Bennis, M. L. and Enquist, L. W., *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory: Cold Spring Harbor, N.Y. (1984); and by Ausubel, F. M. et al., *Current Protocols in Molecular Biology*, published by Greene Publishing Assoc. and Wiley-Interscience, Hoboken, N.J. (1987).

Materials Used

[0050] *Escherichia coli* strains DH10B, TOP10 (Invitrogen; Carlsbad, CA, USA), NEB10 β (NEB; Beverly, MA, USA) were used for routine maintenance and preparation of plasmids and were grown in LB medium (Sambrook and Russell, 2001, *Molecular Cloning: A Laboratory Manual*. 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press). Plasmids and strains are listed in Table 1 and Table 2. Plasmid DNA was transformed into yeast cells using a

standard lithium acetate method (Gietz and Woods, 2002, Transformation of yeasts by the lithium acetate/single-stranded carrier/polyethylene glycol method. *Methods Enzymol*, 350:87-96).

Synthetic medium consisted of 6.7 g/L Difco yeast nitrogen base (YNB) (United States Biological; Marblehead, MA, USA), and was supplemented with amino acids (Amberg, Burke et al., 2005, *Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual*. 2005 edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press). For maintenance of plasmids, media was made without tryptophan or leucine as necessary for plasmid maintenance. Synthetic medium was filter sterilized. Sterile carbon sources were added separately.

Table 1. Plasmids used

Plasmid	Description	
pRS414	pBluescript II SK+, <i>TRP1</i> , <i>CEN6</i> , <i>ARSH4</i>	Christianson TW, et al., Multifunctional yeast high-copy-number shuttle vectors. <i>Gene</i> 1992, 110:119-122.
pRS415	pBluescript II SK+, <i>LEU2</i> , <i>CEN6</i> , <i>ARSH4</i>	<i>Gene</i> 1992, 110:119-122.
pUC57	Gene synthesis vector	(GenScript)
pJ201	Cloning vector	(DNA2.0)
pRH164	pRS414 + P _{HXT7} -MCS-T _{HXT7}	Hector RE, et al., Engineering industrial <i>Saccharomyces cerevisiae</i> strains for xylose fermentation and comparison for switchgrass conversion. <i>J Ind Microbiol Biotechnol</i> 2011, 38:1193-1202.
pRH457	pJ201 + <i>C. crescentus</i> xylR ^a	(DNA2.0)
pRH463	pRH164 + <i>C. crescentus</i> xylR	This work
pRH467	pRH164 + <i>C. crescentus</i> xylR with N-terminal NLS added	This work
pRH482	pCR2.1 + <i>C. crescentus</i> NLS-xylR-linker-SSN6	This work
pRH483	pRH164 + <i>C. crescentus</i> NLS-xylR-linker-SSN6	This work
pRH499	pUC57 + P _{TEF}	(GenScript)
pRH500	pUC57 + P _{TEF-xylO1}	(GenScript)
pRH501	pUC57 + P _{TEF-xylO2-1}	(GenScript)
pRH502	pUC57 + P _{TEF-xylO2-2}	(GenScript)
pRH503	pRS415 + P _{TEF} - MCS - T _{ADH1}	This work
pRH504	pRS415 + P _{TEF-xylO1} - MCS - T _{ADH1}	This work

pRH505	pRS415 + P _{TEF-xyIO 2-1} – MCS – T _{ADH1}	This work
pRH506	pRS415 + P _{TEF-xyIO 2-2} – MCS – T _{ADH1}	This work
pRH511	pRS415 + P _{TEF} – lacZ – T _{ADH1}	This work
pRH512	pRS415 + P _{TEF-xyIO 1} – lacZ – T _{ADH1}	This work
pRH513	pRS415 + P _{TEF-xyIO 2-1} – lacZ – T _{ADH1}	This work
pRH514	pRS415 + P _{TEF-xyIO 2-2} – lacZ – T _{ADH1}	This work
pRH531	pUC57 + P _{TEF-UAS-xyIO1}	(GenScript)
pRH532	pUC57 + P _{TEF-UAS-xyIO2}	(GenScript)
pRH534	pUC57 + P _{TEF-xyIO4}	(GenScript)
pRH546	pRS415 + P _{TEF-UAS-xyIO 1} – lacZ – T _{ADH1}	This work
pRH547	pRS415 + P _{TEF-UAS-xyIO 2} – lacZ – T _{ADH1}	This work
pRH549	pRS415 + P _{TEF-xyIO 4} – lacZ – T _{ADH1}	This work

^a The *C. crescentus* xylR gene used throughout this work was codon-optimized for expression in *S. cerevisiae*

Table 2. Microorganisms used

Strain	Genotype (description)	Reference
CEN.PK2-1C	<i>S. cerevisiae</i> MATa ura3-52 trp1-289 leu2-3,112 his3Δ1 MAL2-8 ^C SUC2	Euroscarf
YRH1054	CEN.PK2-1C [pRH511 (P _{TEF} – lacZ – T _{ADH1}) + pRS414]	This work
YRH1055	CEN.PK2-1C [pRH511 (P _{TEF} – lacZ – T _{ADH1}) + pRH483 (P _{HXT7} –NLS-xylR-SSN6)]	This work
YRH1056	CEN.PK2-1C [pRH512 (P _{TEF-xyIO1} – lacZ – T _{ADH1}) + pRS414]	This work
YRH1057	CEN.PK2-1C [pRH512 (P _{TEF-xyIO 1} – lacZ – T _{ADH1}) + pRH483 (P _{HXT7} –NLS-xylR-SSN6)]	This work
YRH1058	CEN.PK2-1C [pRH513 (P _{TEF-xyIO 2-1} – lacZ – T _{ADH1}) + pRS414]	This work
YRH1059	CEN.PK2-1C [pRH513 (P _{TEF-xyIO 2-1} – lacZ – T _{ADH1}) + pRH483 (P _{HXT7} –NLS-xylR-SSN6)]	This work
YRH1060	CEN.PK2-1C [pRH514 (P _{TEF-xyIO 2-2} – lacZ – T _{ADH1}) + pRS414]	This work
YRH1061	CEN.PK2-1C [pRH514 (P _{TEF-xyIO 2-2} – lacZ – T _{ADH1}) + pRH483 (P _{HXT7} –NLS-xylR-SSN6)]	This work
YRH1156	CEN.PK2-1C [pRH546 (P _{TEF-UAS-xyIO 1} – lacZ – T _{ADH1}) + pRS414]	This work
YRH1157	CEN.PK2-1C [pRH546 (P _{TEF-UAS-xyIO 1} – lacZ – T _{ADH1}) + pRH483 (P _{HXT7} –NLS-xylR-SSN6)]	This work
YRH1158	CEN.PK2-1C [pRH547 (P _{TEF-UAS-xyIO 2} – lacZ – T _{ADH1}) + pRS414]	This work
YRH1159	CEN.PK2-1C [pRH547 (P _{TEF-UAS-xyIO 2} – lacZ – T _{ADH1}) + pRH483 (P _{HXT7} –NLS-xylR-SSN6)]	This work
YRH1162	CEN.PK2-1C [pRH549 (P _{TEF-xyIO 4} – lacZ – T _{ADH1}) + pRS414]	This work
YRH1163	CEN.PK2-1C [pRH549 (P _{TEF-xyIO 4} – lacZ – T _{ADH1}) + pRH483 (P _{HXT7} –NLS-xylR-SSN6)]	This work
YRH1227	CEN.PK2-1C [pRH514 (P _{TEF-xyIO 2-2} – lacZ – T _{ADH1}) + pRH467 (P _{HXT7} – NLS – xylR)]	This work
YRH1276	CEN.PK2-1C [pRH514 (P _{TEF-xyIO 2-2} – lacZ – T _{ADH1}) + pRH463 (P _{HXT7} – xylR)]	This work

Example 1: Constructing Xylose Regulated Promoters

[0051] The constitutive promoter for the translation elongation factor 1α gene (*TEF*) from *Ashbya gossypii* (referred to as the *AgTEF*, or *TEF*, promoter) was modified to include DNA sequences that bind to a sequence-specific DNA-binding protein from *Caulobacter crescentus* (*xylR*). Furthermore, the nucleotide sequence of the *AgTEF* promoter was modified to 1)

remove sequences that direct cleavage of the DNA by the restriction endonucleases EcoRI and BssHII, and 2) remove an alternative TATA sequence to disable aberrant transcription initiation under conditions that inhibit transcription initiation from the main TATA sequence located at -108. To accomplish these goals the following nucleotides were changed: (1) T(-69)→C, (2) G(-127)→C, and C(-136)→G. A nucleotide sequence with these changes (SEQ ID NO: 1) was synthesized (GenScript USA, Piscataway, NJ).

[0052] To generate synthetic promoters capable of binding the xylR protein, 20 nucleotide segments located around the TATA sequence or upstream activation sequence elements were strategically replaced (FIG. 1 and FIG. 4) with the 20 nucleotide sequence (SEQ. ID. NO: 12 ACATGTTAGCGCTACCAAGT) that has been demonstrated to interact with xylR in a xylose-dependent manner (Stephens C., et al., 2007, Regulation of D-xylose metabolism in *Caulobacter crescentus* by a LacI-type repressor. *J Bacteriol*, 189:8828-8834). Promoter variation P_{TEF-xylO1} replaced nucleotides -83 to -102 (SEQ ID NO: 2). Promoter variation P_{TEF-xylO2-1} replaced nucleotides -83 to -102 and -112 to -131 (SEQ ID NO: 3). Promoter variation P_{TEF-xylO2-2} replaced nucleotides -63 to -82 and -83 to -102 (SEQ ID NO: 4). Promoter variation P_{TEF-UAS-xylO1} replaced nucleotides -241 to -256 (SEQ ID NO: 9). Promoter variation P_{TEF-UAS-xylO2} replaced nucleotides -215 to -234 and -241 to -256 (SEQ ID NO: 10). Promoter variation P_{TEF-xylO4} replaced nucleotides -63 to -82 and -83 to -102, -215 to -234, and -241 to -256 (SEQ ID NO: 11). These nucleotide sequences were synthesized (GenScript USA, Piscataway, NJ). The nucleotide sequences were sub-cloned using restriction endonucleases SacI and SpeI and ligated into a vector that is able to replicate in *S. cerevisiae*. This vector also contained nucleotide sequences for additional restriction endonucleases for ease of adding heterologous genes that are operably linked to the promoter sequence. The vector also contained nucleotide sequence from the *S.*

cerevisiae ADHI gene 3' untranslated region (3' UTR) to direct the 3'-end cleavage and polyadenylation of the transcribed RNA that is necessary to generate a stable mRNA capable of efficient nucleocytoplasmic transport.

[0053] In the absence of expression of the xylR repressor protein, each of the synthetic promoters provides constitutive expression (**FIG. 2** and **FIG. 4**). Regulation by xylose is achieved by expression of the xylR protein. The gene for xylR from *C. crescentus* was codon-optimized for enhanced expression in *S. cerevisiae* and synthesized (DNA2.0, Menlo Park, CA)(SEQ ID NO: 6). This gene was operably linked to a constitutive promoter for expression in *S. cerevisiae*. Since bacterial proteins do not have to cross a nuclear membrane to interact with the DNA, one embodiment of this invention [SEQ ID NO: 7] has the xylR gene connected to a nuclear localization signal (NLS of SEQ ID NO: 5) to facilitate transport to the *S. cerevisiae* nucleus where the mode of transcription regulation is occurring. Another embodiment of the invention [SEQ ID NO: 8] comprises xylR connected to an NLS and the *S. cerevisiae SSN6* gene. The Ssn6p protein is a chromatin modifying protein that mediates transcriptional repression. Addition of either the NLS or the NLS and *SSN6* connected to xylR increased repression (FIG. 3).

Example 2: Transcriptional Activity assay

[0054] Transcriptional activity of the promoter was assayed by operably linking the β -galactosidase gene (*lacZ*) to the promoter. The Beta-Glo Assay system (Promega; Madison, WI, USA) was used to determine the level of transcriptional activity from promoter::*lacZ* constructs, essentially as reported in (Hector R. E., et al., 2009, The *Saccharomyces cerevisiae YMR315W* gene encodes an NADP(H)-specific oxidoreductase regulated by the transcription factor Stb5p in response to NADPH limitation, *N Biotechnol*, 26:171-180). Expression of the *lacZ* gene was

assayed from cells grown in the presence or absence of xylose (**FIG. 2, FIG. 3, and FIG. 4**).

Cells from cultures in log-phase (0.2 to 0.5 OD₆₆₀) were diluted in fresh medium to a final OD₆₆₀ = 0.004. Assays were started by adding 50 µL of diluted cells to 50 µL of Beta-Glo reagent, mixed thoroughly, and incubated at room temperature. The Beta-Glo reagent contains a detergent that lyses cells to release the β-galactosidase present. Using these conditions, activity measurements were stable from 60 to 120 min. All assays were performed in 96-well, opaque (white), flat-bottomed microtiter plates. At 60 min the samples were read using the luminescence mode of a SpectraMax M5 microplate reader (Molecular Devices; Sunnyvale, CA, USA). β-galactosidase activities reported in the figures are based on the Relative Light Units (RLU) measured. Each assay was initiated using the same amount of cell mass to minimize variation due to differing cell concentrations.

[0055] To the extent that the term “includes” or “including” is employed in the detailed description or the claims, it is intended to be inclusive in a manner similar to the term “comprising” as that term is interpreted when employed as a transitional word in a claim. Furthermore, to the extent that the term “or” is employed in the detailed description or claims (e.g., A or B) it is intended to mean “A or B or both”. When the applicants intend to indicate “only A or B but not both” then the term “only A or B but not both” will be employed. Thus, use of the term “or” herein is the inclusive, and not the exclusive use. See, Bryan A. Garner, A Dictionary of Modern Legal Usage 624 (2d. Ed. 1995). Also, to the extent that the terms “in” or “into” are used in the specification or the claims, it is intended to additionally mean “on” or “onto.” Furthermore, to the extent the term “connect” is used in the specification or claims, it is intended to mean not only “directly connected to,” but also “indirectly connected to” such as connected through another component or components.

[0056] While the invention has been described with reference to details of the illustrated embodiment, these details are not intended to limit the scope of the invention as defined in the appended claims. All cited references and published patent applications cited in this application are incorporated herein by reference. The embodiment of the invention in which exclusive property or privilege is claimed is defined as follows:

CLAIMS

1. An isolated nucleic acid molecule that has promoter activity specific to xylose and that comprises a DNA sequence selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4, wherein said isolated nucleic acid molecule is operatively linked to at least one heterologous nucleic acid sequence of interest.
2. A vector comprising the isolated polynucleotide of claim 1.
3. A cell comprising the vector of claim 2.
4. The cell of claim 3, wherein the vector is stably integrated into the genome of the cell.
5. A vector comprising a promoter and a heterologous nucleic acid sequence, wherein the promoter comprises a polynucleotide of SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4.
6. A method for expressing a coding region of interest in a transformed yeast cell comprising: a) providing a transformed yeast cell having a recombinant construct, wherein the recombinant construct comprises: (1) a promoter region comprising SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4, and (2) a coding region of interest which is expressible in the yeast cell; wherein the promoter region is operably linked to the coding region of interest; and b) growing the transformed yeast cell of step (a) under conditions whereby the recombinant construct of step (a) is expressed.
7. The method of claim 6, wherein the yeast cell is a member of a genus selected from the group consisting of *Saccharomyces*, *Kluyveromyces*, *Candida*, *Scheffersomyces*, *Spathaspora*, *Yarrowia*, *Schizosaccharomyces*, *Zygosaccharomyces*, *Brettanomyces*, *Debaryomyces*, *Schwanniomyces*, *Pachysolen*, *Torulaspora*, *Hansenula*, or *Pichia*.
8. The method according to claim 6, wherein the coding region of interest encodes a polypeptide, wherein the polypeptide is selected from the group consisting of:

xylanases, xylose reductases, xylose dehydrogenases, xylitol dehydrogenases, xylulokinases, xylose transporters, glucose transporters, galactose transporters, myo-inositol transporters, xylose isomerases, transhydrogenases, NADH kinases, NADP-dependent d-glyceraldehyde-3-phosphate dehydrogenases, transketolases, transaldolases, glucose-6-phosphated dehydrogenases, ribulose-5-phosphate isomerase, ribulose-5-phosphate epimerases, phosphoglucose isomerases, alcohol dehydrogenases, aldehyde dehydrogenases, 2-pyrone synthases, beta-xylosidases, acetyl-CoA synthases, acetyl-CoA carboxylase, phosphoketolases, acetate kinases, transcription factors, and phosphotransacetylases.

9. A transformed yeast comprising a promoter and a heterologous gene encoding a protein, wherein the promoter comprises a polynucleotide of SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4, wherein the polynucleotide is operably linked to said heterologous gene and causes transcription of said heterologous gene when xylose is available.
10. The transformed yeast of claim 9 wherein the transformed yeast is a *Saccharomyes cerevisiae*, *Kluyveromyces*, *Candida*, *Scheffersomyces*, *Spathaspora*, *Yarrowia*, *Schizosaccharomyces*, *Zygosaccharomyces*, *Brettanomyces*, *Debaryomyces*, *Schwanniomyces*, *Pachysolen*, *Torulaspora*, *Hansenula*, or *Pichia*.

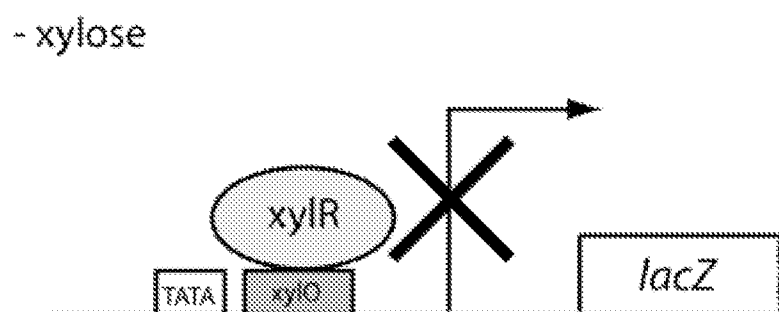


FIG. 1A

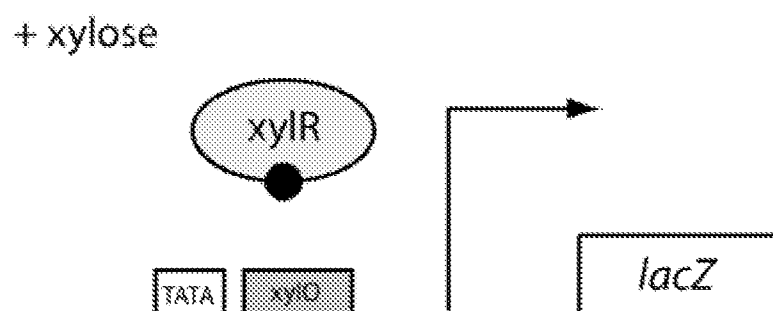


FIG. 1B

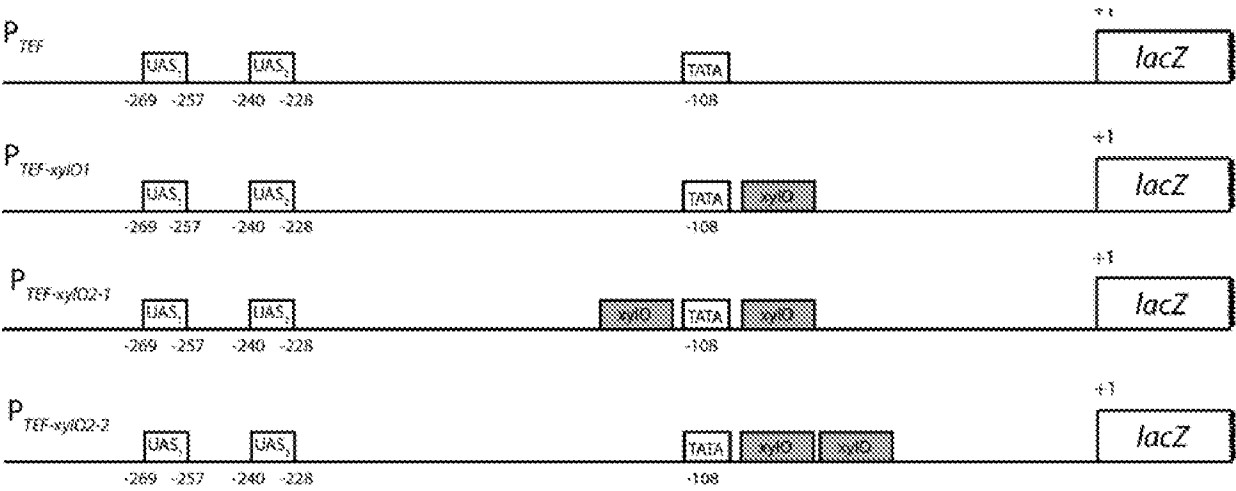


FIG. 1C

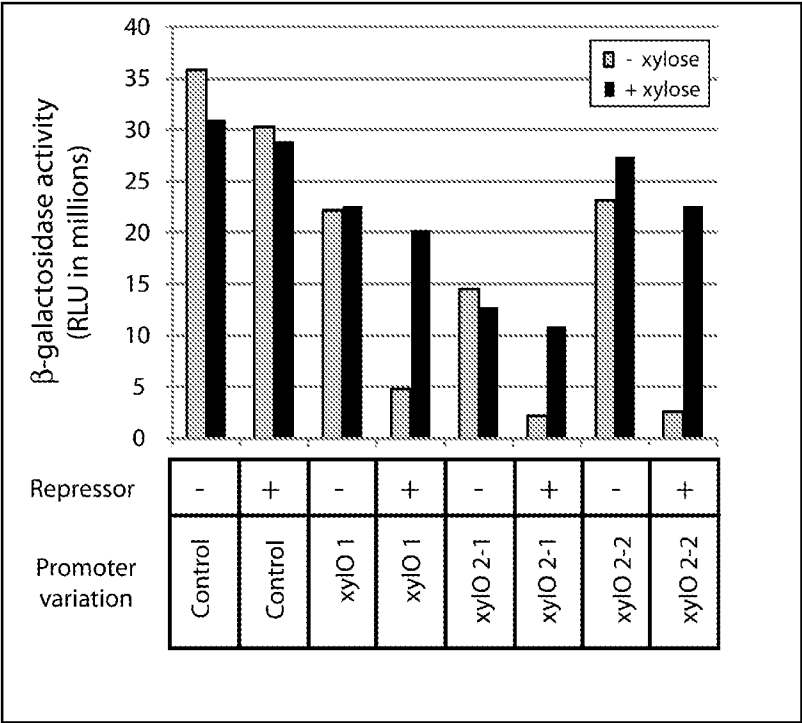


FIG. 2

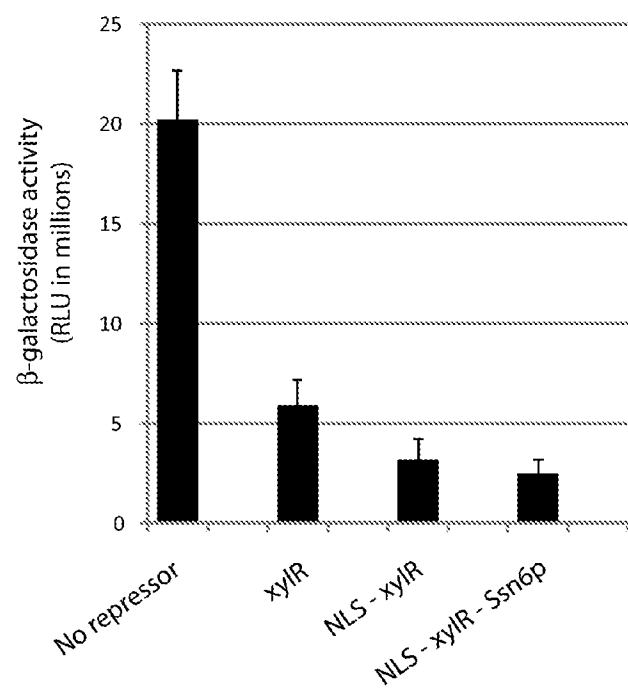


FIG. 3

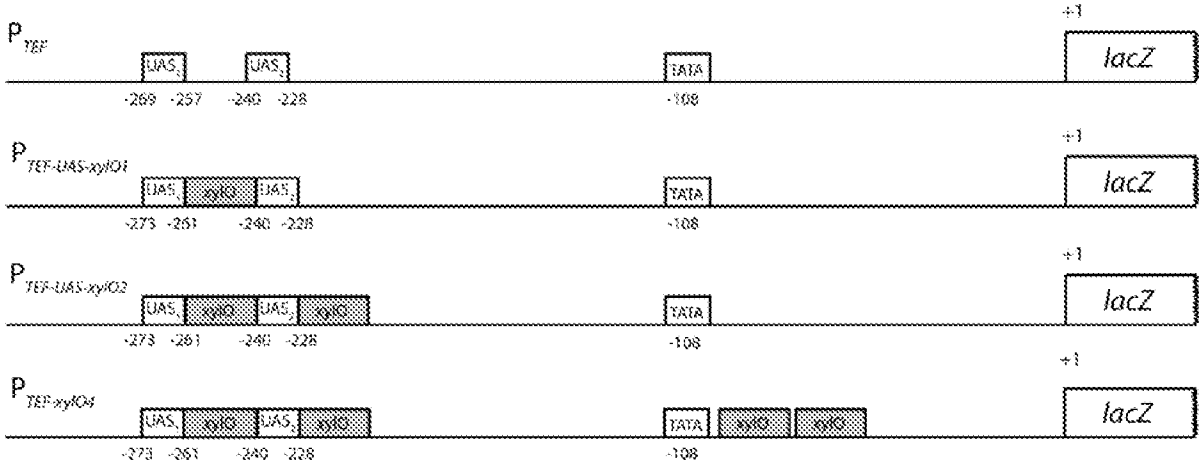


FIG. 4A

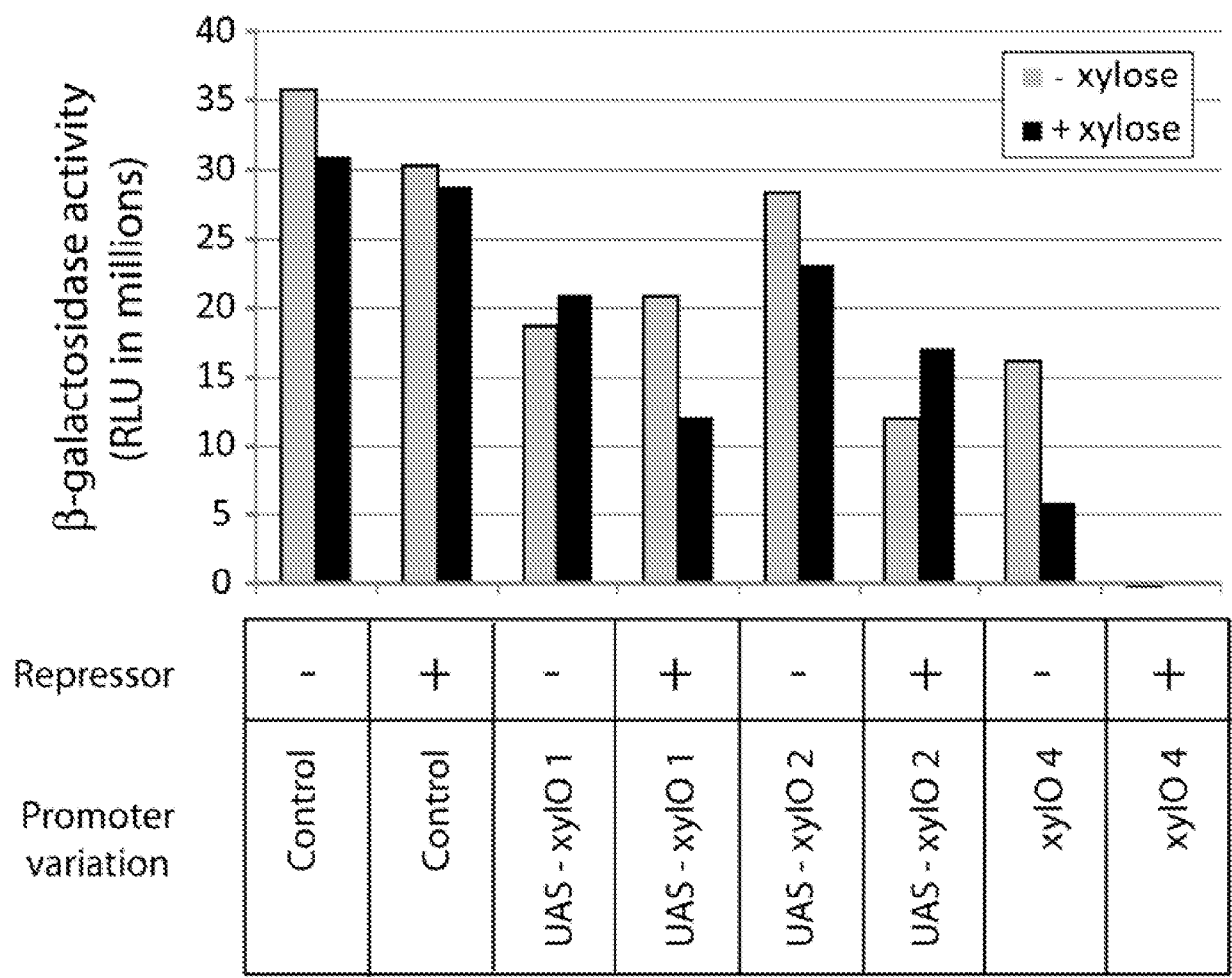


FIG. 4B

A. CLASSIFICATION OF SUBJECT MATTER**C12N 15/81(2006.01)i, C12N 1/16(2006.01)i**

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12N 15/81; C12N 15/63; C12N 15/70; C12N 1/21; C12N 1/16

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS(KIPO internal) & Keywords: xylose, promoter, vector

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LEE, SUN-MI et al., 'Systematic and evolutionary engineering of a xylose isomerase-based pathway in <i>Saccharomyces cerevisiae</i> for efficient conversion yields', <i>Biotechnology for Biofuels</i> , 2014, Vol.7, Article No.122 (internal pages 1-8) See abstract; page 7; table 1; figs. 1-2.	1-10
A	NCBI, GenBank accession no. X73978.1 (18 April 2005) See the sequence.	1-10
A	KR 10-2011-0124739 A (KYUNGPOOK NATIONAL UNIVERSITY INDUSTRY-ACADEMIC COOPERATION FOUNDATION) 17 November 2011 See abstract; claims 1-3; paragraph [0006].	1-10
A	STEINER, SABINE et al., 'Sequence and promoter analysis of the highly expressed TEF gene of the filamentous fungus <i>Ashbya gossypii</i> ', <i>Molecular and General Genetics MGG</i> , 1994, Vol.242, Issue 3, pages 263-271 See the whole document.	1-10
A	KR 10-2012-0099202 A (KOREA RESEARCH INSTITUTE OF BIOSCIENCE AND BIOTECHNOLOGY) 07 September 2012 See the whole document.	1-10



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

20 October 2016 (20.10.2016)

Date of mailing of the international search report

20 October 2016 (20.10.2016)

Name and mailing address of the ISA/KR

International Application Division

Korean Intellectual Property Office

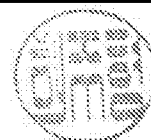
189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

HEO, Joo Hyung

Telephone No. +82-42-481-8150



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2016/040050

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
KR 10-2011-0124739 A	17/11/2011	KR 10-1207629 B1	03/12/2012
KR 10-2012-0099202 A	07/09/2012	CN 102459650 A	16/05/2012
		CN 102459650 B	01/10/2014
		EP 2465946 A2	20/06/2012
		KR 10-1128080 B1	29/03/2012
		KR 10-1222056 B1	14/01/2013
		US 2012-0238470 A1	20/09/2012
		WO 2010-143871 A2	16/12/2010
		WO 2010-143871 A3	21/04/2011
		WO 2010-143871 A9	03/03/2011