Metabolic Engineering Communications 12 (2021) e00160

Contents lists available at ScienceDirect

Metabolic Engineering Communications

journal homepage: www.elsevier.com/locate/mec

The identification of novel promoters and terminators for protein expression and metabolic engineering applications in *Kluyveromyces marxianus*

Pradeep Kumar [a], Debendra Kumar Sahoo [a], Deepak Sharma [a,b,*]

[a] Council of Scientific and Industrial Research-Institute of Microbial Technology, India
[b] Academy of Scientific & Innovative Research, India

ARTICLE INFO

Keywords:
Kluyveromyces marxianus
Protein expression
Metabolic engineering
Promoter
Terminator
β-Galactosidase

ABSTRACT

The *K. marxianus* has emerged as a potential yeast strain for various biotechnological applications. However, the limited number of available genetic tools has hindered the widespread usage of this yeast. In the current study we have expanded the molecular tool box by identifying novel sets of promoters and terminators for increased recombinant protein expression in *K. marxianus*. The previously available transcriptomic data were analyzed to identify top 10 promoters of highest gene expression activity. We further characterized and compared strength of these identified promoters using eGFP as a reporter protein, at different temperatures and carbon sources. To examine the regulatory region driving protein expression, serially truncated shorter versions of two selected strong promoters were designed, and examined for their ability to drive eGFP protein expression. The activities of these two promoters were further enhanced using different combinations of native transcription terminators of *K. marxianus*. We further utilized the identified DNA cassette encoding strong promoter in metabolic engineering of *K. marxianus* for enhanced β-galactosidase activity. The present study thus provides novel sets of promoters and terminators as well as engineered *K. marxianus* strain for its wider utility in applications requiring lactose degradation such as in cheese whey and milk.

1. Introduction

The use of microorganisms for various biotechnological applications has been greatly expanded from producing beverages to food, enzymes, biofuels, flavors and pharmaceuticals. The increased use of these organisms is primarily due to the advancement of various genetics and molecular biology tools, which has facilitated metabolic engineering of biochemical pathways for desired applications (Lian et al., 2018; Nielsen and Jewett, 2008). The yeast-based system has gained more potential as it is amenable to genetic manipulations and also possesses various biochemical pathways found in higher eukaryotes such as mammalian or plant cells (Chakraborty et al., 2016; Mohammadi et al., 2015). Some of the commonly employed yeasts for research as well as industrial processes are *Saccharomyces cerevisiae, Pichia pastoris*, *Hansenula polymorpha*, *Kluyveromyces lactis* and *Yarrowia lipolytica.* Recently another yeast *Kluyveromyces marxianus*, a sister species of *K. lactis*, has also emerged as a valuable strain for biotechnological applications (Karim et al., 2020; Lane and Morrissey, 2010).

The *K. marxianus* is a non-conventional, non-pathogenic and GRAS (generally regarded as safe) organism. *K. marxianus* possesses various desirable properties which is absent in many other commonly used yeast, and thus the strain is a potential substitute for biotechnological needs. It's a fastest growing organism as compared to any of the other known eukaryotes (Groeneveld et al., 2009). It's a thermotolerant yeast that could grow up to 52 °C (Banat et al., 1992). *K. marxianus* can assimilate various sugars such as glucose, lactose, mannose, galactose and xylose (Fonseca et al., 2013) (Rodrussamee et al., 2011). The availability of its whole genome sequence could be used for desired genetic manipulations.

In contrast to extensively studied *S. cerevisiae*, not much is established in *K. marxianus*. The lack of efficient genetic tools, limited number of auxotrophic markers, limited choice of constitutive or inducible promoters is the major drawback for its metabolic engineering, and usage for biotechnological applications. The presence of dominant nonhomologous end joining (NHEJ) over homologous recombination in *K. marxianus* is another limitation for construction of specific gene knockout strain (Abdel-Banat et al., 2010). The limitation has been partially overcome by the development of CRISPR-Cas9 system that has improved the efficiency to construct specific gene knockout strain (Löbs et al.,

* Corresponding author. CSIR-Institute of Microbial Technology, Sector 39A, Chandigarh, India.
E-mail address: deepaks@imtech.res.in (D. Sharma).

https://doi.org/10.1016/j.mec.2020.e00160
Received 7 September 2020; Received in revised form 23 December 2020; Accepted 24 December 2020

2017). Similarly, a CRISPR interference (CRISPRi) method based upon inactive Cas9, has been successfully used for simultaneous suppression of multiple genes for improved ethyl acetate production in *K. marxianus* (Löbs et al., 2018). The *K. marxianus* has shown its potential usage in production of various biomolecules including proteins, ethanol, oligosaccharides and oligopeptides (Belem et al., 1999; Cheng et al., 2006; Yanase et al., 2010). Various enzymes, such as Inulinase, aminopeptidases, β-galactosidase have been successfully produced in their active state from *K. marxianus* (Bansal et al., 2008; Ramírez-Zavala et al., 2004; Zhou et al., 2014). To fully exploit the unique properties of *K. marxianus*, it is desirable to expand the tool-box for carrying out genetic manipulations. The ability to modulate a biochemical pathway by varying the expression level of its enzymes and proteins is required for the development of yeast strain for desired application. Thus the identification of novel promoters that could be used in the presence of various sugars and different temperatures would further enhance the potential of *K. marxianus* as a substitute for *P. pastoris* or other yeasts.

Various approaches such as optimization of signal sequences and exploration of strong promoters are currently underway for metabolic engineering applications in *K. marxianus*. Both types of promoters, native as well as those derived from budding yeast *S. cerevisiae* have been examined in *K. marxianus*. Promoters of *S. cerevisiae* origin used for expression in *K. marxianus* are of galactokinase (P_{ScGAL1}) (Almeida et al., 2003), glyceraldehyde-3-dehydrogenase (P_{ScTDH3}) (Nonklang et al., 2009), 3-phosphoglycerate kinase (P_{ScPGK1}) (Ball et al., 1999; Pecota et al., 2007) and alcohol dehydrogenase (P_{ScADH1}) (Lee et al., 2013). Among these promoters, the P_{ScGAL1} promoter is galactose regulatable, and other such as P_{ScTDH3}, P_{ScPGK1} and P_{ScADH1} are constitutively expressed. Similarly, native promoters of *K. marxianus* origin such as P_{KmINU1} (promoter of inulinase enzyme), P_{KmGAL1}, P_{KmTDH3}, P_{KmPGK1} and P_{KmADH1} have been employed to express recombinant proteins (Akada et al., 2014; Bergkamp et al., 1993; Yang et al., 2015; Zhang et al., 2015). The expression from P_{KmINU1} and P_{KmGAL1} is carbon source dependent. Furthermore, attempts have been made to improve strength of native promoters such as of P_{KmINU1} (Zhou et al., 2018). Studies show that under growth conditions where glucose is used as carbon source, the native promoters of *K. marxianus* show higher strength than homologous promoter derived from *S. cerevisiae* paving way for further exploration of native promoters of *K. marxianus* (Yang et al., 2015).

In addition to promoters, terminators are known to play critical role in protein expression by modulating the stability of mRNA. Only limited number of terminators sequences have been examined in *K. marxianus*. The different gene-terminators of *K. marxianus* and *S. cerevisiae* origin have been used to modulate promoter strength (Rajkumar et al., 2019). These studies suggest that further expansion of number of terminators sequences would provide additional control for gene expression for metabolic engineering applications in *K. marxianus*.

Overall there are only limited number of promoters as well as terminator sequences available for protein production and metabolic engineering applications in *K. marxianus*, and thus there is a need to identify novel promoters and terminators for modulation of desirable biochemical pathways. Thus in the current study, we carried out analysis of available transcriptomic profile of *K. marxianus*, and identified novel promoters for high expression of proteins, enzymes and metabolic engineering applications. Also, using different sets of terminator sequences, we further enhanced promoter strength. Using the identified combination of these promoters and terminators, we expressed HA antigen of influenza virus. We further used the promoter cassette to engineer *K. marxianus* with improved β-galactosidase production. The study thus provides novel sets of promoters and terminators for high level production of heterologous proteins and thus contributes to novel genetic tools for use of *K. marxianus* for metabolic engineering and biotechnological applications. The genetic tools developed in the present study would aid to further understanding of cellular biology of *K. marxianus*.

2. Materials and methods

2.1. Plasmid, strains and growth media

The *K. marxianus* NBRC1777 was obtained from Biological Resource Centre, NITE (NBRC), Japan. The *S. cerevisiae* strain BY4741 (*MATa his3Δ1, leu2Δ0, met15Δ0, ura3Δ0)* was used. The plasmid used in the study are listed in Table 3. SD (Synthetic-defined) media was composed of yeast nitrogen base (YNB) with ammonium sulfate (6.7 g/l) and 2% dextrose (w/v). SX media composition is similar to SD media except that 2% (w/v) xylose instead of dextrose provided carbon source. Amino acids were supplemented as required. YPD media is composed of 1% yeast extract, 2% peptone and 2% dextrose (w/v). YPX and YPL media is same as YPD except that 2% xylose or lactose was used instead of dextrose.

2.2. Development of K. marxianus expression vectors

The integrating shuttle vector pRS306 was used as a backbone for construction of recombinant yeast expression vector. First, 1265 bp comprising of ARS/CEN sequence was PCR amplified from the genomic DNA of *K. marxianus* strain NBRC1777. Amplified product was inserted into the vector pRS306 by Transfer-PCR method (Erijman et al., 2011). The resultant expression vector is referred to as pPKM316. Subsequently, the DNA encoding eGFP and T_{ScCYC1} terminator (of *S. cerevisiae*) was subcloned into pPKM316 using restriction sites BamHI and KpnI to construct pPKM316-eGFP-T_{ScCYC1}. The DNA encoding desired promoter (Table S2) was PCR amplified from genomic DNA of *K. marxianus* strain NBRC1777 using promoter specific forward and reverse primers (Table S1). The amplified DNA sequences were digested with restriction enzymes SacI and XbaI. The digested product was ligated into pPKM316-eGFP-T_{ScCYC1} to develop expression vector encoding eGFP under the control of desired promoter. The plasmid encodes in 5′ to 3’ direction, the desired promoter, eGFP and T_{ScCYC1} terminator.

For studying the role of different terminators, the T_{ScCYC1} terminator was later replaced by desired terminator sequences. The desired terminator sequences ($T_{KmIMTT1}$ or $T_{KmIMTT2}$) were PCR amplified from *K. marxianus* genome, and subcloned into pPKM316-$P_{KmIMTCP2}$-eGFP plasmid using XhoI and KpnI restriction sites.

Four different signal sequences (signal sequence of Inulinase with P10L substitution ($ssInu1_{Km}$) (Zhou et al., 2018) and mating α-factor of *K. marxianus* ($ssαMF_{Km}$), Pho1 signal sequence of *P. pastoris* ($ssPho1_{Pp}$), mating α-factor signal sequence with D83E of *S. cerevisiae* ($ssαMF_{Sc}$)) were examined in the present study. The DNA encoding $ssαMF_{Sc}$ was PCR amplified from *P. pastoris* plasmid pPic9 (Invitrogen #K1710-01). The DNA encoding $ssαMF_{Km}$ was PCR amplified from the *K. marxianus* genomic DNA. The DNA encoding $ssInu1_{Km}$ and $ssPho1_{Pp}$ signal sequence were constructed using overlap PCR. The DNA cassette encoding from 5′ to 3′, $P_{KmIMTCP2}$ promoter and desired signal sequences were constructed using overlap PCR and further subcloned using restriction sites SacI and XbaI into expression plasmid at 5’ end of gene encoding eGFP.

For β-galactosidase expression, the β-galactosidase gene sequence was PCR amplified from the genomic DNA of *K. marxianus* strain NBRC1777 and subcloned at SpeI and XhoI restriction sites under desired promoter and terminator.

For PCR amplification, the initial denaturation was carried out for 5 min at 98 °C, followed by additional 30 cycles of denaturation for 30 s at 98 °C, primer annealing for 30 s at 55–60 °C and extension at 72 °C. Then one final extension time of 10 min at 72 °C was carried out.

2.3. Construction of K. marxianus strain with uracil auxotrophy

The uracil auxotrophy of *K. marxianus* was constructed using homologous recombination of DNA cassette lacking Ura3 encoding gene at Ura3 locus. The DNA cassette lacking *URA3* and having homology of 250

Table 1
Fold change in mRNA level of 10 most abundant transcripts. The fold change was calculated by dividing the transcript value of the gene by the lowest non-zero transcript value (0.03) reported in (Lertwattanasakul et al., 2015).

Promoter/Gene notation	Name	CDS product	30 °C (Fold change)	45 °C (Fold change)	Length of Promoter (bp)
$P_{KmIMTCP1}$/IMTCP1	KLMA_10040	Uncharacterized protein YDR524C–B	17169530	7149637	1000
$P_{KmIMTCP2}$/IMTCP2	KLMA_40174	Uncharacterized cell wall protein YDR134C	1362106	1221608	999
$P_{KmIMTCP3}$/IMTCP3	KLMA_40094	40 S ribosomal protein S13	213216.4	438075.7	438
$P_{KmIMTCP4}$/IMTCP4	KLMA_10830	Ribosyldihydronicotinamide dehydrogenase [quinone]	101348.9	451185.9	594
$P_{KmIMTCP5}$/IMTCP5	KLMA_20547	Protein BTN2	131705.2	602783.1	479
$P_{KmIMTCP6}$/IMTCP6	KLMA_10518	Inulinase	431333.3	20887.56	930
$P_{KmIMTCP7}$/IMTCP7	KLMA_40218	Glyceraldehyde-3-phosphate dehydrogenase 1	420417	79451.56	1008
$P_{KmIMTCP8}$/IMTCP8	KLMA_70281	Asparagine synthetase [glutamine-hydrolyzing] 1	23655.22	353828.3	1000
$P_{KmIMTCP9}$/IMTCP9	KLMA_10179	Glycerol-3-phosphate dehydrogenase [NAD+] 1	76492.56	40412	1000
$P_{KmIMTCP10}$/IMTCP10	KLMA_10462	Enolase	143882.2	26598.33	977
P_{KmTDH3}/TDH3	KLMA_80059	Glyceraldehyde-3-phosphate dehydrogenase 3	489093.7	81713.78	1060
P_{KmPGK1}/PGK1	KLMA_10540	Phosphoglycerate kinase	15041.11	3238.889	1035

Table 2
Promoter and terminator combinations used in this study.

S. No.	Combination
1.	$P_{KmIMTCP1}$-T_{ScCYC1}
2.	$P_{KmIMTCP2}$-T_{ScCYC1}
3.	$P_{KmIMTCP3}$-T_{ScCYC1}
4.	$P_{KmIMTCP4}$-T_{ScCYC1}
5.	$P_{KmIMTCP5}$-T_{ScCYC1}
6.	$P_{KmIMTCP6}$-T_{ScCYC1}
7.	$P_{KmIMTCP7}$-T_{ScCYC1}
8.	$P_{KmIMTCP8}$-T_{ScCYC1}
9.	$P_{KmIMTCP9}$-T_{ScCYC1}
10.	$P_{KmIMTCP10}$-T_{ScCYC1}
11.	P_{KmTDH3}-T_{ScCYC1}
12.	P_{KmPGK1}-T_{ScCYC1}
13.	$P_{KmIMTCP1}$-$T_{KmIMTT1}$
14.	$P_{KmIMTCP1}$-$T_{KmIMTT2}$
15.	$P_{KmIMTCP2}$-$T_{KmIMTT1}$
16.	$P_{KmIMTCP2}$-$T_{KmIMTT2}$
17.	P_{ScGPD}-T_{ScCYC1}
18.	P_{ScGPD}-$T_{KmIMTT1}$
19.	P_{ScGPD}-$T_{KmIMTT2}$

base pair (bp) upstream and 250bp downstream sequence of URA3 gene was synthesized from GenScript USA, Inc. The cassette was transformed into *K. marxianus* strain. The transformants with uracil auxotropy was identified by their inability to grow on solid media lacking uracil as well as growth onto 5-fluoroorotic acid (5-FOA) containing growth media.

2.4. Immunoblot analysis

NBRC1777 *ura3Δ* cells (10 ml culture volume in 50 ml flasks) harboring desired plasmid were grown until mid-log phase. The cells were harvested by centrifugation and washed with PBS buffer. The cells were lysed, and cell-debris was removed by centrifugation at 3000*g* for 2 min. The lysate supernatant was used for immunoblotting. About 10–15 μg of total protein from cell lysate was subjected to electrophoresis onto a 12% SDS–polyacrylamide gel. The proteins were then transferred onto PVDF membrane. A 1:5000 dilution of an anti-GFP (cat. No. MA5-15256; thermo fisher scientific) or anti-Pgk1 (Life Technologies, USA-459250) or anti- His_6 (Pierce, USA-MA1-21315) antibody was used for detection.

To detect eGFP in culture growth media, the cells were separated from growth media by centrifugation at 12000*g* for 30 min. The supernatant was collected, and 20 μl of supernatant was used for immunoblotting.

2.5. In vivo luciferase assay

The gene encoding luciferase (FLuc) was PCR amplified and subcloned into plasmid downstream of desired promoter cassette using restriction sites BamHI and XhoI. The designed plasmids encoding luciferase were transformed into yeast *K. marxianus* or *S. cerevisiae*. The 3–4 transformants were pooled and grown overnight into SD liquid growth media lacking uracil. The overnight grown primary culture was

Table 3
Plasmids constructed in the present study.

S. No.	Plasmid	Marker
1.	pPKM316-$P_{KmIMTCP1}$-eGFP-T_{ScCYC1}	URA3
2.	pPKM316-$P_{KmIMTCP2}$-eGFP-T_{ScCYC1}	URA3
3.	pPKM316-$P_{KmIMTCP3}$-eGFP-T_{ScCYC1}	URA3
4.	pPKM316-$P_{KmIMTCP4}$-eGFP-T_{ScCYC1}	URA3
5.	pPKM316-$P_{KmIMTCP5}$-eGFP-T_{ScCYC1}	URA3
6.	pPKM316-$P_{KmIMTCP6}$-eGFP-T_{ScCYC1}	URA3
7.	pPKM316-$P_{KmIMTCP7}$-eGFP-T_{ScCYC1}	URA3
8.	pPKM316-$P_{KmIMTCP8}$-eGFP-T_{ScCYC1}	URA3
9.	pPKM316-$P_{KmIMTCP9}$-eGFP-T_{ScCYC1}	URA3
10.	pPKM316-$P_{KmIMTCP10}$-eGFP-T_{ScCYC1}	URA3
11.	pPKM316-P_{KmTDH3}-eGFP-T_{ScCYC1}	URA3
12.	pPKM316-P_{KmPGK1}-eGFP-T_{ScCYC1}	URA3
13.	pPKM316-$P_{KmIMTCP1\text{-}787}$-eGFP-T_{ScCYC1}	URA3
14.	pPKM316-$P_{KmIMTCP1\text{-}612}$-eGFP-T_{ScCYC1}	URA3
15.	pPKM316-$P_{KmIMTCP1\text{-}400}$-eGFP-T_{ScCYC1}	URA3
16.	pPKM316-$P_{KmIMTCP1\text{-}194}$-eGFP-T_{ScCYC1}	URA3
17.	pPKM316-$P_{KmIMTCP2\text{-}699}$-eGFP-T_{ScCYC1}	URA3
18.	pPKM316-$P_{KmIMTCP2\text{-}499}$-eGFP-T_{ScCYC1}	URA3
19.	pPKM316-$P_{KmIMTCP2\text{-}408}$-eGFP-T_{ScCYC1}	URA3
20	pPKM316-$P_{KmIMTCP2\text{-}299}$-eGFP-T_{ScCYC1}	URA3
21.	pPKM316-$P_{KmIMTCP2\text{-}199}$-eGFP-T_{ScCYC1}	URA3
22.	pPKM316-$P_{KmIMTCP2\text{-}111}$-eGFP-T_{ScCYC1}	URA3
23.	pPKM316-$P_{KmIMTCP2}$-eGFP-$T_{KmIMTT1}$	URA3
24.	pPKM316-$P_{KmIMTCP2}$-eGFP-$T_{KmIMTT2}$	URA3
25.	pPKM316-$P_{KmIMTCP1}$-FLuc-$T_{KmIMTT1}$	URA3
26.	pPKM316-$P_{KmIMTCP1}$-FLuc-$T_{KmIMTT2}$	URA3
27.	pPKM316-$P_{KmIMTCP2}$-FLuc-$T_{KmIMTT1}$	URA3
28.	pPKM316-$P_{KmIMTCP2}$-FLuc-$T_{KmIMTT2}$	URA3
29.	pPKM316-$P_{KmIMTCP1}$-FLuc-T_{ScCYC1}	URA3
30.	pPKM316-$P_{KmIMTCP2}$-FLuc-T_{ScCYC1}	URA3
31.	pPKM316-P_{KmTDH3}-FLuc-T_{ScCYC1}	URA3
32.	pPKM316-P_{KmPGK1}-FLuc-T_{ScCYC1}	URA3
33.	pPKM316-$P_{KmIMTCP2}$-ss$Inu1_{Km}$-eGFP-$T_{KmIMTT1}$	URA3
34.	pPKM316-$P_{KmIMTCP2}$-ss$Pho1_{Pp}$-eGFP-$T_{KmIMTT1}$	URA3
35.	pPKM316-$P_{KmIMTCP2}$-ssαMF$_{Sc}$-eGFP-$T_{KmIMTT1}$	URA3
36.	pPKM316-$P_{KmIMTCP2}$-ssαMF$_{Km}$-eGFP-$T_{KmIMTT1}$	URA3
37.	pPKM316-$P_{KmIMTCP2}$-HA-$T_{KmIMTT1}$	URA3
38.	pPKM316-P_{KmPGK1}-β-Gal-T_{ScCYC1}	URA3
39.	pPKM316-P_{KmTDH3}- β-Gal-T_{ScCYC1}	URA3
40.	pPKM316-$P_{KmIMTCP2}$- β-Gal-$T_{KmIMTT1}$	URA3
41.	pRS316-P_{ScGPD}-FLuc-T_{ScCYC1}	URA3
42.	pRS316-$P_{KmIMTCP1}$-FLuc-T_{ScCYC1}	URA3
43.	pRS316-$P_{KmIMTCP2}$-FLuc-T_{ScCYC1}	URA3
44.	pRS316-P_{ScGPD}-FLuc-$T_{KmIMTT1}$	URA3
45.	pRS316-P_{ScGPD}-FLuc-$T_{KmIMTT2}$	URA3
46.	pRS316-$P_{KmIMTCP1}$-FLuc-$T_{KmIMTT1}$	URA3
47.	pRS316-$P_{KmIMTCP1}$-FLuc-$T_{KmIMTT2}$	URA3
48.	pRS316-$P_{KmIMTCP2}$-FLuc-$T_{KmIMTT1}$	URA3
49.	pRS316-$P_{KmIMTCP2}$-FLuc-$T_{KmIMTT2}$	URA3
50.	pRS316-P_{ScGPD}-HA-T_{ScCYC1}	URA3

diluted to 0.05 O.D.$_{600}$ using 20 ml of sterile growth media (in 100 ml flask) with dextrose or xylose as carbon source. Cells were grown at either 30 °C or 45 °C at 200 rpm until cell O.D.$_{600}$ reached 0.8–1.0. The cells were harvested by centrifugation and re-suspended into 1 ml of YNB media. For measuring luciferase activity, 200 μl of 0.3 O.D.$_{600}$ cells were added with 50 μl of 1 mg/ml D-Luciferin (Sigma-Aldrich Cat #L9504), and the luminescence was measured in Multimode Plate Readers (TECAN Infinite M200 PRO, Switzerland).

2.6. Quantitative real-time PCR

Cells, grown in selective SD media (20 ml in 100 ml flasks) at 30 °C and 200 rpm until mid-log phase, were harvested and total RNA was isolated using HiPurA Yeast RNA Purification Kit (HiMedia, India-MB611). The 100 ng of isolated RNA was used to prepare cDNA using a cDNA synthesis kit (BIO-RED Cat#170–8891). For quantitative Real-Time PCR (qRT-PCR), cDNA (1 μl) was amplified using iTaq™ Universal SYBR® Green Supermix qPCR kit (BIO-RED, CAT#172–5124) by using eGFP specific primer (Table S1) and following manufacturer's protocol. The reaction was subjected to an initial denaturation at 95 °C for 2 min and 40 repetitive cycles of 15 s at 95 °C, 30 s at 60 °C in *CFX96* Touch *Real-Time PCR* System (BIO-RED). The 18 S rRNA was selected as a reference gene.

2.7. β-Galactosidase activity assay

The *K. marxianus* cells harboring plasmid encoding β-galactosidase were grown until mid-log phase as 10 ml culture in 50 ml flasks (shaking speed 200 rpm). 1 O.D.$_{600nm}$ cells were treated with 3 rounds of freeze thaw cycle for permeabilization. Cells were further incubated with 0.2 ml of ortho-Nitrophenyl-β-galactoside (ONPG 4 mg/ml) for 15 min. The reaction was stopped with 0.3 ml of 1 M Na_2CO_3. Cells were centrifuged at 12000g for 10 min and the enzyme activity was monitored in supernatant by measuring absorbance at 420 nm.

3. Results

3.1. The transcriptomic analysis of gene expression reveals the promoters of higher strength

As promoter act as a regulatory region for gene transcription, the relatively higher number of a gene transcript is likely to be expressed from a stronger promoter. Thus, to identify strong promoters in *K. marxianus*, we analyzed the transcription profile of cellular genes as a reporter of the relative strength of promoters. A previous study has reported transcriptomic data of *K. marxianus* cells grown at 30 °C in liquid YPD media under shaking condition (Lertwattanasakul et al., 2015). To understand the relative increase in the number of transcripts, we normalized the available gene expression data of each gene by dividing its number of transcripts by the lowest positive transcript number (0.03 ppm at 30 °C).

Table 1 provides fold increase of the top 10 highly expressed gene transcripts. Interestingly the top 2 highly expressed genes, IMTCP1 and IMTCP2, encode for yet uncharacterized proteins. One of the gene, IMTCP3 codes for a subunit of 40 S ribosomal protein. The ribosomal proteins are known to be abundantly present in the cell (Ho et al., 2018). The gene products driven by previously used strong promoters such as glyceraldehyde 3-phosphate dehydrogenase 1, inulinase, enolase were also found to be among top 10 identified transcripts suggesting that these identified promoters are likely to be strong promoters for use in *K. marxianus*.

3.2. Promoter driving IMTCP1 or IMTCP2 shows the highest expression of reporter protein eGFP

To examine the relative strength of the above-identified promoters, we constructed an expression plasmid compatible for use in *K. marxianus*. The plasmid was further used to express eGFP encoded under the identified promoters.

For the construction of the recombinant yeast expression vector, the ARS/CEN sequence of *K. marxianus* strain NBRC1777 was inserted into integrating shuttle vector pRS306 (ATCC 77141) to generate plasmid pPKM316 (Fig. 1) (Ball et al., 1999). To examine the strength of individual promoter, the gene encoding reporter protein eGFP was subcloned into pPKM316. The DNA encoding each of the identified promoters was further subcloned into the plasmid at region 5' to gene encoding eGFP. Previous studies show that 500–700 bp upstream region is sufficient to drive gene expression (Lang et al., 2020; Lee et al., 2015; Ohler and Niemann, 2001). Thus a minimum length of 500bp was selected as promoter region. We selected either intergenic or 1000 bp upstream region (except for $P_{KmIMTCP4}$) if intergenic region is more than 1000bp, as the promoter length. For $P_{KmIMTCP4}$, instead of 1000bp only 594 bp were selected (due to the presence of a desired restriction site just upstream of 594bp) as the promoter region.

The plasmids encoding eGFP under desired promoters were transformed into *K. marxianus* cells. The transformants were grown into liquid SD media lacking uracil at 30 °C. The cellular lysate prepared from equal O.D.$_{600nm}$ cells was probed with anti-GFP antibodies. The Pgk1 expressed from chromosomal *PGK1* encoded under its native promoter was used as an internal control. As seen in Fig. 2, though Pgk1 is expressed at a similar level, the cellular abundance of eGFP varied with different promoters. We observed that as compared to previously known strong constitutive promoter P_{KmTDH3}, the eGFP expression from $P_{KmIMTCP2}$ was about 2 fold higher. The $P_{KmIMTCP1}$ showed second highest expression as compared to other promoters selected from transcriptomic analysis. The other chosen promoters showed much lower expression than $P_{KmIMTCP2}$.

To examine whether the relative strength of identified promoters varies with temperature, we examined eGFP level from cells grown under shaking conditions in SD media at 45 °C. The cellular lysate was prepared similarly to the method described above, and the eGFP level was measured on immunoblot using anti-GFP antibodies. Similar to as seen above for cell growth at 30 °C, among all promoters that were examined, $P_{KmIMTCP2}$ was found to be the strongest promoter followed by $P_{KmIMTCP1}$ whose strength was found be similar to as of previously used P_{KmTDH3}. The strong activity of $P_{KmIMTCP2}$ is in agreement with a recent study that identified NC1 (smaller fragment of $P_{KmIMTCP2}$) as one of the strong promoters in *K. marxianus* (Lang et al., 2020).

3.3. $P_{KmIMTCP1}$ and $P_{KmIMTCP2}$ show the strongest eGFP expression in growth media containing xylose as carbon source

The type of carbon source varies depending upon the biotechnological applications which influences microbial growth. For example, xylose is one of the primary sources of carbon when lignocellulose hydrolysates is used for biofuel production. We thus examined the relative strength of our identified promoter in media containing xylose instead of dextrose as carbon source, and compared with that of previously used strong promoters P_{KmTDH3} and P_{KmPGK1}. The cells harboring plasmid encoding eGFP under various promoters were grown into liquid growth media with xylose (SX) as the carbon source and the cellular lysate was examined for eGFP expression. We observed that in xylose containing media, the expression of eGFP was higher from promoters $P_{KmIMTCP1}$, $P_{KmIMTCP2}$, $P_{KmIMTCP6}$, and $P_{KmIMTCP9}$ as compared to that from previously used strong promoters, P_{KmTDH3} and P_{KmPGK1} (Fig. 2B). The results from SD and SX media showed that Inulinase promoter ($P_{KmIMTCP6}$) has relatively higher activity in the presence of xylose than dextrose. Among the newly identified promoters, the eGFP expression was highest from $P_{KmIMTCP2}$ followed by from $P_{KmIMTCP1}$. We further compared the ability of two strongest promoters ($P_{KmIMTCP1}$ and $P_{KmIMTCP2}$) to drive eGFP expression in dextrose or xylose containing growth media. As seen in Fig. 2C, though the strength of P_{KmTDH3} varied with the carbon source, the identified strong promoters showed similar eGFP expression in both media

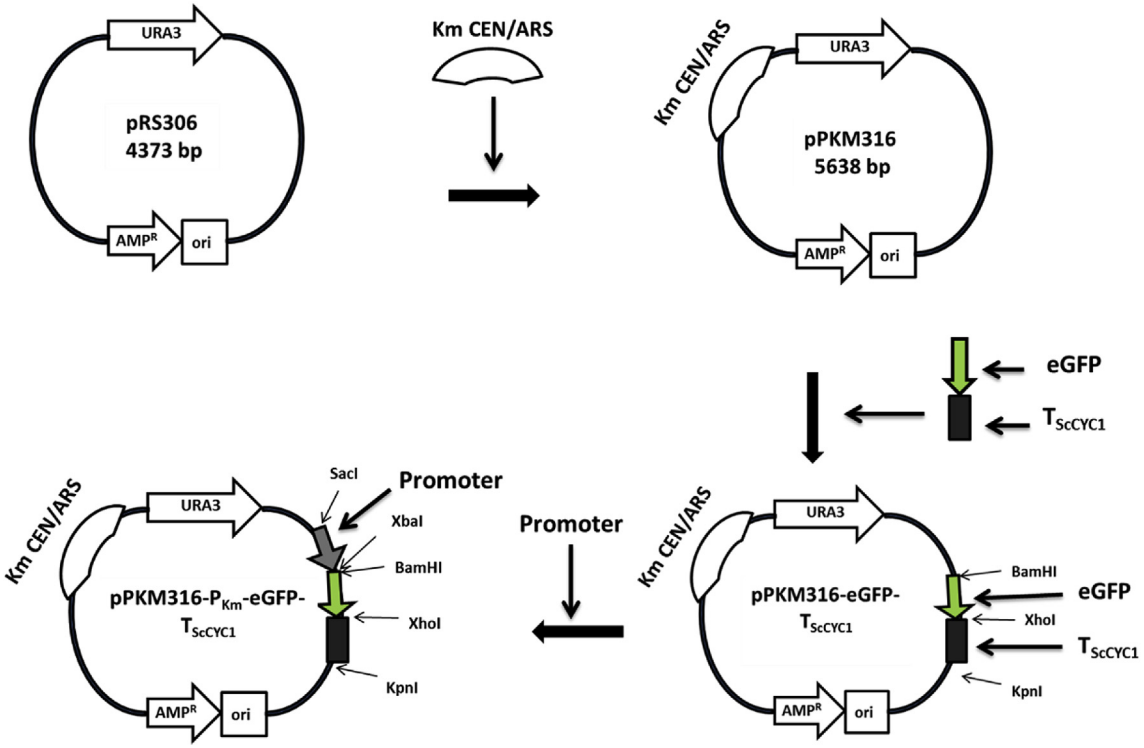

Fig. 1. Schematics showing the methodology used for development of *K. marxianus* expression plasmid. The CEN/ARS origin sequence of *K. marxianus* was inserted into the vector pRS306 to generate pPKM316. The pPKM316 was further used for insertion of cassette encoding desired promoters and terminators.

suggesting that the strength of these newly identified promoters is independent of dextrose or xylose as carbon source. Overall above data suggest that the identified $P_{KmIMTCP1}$ and $P_{KmIMTCP2}$ promoters are more efficient at different temperatures and carbon sources than the previously known promoters in *K. marxianus* (Fig. S1).

3.4. The 388 bp at 5' end of $P_{KmIMTCP1}$ are indispensable for its activity

We further focused on $P_{KmIMTCP1}$ and $P_{KmIMTCP2}$ promoters that showed relatively higher eGFP expression as compared to the other promoters. Promoter region binds to RNA polymerase as well as with other regulatory factors required for the polymerase recruitment. To determine the regulatory region in $P_{KmIMTCP1}$ and $P_{KmIMTCP2}$, we constructed various derivatives of these promoters by sequentially truncating stretch of DNA bases from the 5′ region of the promoter as shown in Fig. 3A and 3B. For $P_{KmIMTCP1}$, 4 truncated derivatives of different lengths were PCR amplified, and subcloned 5′ to gene encoding eGFP in plasmid pPKM316-eGFP-T_{ScCYC1}. The capability of truncated promoters to drive eGFP expression was monitored in *K. marxianus* as described above for full-length promoter at 30 °C in media containing dextrose as a carbon source. The varying level of eGFP expression was observed with promoters of different length (Fig. 3C and 3D). The results show that $P_{KmIMTCP1\text{-}1000}$, $P_{KmIMTCP1\text{-}787}$ and $P_{KmIMTCP1\text{-}612}$ have relatively similar strength. Further truncation of promoter length led to a progressive decrease in eGFP cellular abundance with $P_{KmIMTCP1\text{-}194}$ showing least expression of the protein. Overall, the data suggest that the extreme 388 bp at 5' region of the $P_{KmIMTCP1}$ does not contribute much to its ability to drive gene expression.

Similarly, six truncated promoters of varying lengths based upon the parent $P_{KmIMTCP2}$ promoter were constructed and examined for their strength using eGFP as reporter protein. The expression of eGFP decreased with an increase in truncation of 5' region of $P_{KmIMTCP2}$ suggesting the full-length promoter was required for the maximal protein expression.

3.5. The terminators $T_{KmIMTT1}$ and $T_{KmIMTT2}$ enhance promoter activity by stabilizing mRNA

The cellular abundance of a protein not only depends upon the strength of promoter but also on the terminator sequence that follows 3′ to the encoding gene. The terminator sequence affects gene expression by modulating mRNA stability (Curran et al., 2013; Rajkumar et al., 2019). Thus the relatively higher transcript level of the chromosomally encoded native gene under $P_{KmIMTCP1}$ or $P_{kmIMTCP2}$ could be dependent upon its native terminators. To examine the effect of native terminators of $P_{kmIMTCP1}$ ($T_{KmIMTT1}$) and $P_{kmIMTCP2}$ ($T_{KmIMTT2}$) on $P_{kmIMTCP2}$ strength, we replaced T_{ScCYC1} terminator from 3' end of gene encoding eGFP with that of $T_{KmIMTT1}$ and $T_{KmIMTT2}$. It is reported that terminator sequences generally contains characteristic elements such as AT-rich efficiency elements and A-rich positioning elements (AAG/TAA), and about 200bp downstream sequence from stop codon is sufficient for termination (Helden et al., 2000). To include such characteristic features, we used minimum of 250bp downstream of stop codon (269bp and 290bp for $T_{KmIMTT1}$ and $T_{KmIMTT2}$ respectively) as the terminator sequence.

The plasmid encoding from 5′ to 3' direction $P_{kmIMTCP2}$, eGFP, $T_{KmIMTT1}$/$T_{KmIMTT2}$ was transformed into *K. marxianus* strain. The transformants were grown in liquid SD media at 30 °C, and examined for eGFP expression using immunoblot as described above. Interestingly, the expression of eGFP from $P_{kmIMTCP2}$ was enhanced when $T_{KmIMTT1}$ was used as terminator instead of T_{ScCYC1} (Fig. 4A). Similar increase of eGFP expression was observed with $T_{KmIMTT2}$ terminator as compared to T_{ScCYC1}. To examine whether increase in eGFP level is because of enhanced mRNA stability, we carried out qRT-PCR to examine the transcript level (Fig. 4B). Similar to as seen for increase in eGFP protein abundance, about 4 fold increase of mRNA level was observed when T_{ScCYC1} was replaced with either $T_{KmIMTT1}$/$T_{KmIMTT2}$. Further the eGFP expression from $P_{KmIMTCP2}$-$T_{KmIMTT1}$ was found to similar at different temperatures (18 °C, 30 °C and 37 °C) used for *K. marxianus* growth (Fig. 4C). We also examined the effect of the protein overexpression on

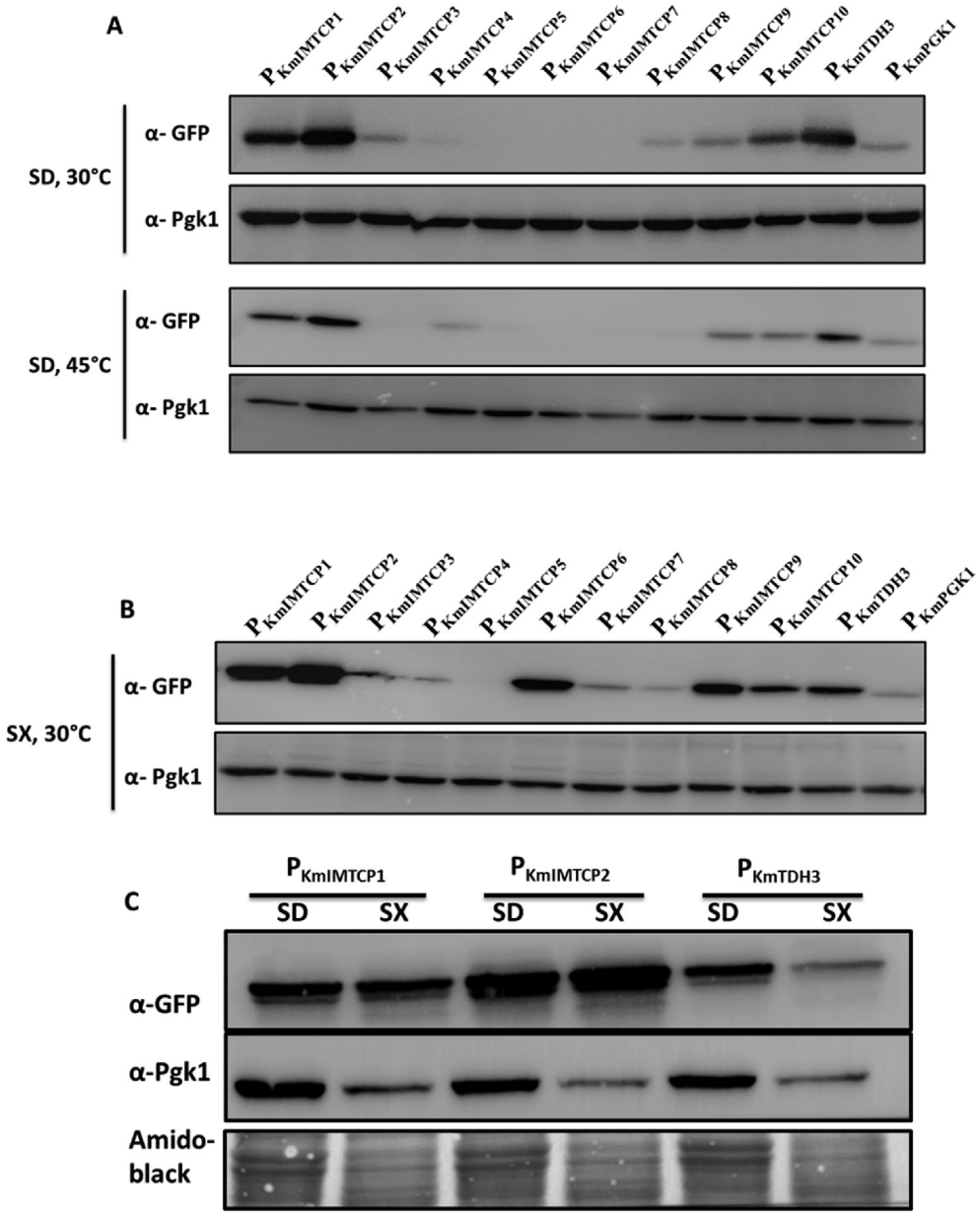

Fig. 2. $P_{KmIMTCP1}$ and $P_{KmIMTCP2}$ show relatively higher expression of reporter protein eGFP. (A) Immunoblot analysis to compare the strength of different promoters using eGFP as a reporter protein and T_{ScCYC1} terminator. The cells were transformed with plasmid encoding indicated promoters. Transformants were grown in liquid growth media containing dextrose (SD) at 30 °C or 45 °C. The cellular lysate was immunoblotted with anti-GFP or anti-PGK1 antibody. Previously known P_{KmTDH3} and P_{KmPGK1} promoters of *K. marxianus* origin were also examined for comparison. **(B)** The promoter strength was analyzed similar to as descrived in Panel "A" except that xylose was used as carbon source instead of dextrose. **(C)** The eGFP expression from indicated promoters. Cells were grown at 30 °C in media containing dextrose or xylose as carbon source. The endogenous Pgk1 showed lower expression when cells were grown in xylose containing media.

cells growth, and seen in Fig. S2, no significant growth defect was observed upon eGFP overexpression from $P_{kmIMTCP2}$-$T_{KmIMTT1}$ cassette. Overall, the designed sets of promoters and terminators showed significantly higher expression than any of the other native promoters used for protein expression in *K. marxianus*.

3.6. The cassette $P_{KmIMTCP2}$-$T_{KmIMTT1}$ shows relatively higher expression of luciferase gene

We further examined the effect of $T_{KmIMTT1}$/$T_{KmIMTT2}$ terminator sequences on the expression of gene encoding firefly luciferase driven by $P_{KmIMTCP1}$ and $P_{KmIMTCP2}$ promoters. The luciferase encoding gene was subcloned in-between 5′ and 3′ region of desired promoter and terminator respectively. The plasmid encoding from 5′ to 3' direction, the desired promoter, gene encoding firefly luciferase, and desired terminator was transformed into *K. marxianus* strain. The transformants were examined for in vivo luciferase activity by monitoring increase in luminescence as a measure of luciferase expression. As seen in Fig. 5, when cells were grown at 30 °C (Fig. 5A) or 45 °C (Fig. 5B) in dextrose containing liquid growth media, varying level of luciferase activity was observed from different combinations of promoters and terminators. The strongest luminescence signal was obtained when luciferase was expressed from $P_{KmIMTCP2}$ promoter in combination with $T_{KmIMTT1}$ or $T_{KmIMTT2}$ terminator sequences, suggesting that among the different combinations, $P_{KmIMTCP2}$-$T_{KmIMTT1}$ or $P_{KmIMTCP2}$-$T_{KmIMTT2}$ drive the strongest expression of the protein. Similarly, there was ~1.7 fold increase in luminescence when the luciferase was expressed from $P_{KmIMTCP1}$-$T_{KmIMTT1}$ or $P_{KmIMTCP2}$-$T_{KmIMTT1}$ as compared to $P_{KmIMTCP1}$-T_{ScCYC1} or $P_{KmIMTCP2}$-T_{ScCYC1}. The data shows that the order of promoter-terminator combination strength with regard to their ability to drive luciferase expression is $P_{KmIMTCP2}$-$T_{KmIMTT1}$ ~ $P_{KmIMTCP2}$-$T_{KmIMTT2}$ > $P_{KmIMTCP1}$-$T_{KmIMTT1}$ ~ $P_{KmIMTCP1}$-$T_{KmIMTT2}$ > $P_{KmIMTCP2}$-T_{ScCYC1} ~ $P_{KmIMTCP1}$-T_{ScCYC1} > P_{KmTDH3}-T_{ScCYC1} >P_{KmPGK1}-T_{ScCYC1}.

In order to explore whether the relative strength of above DNA cassette with identified promoters and terminators is affected by a different carbon source, we grew cells harbouring plasmid encoding these cassettes in media containing xylose instead of dextrose as source of carbon. The luciferase activity was measured as described above. Similar

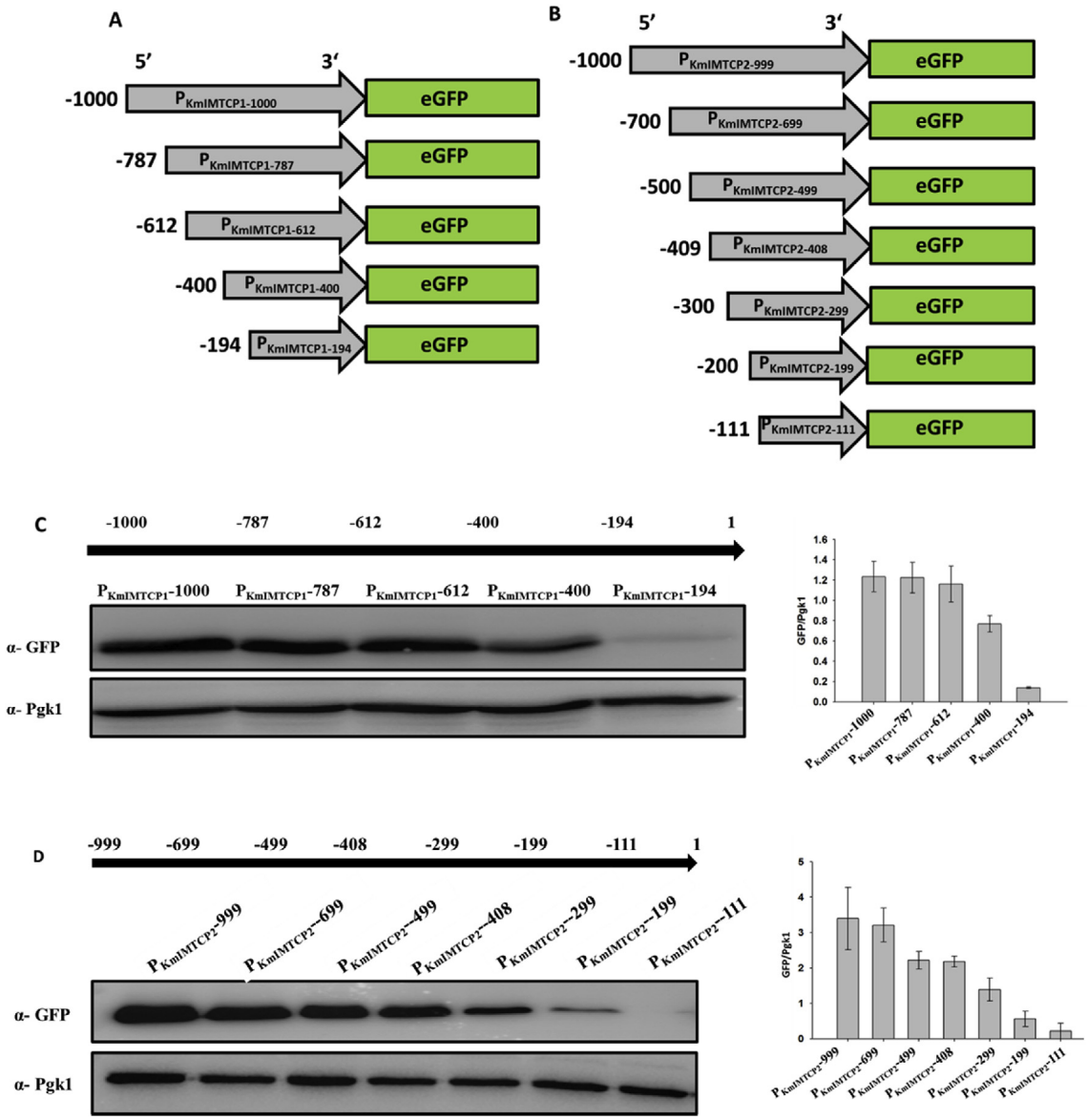

Fig. 3. eGFP expression varies with the promoter length. (A and B) Schematics of designed derivatives of $P_{KmIMTCP1}$ and $P_{KmIMTCP2}$. **(C)** Cells were transformed with plasmid encoding varying length of $P_{KmIMTCP1}$ promoter. Transformants were grown in SD liquid media, and cellular lysate was immunoblotted with anti-GFP antibody. Pgk1 was probed as loading control. **(D)** Cells were examined for eGFP expression using method similar to as mentioned above in Panel "C". The eGFP was expressed under indicated derivatives of $P_{KmIMTCP2}$.

to as observed for dextrose containing media, when cells were grown in media containing xylose, $P_{KmIMTCP2}$-$T_{KmIMTT1}$ ~ $P_{KmIMTCP2}$-$T_{KmIMTT2}$ showed highest luciferase activity followed by that shown by $P_{KmIMTCP1}$-$T_{KmIMTT1}$ ~ $P_{KmIMTCP1}$-$T_{KmIMTT2}$ (Fig. 5C). Further, the identified promoters, $P_{KmIMTCP1}$ and $P_{KmIMTCP2}$ with any of the three terminators T_{ScCYC1}, $T_{KmIMTT1}$ or $T_{KmIMTT2}$ showed relatively higher luciferase expression than P_{KmPGK1}-T_{ScCYC1} or P_{KmTDH3}-T_{ScCYC1}.

3.7. The terminators $T_{KmIMTT1}$ and $T_{KmIMTT2}$ further improved promoter strength in S. cerevisiae

The promoters of *S. cerevisiae* origin have been successfully used to express proteins in *K. marxianus* suggesting that promoters from one yeast species could also be active in others. We thus examined the ability of identified *K. marxianus* promoter to drive protein expression in *S. cerevisiae*. Further as terminators provide an additional regulation for protein expression, we examined whether the above identified terminators affect P_{ScGPD} in *S. cerevisiae*. Various combinations of promoters and terminators (Fig. 6), were constructed as described in Table 2. Overall, we used three promoters, P_{ScGPD} from *S. cerevisiae,* $P_{KmIMTCP1}$ and $P_{KmIMTCP2}$ from *K. marxianus*; and three terminator sequences, T_{ScCYC1} from *S. cerevisiae,* $T_{KmIMTT1}$ and $T_{KmIMTT2}$ from *K. marxianus*. The firefly luciferase was used as reporter protein to examine the strength of these expression cassettes.

The pRS316 either alone or encoding any of above expression cassettes was transformed into *S. cerevisiae* strain BY4741, and transformants were examined for the luciferase activity using method similar to as used above for *K. marxianus* strain. Both $P_{KmIMTCP1}$-T_{ScCYC1} and $P_{KmIMTCP2}$-T_{ScCYC1} showed activity in *S. cerevisiae* as seen by enhanced luciferase activity as compared to cells harboring vector alone however the luciferase activity was slightly lower than when luciferase was expressed from native promoter and terminator of *S. cerevisiae* (P_{ScGPD}-T_{ScCYC1}). The substitution of T_{ScCYC1} in P_{ScGPD}-T_{ScCYC1} with $T_{KmIMTT1}$ or $T_{KmIMTT2}$ improved the promoter ability to drive luciferase expression as seen by further increase in luminescence by 1.6 and 1.8 fold respectively. Similar increase in luciferase activity was detected when

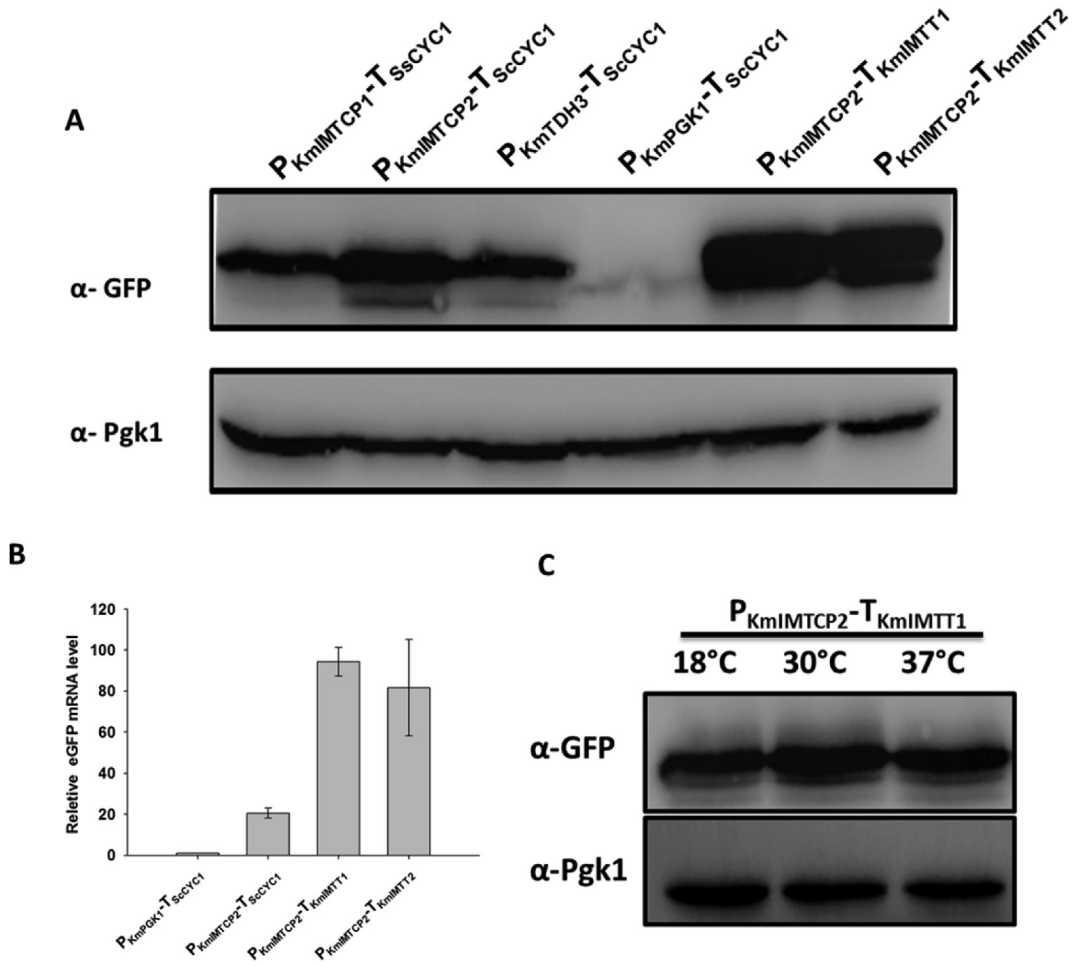

Fig. 4. The eGFP expression improves when $P_{KmIMTCP2}$ is used in combination with $T_{KmIMTT1}$ or $T_{KmIMTT2}$(A) Cells were transformed with plasmids encoding indicated combinations of promoters and terminators. Transformants were grown in SD liquid media at 30 °C, and examined for eGFP expression in cellular lysate using anti-GFP antibodies. Pgk1 level was examined as loading control. **(B)** Cells were grown as in Panel A. The mRNA was isolated, and eGFP transcript level was examined using qRT-PCR with primers specific for the genes encoding eGFP or 18 S (internal control). **(C)** The strength of indicated DNA cassette was examined by monitoring eGFP level expressed at 18 °C, 30 °C or 37 °C in *K. marxianus*.

T_{ScCYC1} was substituted with $T_{KmIMTT1}$ or $T_{KmIMTT2}$ in $P_{KmIMTCP1}$-T_{ScCYC1} and $P_{KmIMTCP2}$-T_{ScCYC1}. The cells encoding designed P_{ScGPD}-$T_{KmIMTT1}$ or P_{ScGPD}-$T_{KmIMTT2}$ showed marginally higher luminescence than those encoding $P_{KmIMTCP2}$-$T_{KmIMTT1}$ or $P_{KmIMTCP2}$-$T_{KmIMTT2}$.

3.8. *The cassette $P_{KmIMTCP2}$-$T_{KmIMTT1}$ showed higher strength than P_{scGPD}-T_{ScCYC1} for influenza HA antigen expression*

We further examined the application of above designed expression cassette from *K. marxianus* for its ability to drive expression of influenza protein Hemagglutinin (HA). The HA antigen is a potential candidate for the design of influenza vaccine. We examined HA expression driven by the designed DNA cassette $P_{KmIMTCP2}$-$T_{KmIMTT1}$ in *K. marxianus*, and compared with that produced from widely used strong constitutive promoter cassette P_{ScGPD}-T_{ScCYC1} in *S. cerevisiae*. The gene encoding His_{10}-HA was subcloned with desired combination of promoters and terminators in centromeric plasmid for expression in *K. marxianus* or *S. cerevisiae*. The cells harboring expression plasmid were grown until mid log phase, and cellular lysate was probed with anti-His tag antibodies on an immunoblot. As seen in Fig. 7, the HA expression from *K. marxianus* was found to be 2.5 fold higher than that from P_{ScGPD} promoter in *S. cerevisiae*, further suggesting that the identified promoter cassette could not only drive stronger expression of reporter proteins such as eGFP or firefly luciferase but also of other heterologous proteins.

3.9. *S. cerevisiae α-factor signal sequence efficiently secretes eGFP in K. marxianus*

We further examined the ability of the identified $P_{KmIMTCP2}$-$T_{KmIMTT1}$ cassette on expression of eGFP secreted out into extracellular growth media. The secretion of a protein into extracellular space requires the presence of secretory signal sequence that traffics the protein into media supernatant through endoplasmic reticulum (ER).

To examine the secretion of eGFP in extracellular growth media, the protein was expressed in fusion with various signal sequences under $P_{KmIMTCP2}$. The various signal sequence that are compared in the present study are; Inulinase signal sequence ($ssInu1_{Km}$) with P10L substitution of *K. marxianus*, Pho1 signal sequence ($ssPho1_{Pp}$) of *P. pastoris*, mating α-factor signal sequence ($ss\alpha MF_{Sc}$) containing D83E substitution of *S. cerevisiae* and mating α-factor signal sequence ($ss\alpha MF_{Km}$) of *K. marxianus*. The plasmid encoding eGFP in fusion with desired signal sequence was transformed in *K. marxianus*. The media supernatant was collected and examined for eGFP expression. As seen in Fig. 8A, among the various signal sequences examined, eGFP in media supernatant was significantly higher when expressed in fusion with signal sequence from mating α factor of *S. cerevisiae*.

We further examined the effect of changes in carbon source on promoter strength as well as ability of αMF_{Sc} for extracellular secretion of proteins. The DNA cassette $P_{KmIMTCP2}$-$ss\alpha MF_{Sc}$-eGFP-$T_{KmIMTT1}$ encoding fusion of eGFP with $ss\alpha MF_{Sc}$ expressed under the control of $P_{KmIMTCP2}$

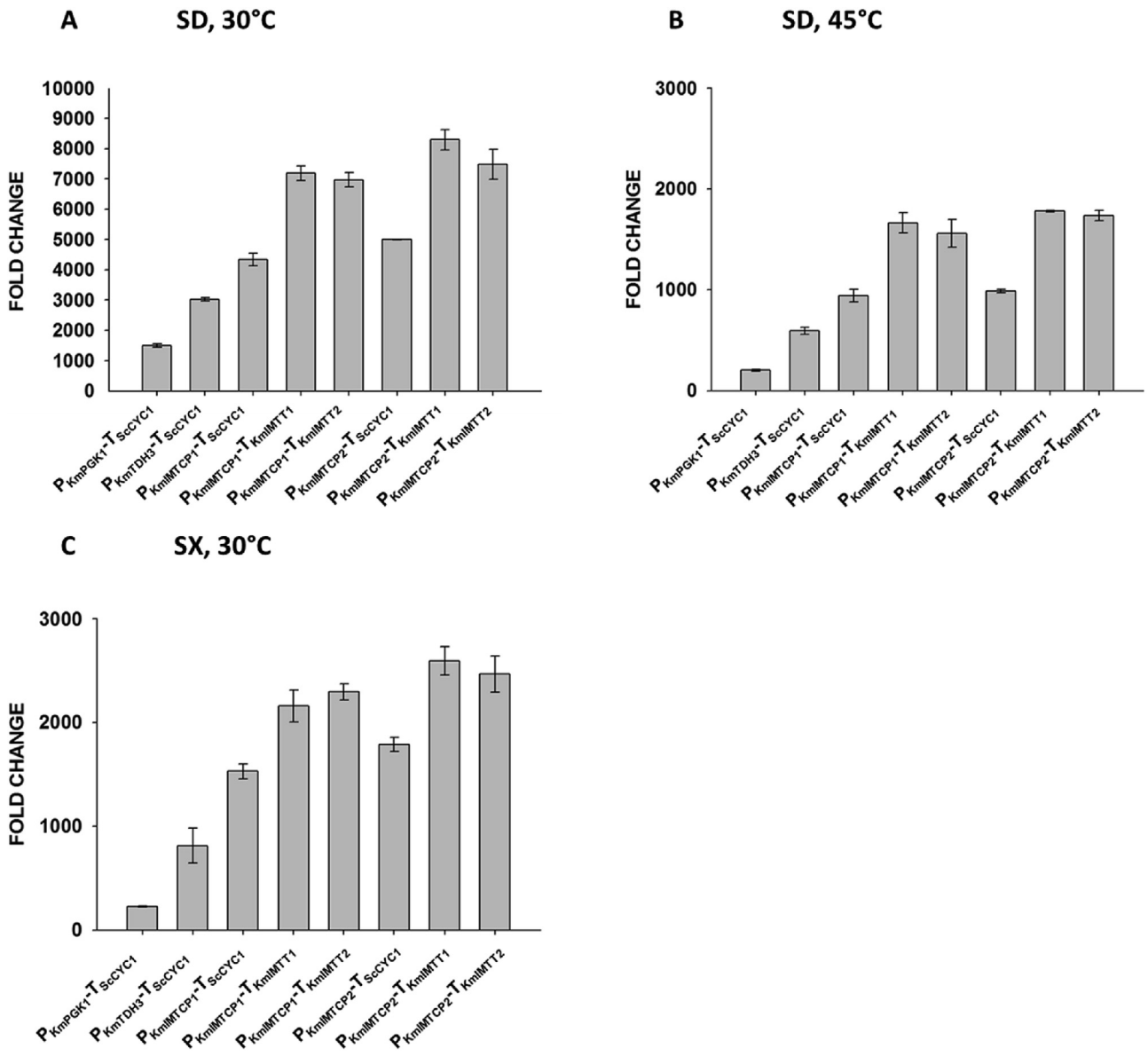

Fig. 5. The $T_{KmIMTT1}$ enhances luciferase expression. Cells were transformed with plasmid encoding luciferase under the control of different combinations of indicated promoters and terminators. Transformants were grown until OD_{600nm} of 0.8 in liquid synthetic media containing dextrose or xylose as carbon source. About 200 μl of 0.3 $O.D._{600nm}$ cells were added with D-Luciferin as luciferase substrate, and increase in luminescence was monitored. Shown is the luciferase activity from cells grown (A) at 30 °C with dextrose or (B) 45 °C with dextrose as carbon source or (C) at 30 °C with xylose as carbon source.

promoter was genomically integrated in *K. marxianus* at *HIS3* locus. The genomically integrated strain was grown in liquid growth media containing dextrose, xylose or lactose as carbon source. The eGFP expression was confirmed in extracellular growth media using immnoublot analysis. As seen in Fig. 8B, the eGFP was expressed similarly in all the three types of growth media suggesting the strength of promoter as well as the secretion signal sequence is independent of the type of carbon source used for culturing cells.

3.10. The cassette $P_{KmIMTCP2}$-$T_{KmIMTT1}$ showed higher β-galactosidase expression

The *K. marxianus* has potential usage for the enzymatic mitigation of lactose in products such as cheese whey and milk, due to its ability to produce β-galactosidase. To further improve the utility of *K. marxianus* for lactose utilization, we engineered the strain for enhanced production of β-galactosidase.

Above data show that the ability of designed $P_{KmIMTCP2}$-$T_{KmIMTT1}$ for protein production is independent of carbon sources such as dextrose, xylose or lactose. We subcloned gene encoding *K. marxianus* β-galactosidase under the $P_{KmIMTCP2}$-$T_{KmIMTT1}$ in Ura3 based plasmid. The β-galactosidase encoding plasmid was transformed into *K. marxianus*, and examined for β-galactosidase activity as mentioned in Materials and Methods. As seen in Fig. 9, wt *K. marxianus* with empty vector showed lower β-galactosidase activity of 10 miller units. The enzyme activity of the β-galactosidase produced from $P_{KmIMTCP2}$-$T_{KmIMTT1}$ showed ~1800 miller units higher activity as compared to that produced chromosomal borne gene. Further the $P_{KmIMTCP2}$-$T_{KmIMTT1}$ produced 1.6 and 24 fold higher activity than previously identified P_{KmTHD3} and P_{KmPGK1} promoters. Overall our data suggest that the *K. marxianus* engineered to produce β-galactosidase from $P_{KmIMTCP2}$-$T_{KmIMTT1}$ produces significantly higher β-galactosidase then the wt organism.

4. Discussion

For wider utilization of *K. marxianus*, there is a need to expand the

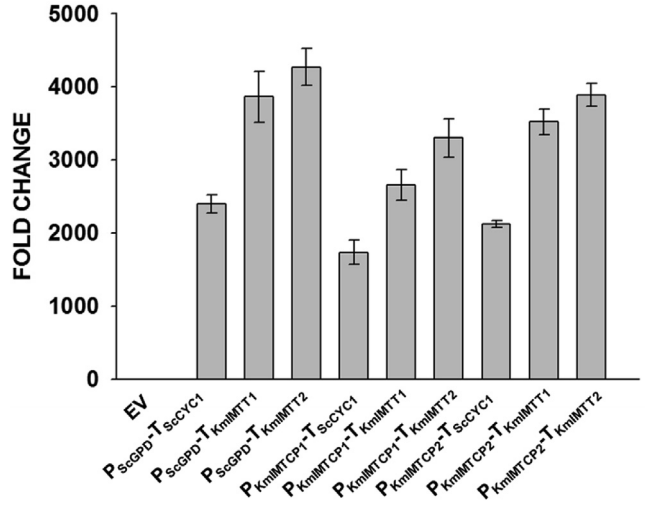

Fig. 6. Comparison of the activity of luciferase expressed using indicated combination of promoter and terminator in *S. cerevisiae*. Cells harboring plasmid encoding luciferase under different promoter and terminator combinations were grown in liquid SD media at 30 °C. The luciferase activity was monitored as described in Materials and Methods. The luciferase substrate D-luciferin was added into 200 μl of culture volume ($O.D._{600nm} = 0.3$) and increase in luminescence was monitored.

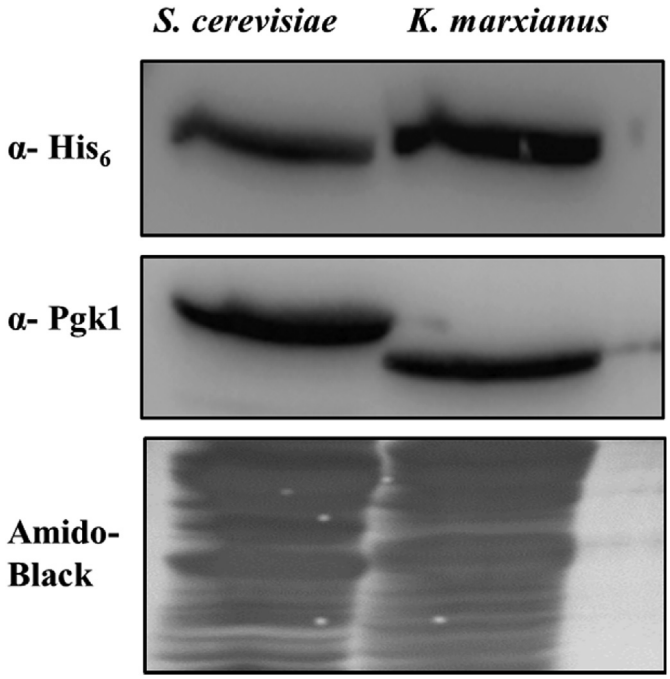

Fig. 7. Comparative expression of HA protein in different yeast strains. Immunoblot analysis of expression level of His tagged HA protein in *S. cerevisiae* and *K. marxianus*. The His_{10}-HA was expressed from strong GPD promoter in *S. cerevisiae*. The $P_{KmIMTCP2}$-$T_{KmIMTT1}$ was used to express His_{10}-HA in *K. marxianus*. Yeast cells were grown at 30 °C with dextrose as carbon source. The cells were harvested and cellular lysate was probed with anti-His_6 or anti-Pgk1 (loading control) antibody.

genetic toolbox required for various applications such as to improve the production of heterologous proteins, engineering of its glycosylation pathways and secretory capacity. Many of such synthetic biology approaches require genetic tools for genomic integration as well as expression of desired proteins in different carbon sources and temperatures. The present study provides novel sets of promoters and terminators, and further used it for metabolic engineering of *K. marxianus* for higher β-galactosidase activity.

The analysis of available transcriptomic data revealed relative strength of various promoters in *K. marxianus*. The transcript level of previously used strong promoter P_{KmPGK1} in *K. marxianus* was found to be lower than the most abundant transcripts at 30 °C. In agreement, the relative eGFP expression driven from identified top 2 promoters was found to be higher than from P_{KmPGK1} suggesting that transcriptomics could potentially provide novel strong promoters with varying level of protein expression. This is also in agreement with previous studies focused to identify novel promoters based upon level of various transcripts (Yuan et al., 2020).

The eGFP or luciferase expression from identified strong promoters $P_{KmIMTCP1}$ and $P_{KmIMTCP2}$ at 30 °C and 45 °C showed consistent higher expression than previously known P_{KmTDH3} or P_{KmPGK1} promoter suggesting that their ability to drive relatively higher protein expression is independent of alteration of growth temperature. The constitutive expression from $P_{KmIMTCP1}$ and $P_{KmIMTCP2}$ is also consistent with the reported transcriptomic data obtained under various conditions (Table S3) (Lertwattanasakul et al., 2015). As evident from Table S3, IMTCP1 and IMTCP2 have highest transcript value than the rest of the genes (IMTCP3-IMTCP10) under all four conditions (Lertwattanasakul et al., 2015). The promoter strength at higher temperature is particularly crucial in processes where cell growth is required at higher temperature such as during alcohol fermentation.

The substitution of dextrose with xylose for carbon source showed alteration of eGFP expression level suggesting that strength of some of the identified promoters are dependent upon the nature of the carbon source. The effect of change in carbon source to xylose was more pronounced for $P_{KmIMTCP6}$ and $P_{KmIMTCP9}$ than $P_{KmIMTCP1}$ or $P_{KmIMTCP2}$. It is possible that the effect of xylose on promoter strength is due to cellular adaptation of metabolic pathways for better utilization of the pentose sugars. This is in agreement with the known influence of change in carbon source on metabolic perturbation and reprograming of gene transcription (Diniz et al., 2017; Paulo et al., 2015). Interestingly, the strength of the best two identified promoters, $P_{KmIMTCP1}$ and $P_{KmIMTCP2}$ remain independent of change in the carbon source whereas the previously used P_{KmTDH3} show relatively lower expression in xylose containing media.

It is known that various regulatory elements interact with the promoter region and govern its ability to drive gene expression. The study examining DNA fragments of $P_{KmIMTCP1}$ or $P_{KmIMTCP2}$ that are still capable to drive protein expression showed that though $P_{KmIMTCP1}$ could be truncated by 388 bp without affecting its ability to regulate protein expression, full length $P_{KmIMTCP2}$ is required for maximal activity. The smaller fragment of $P_{KmIMTCP1}$ having 612 bp showing similar activity as of full length 1000bp promoter is in agreement with previous studies showing that in general 500-700bp upstream sequence is sufficient to drive protein expression (Lang et al., 2020; Lee et al., 2015; Ohler and Niemann, 2001). However since full-length $P_{KmIMTCP2}$ is required, it is possible that different regulatory factors such as transcription factors binds to the extreme 5′ end of the promoter $P_{KmIMTCP2}$ and allosterically influences RNA polymerase binding to the 3' region of the promoter. Similar upstream enhancer/regulatory sequences has been observed before for promoters such POX2 in *Y. lipolytica* (Shabbir Hussain et al., 2016).

The $P_{KmIMTCP1}$ or $P_{KmIMTCP2}$ were found to be strongest among the selected 10 promoters. Promoters generally consist of conserved functional elements such as a core promoter region (consisting of TATA box and the transcriptional start site) and upstream regulatory sequences. In general, these regions lie within first 200bp upstream sequence (Basehoar et al., 2004). We looked into the presence of TATA boxes in all 10 promoters and found TATA boxes in less than 250bp upstream of translation start site of all the promoters. The upstream regulatory elements of these promoters are still not defined. Thus we did not find any apparent correlation between amplified promoter regions of the remaining 8

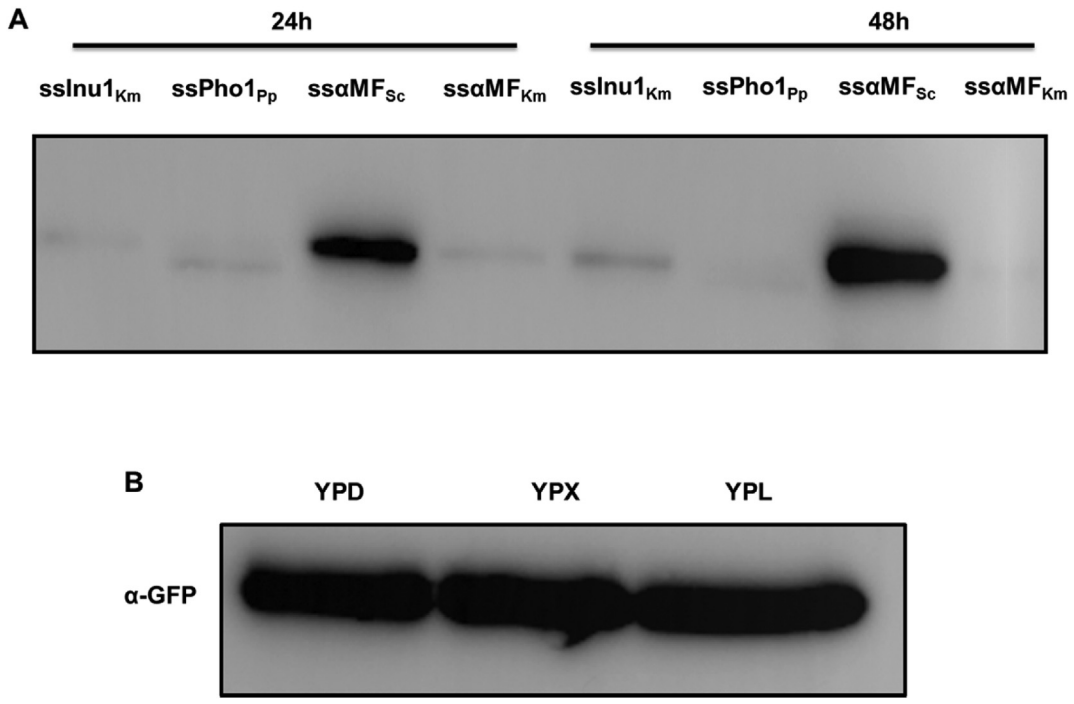

Fig. 8. Comparison of secretory capacity of various signal sequences in *K. marxianus*. (A) Cells harboring plasmid encoding eGFP fused with indicated signal sequences under $P_{KmIMTCP2}$-$T_{KmIMTT1}$ cassette was grown for 24 h and 48 h. The culture was spun at 12000g for 15 min, and media supernatant was collected, and further probed with anti-GFP antibody. The signal sequences examined are: Inulinase signal sequence ($ssInu1_{Km}$) with P10L substitution of *K. marxianus*, $ssPho1_{Pp}$ signal sequence of *P. pastoris*, mating α-factor signal sequence ($ssαMF_{Sc}$) with D83E of *S. cerevisiae* and mating α-factor signal sequence ($ssαMF_{Km}$) of *K. marxianus* **(B)** expression of extracellular eGFP expressed in fusion with $ssαMF_{Sc}$ encoded under the control of $P_{KmIMTCP2}$-$T_{KmIMTT1}$. The DNA $P_{KmIMTCP2}$-$ssαMF_{Sc}$-eGFP-$T_{KmIMTT1}$ cassette was genome integrated at HIS3 locus in *K. marxianus* genome. Cells were grown in media containing dextrose (YPD), xylose (YPX) and lactose (YPL). The eGFP expression was examined as above in Panel (A).

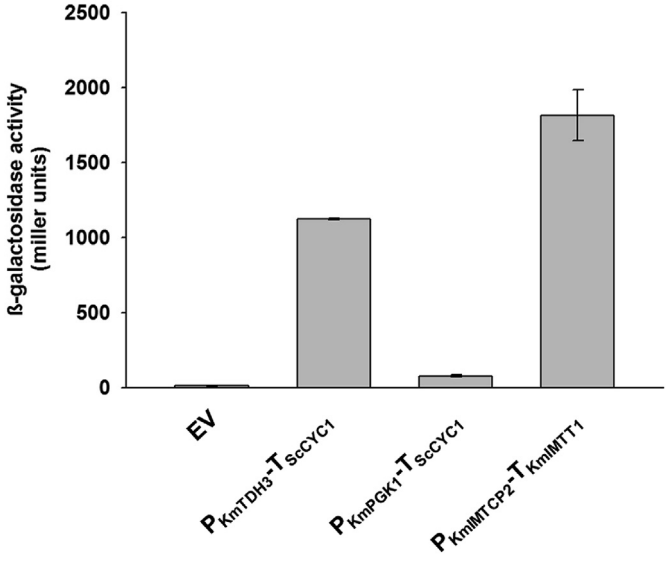

Fig. 9. The engineered *K. marxianus* with β-galactosidase encoded gene under $P_{KmIMTCP2}$-$T_{KmIMTT1}$ shows significantly higher enzyme activity. Cells were transformed with plasmid encoding β-Galactosidase under the control of different combinations of indicated promoters and terminators. The β-galactosidase activity was measured as described in Materials and Methods. As seen, the engineered *K. marxianus* cells that express β-galactosidase from combination of $P_{KmIMTCPP2}$-$T_{KmIMTT1}$ show significantly higher enzyme activity than those that express from chromosomally encoded enzyme.

promoters and their low activities. A detailed study to identify the upstream regulatory elements and their role in promoter strengths would be required to further understand the underlying basis of varying activities of these identified promoters.

The further increase in the ability of $P_{KmIMTCP1}$ and $P_{KmIMTCP2}$ to drive protein expression upon substitution of T_{ScCYC1} with $T_{KmIMTT1}$ or $T_{KmIMTT2}$ indicates that the protein abundance not only depends upon the promoter but also upon terminator which is in agreement with previous studies. The increase in protein expression with identified terminators is not only specific for $P_{KmIMTCP1}$ or $P_{KmIMTCP2}$ but more general as they also enhanced promoter strength of other promoter such as P_{ScGPD}. The enhanced expression is not specific to eGFP but is more general for other proteins such as luciferase. As the effect is not specific to a particular promoter or substrate, it is possible that the ability of the identified terminators to enhance protein expression could be related to enhanced mRNA stability. This is in agreement with increase in transcript level of eGFP when $T_{KmIMTT1}$ or $T_{KmIMTT2}$ were used as terminator sequence.

Our data show that the identified promoter cassette $P_{KmIMTCP2}$-$T_{KmIMTT1}$ functions efficiently at wide range of sugars as carbon source such as dextrose, xylose and lactose. The metabolically engineered *K. marxianus* harboring this cassette showing more than 181 fold higher β-galactosidase activity as compared to the strain expressing from chromosomally encoded enzyme, further enhances the wider utility of the strain in applications requiring lactose degradation such as from cheese whey and milk.

The ability to metabolically engineer microorganism has greatly enhanced their potential for diverse applications. *K. marxianus* provides a unique opportunity in various biotechnological applications and thus number of ongoing studies are focused to develop novel genetic tools and to understand more about its biology such as its glycosylation pathway and metabolomics. The present study has now provided various DNA cassettes consisting of novel strong promoters and terminators and further showed the application of these tools in metabolic engineering of *K. marxianus* for enhanced β-galactosidase activity. We also show that these tools could be efficiently used for the production of potential therapeutic agents such as influenza antigen HA. The promoters and

terminators as well as the engineered *K. marxianus* strain identified in the present study has enhanced the potential of *K. marxianus* for various research, industrial and biotherapeutic applications.

5. Conclusions

In the present study, we have identified novel strong native promoters of *K. marxianus*. The ability of identified promoters for protein production was further enhanced by novel sets of terminators. The identified promoters showed similar strength at wider range of temperatures and different carbon sugars. The identified promoters and terminators were also found to be active in *S. cerevisiae*. Further, we used the identified promoter and terminator DNA cassette to engineer *K. marxianus* for increased production of β-galactosidase. The engineered strain would have wider applications requiring mitigation of lactose such as in cheese whey and milk. Thus, the designed DNA cassettes of novel promoters and terminators has increased the genetic tool box for basic research, protein production and metabolic engineering applications in *K. marxianus*.

Author contributions

P.K., D.K.S. and D.S. designed research. P.K. performed research. D.S., and D.K.S., arranged for reagents and tools. P.K., D.K.S. and D.S.analyzed results. P.K., D.K.S. and D.S. wrote manuscript.

Financial disclosure

The work is supported by the Council of Scientific and Industrial Research (CSIR), India, and the Ramalingaswami Re-entry Fellowship (BT/RLF/RE-entry/33/2010) from the Department of Biotechnology (DBT), India to Deepak Sharma. The current study was also supported by funds received from BIRAC, Department of Biotechnology, India.

Additional information

A patent on part of this study has been filed: "VECTOR FOR HIGH-EXPRESSION OF PROTEINS IN YEAST" Application No. PCT/IN 2020/050640.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank Prof. Raghavan Varadarajan, IISc, Bangalore, India for kind gift of plasmid encoding HA. We thank Ms. Laxmi Swami for help in standardization of transformation protocol and Dr. Anubhav for help in plasmid design. Deepak Sharma acknowledges funding support from Council of Scientific and Industrial Research (CSIR), India, Biotechnology Industry Research Assistance Council, India and the Ramalingaswami Re-entry Fellowship from Department of Biotechnology (DBT) India.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.mec.2020.e00160.

References

Abdel-Banat, B.M., Nonklang, S., Hoshida, H., Akada, R., 2010. Random and targeted gene integrations through the control of non-homologous end joining in the yeast Kluyveromyces marxianus. Yeast 27, 29–39.

Akada, R., Hoshida, H., Ide, M., 2014. High-expression Promoter Derived from Kluyveromyces Marxianus (Google Patents).

Almeida, C., Queirós, O., Wheals, A., Teixeira, J., Moradas-Ferreira, P., 2003. Acquisition of flocculation phenotype by Kluyveromyces marxianus when overexpressing GAP1 gene encoding an isoform of glyceraldehyde-3-phosphate dehydrogenase. J. Microbiol. Methods 55, 433–440.

Ball, M.M., Raynal, A., Guérineau, M., Iborra, F., 1999. Construction of efficient centromeric, multicopy and expression vectors for the yeast Kluyveromyces marxianus using homologous elements and the promoter of a purine-cytosine-like permease. J. Mol. Microbiol. Biotechnol. 1, 347–353.

Banat, I.M., Nigam, P., Marchant, R., 1992. Isolation of thermotolerant, fermentative yeasts growing at 52 C and producing ethanol at 45 C and 50 C. World J. Microbiol. Biotechnol. 8, 259–263.

Bansal, S., Oberoi, H.S., Dhillon, G.S., Patil, R., 2008. Production of β-galactosidase by Kluyveromyces marxianus MTCC 1388 using whey and effect of four different methods of enzyme extraction on β-galactosidase activity. Indian J. Microbiol. 48, 337–341.

Basehoar, A.D., Zanton, S.J., Pugh, B.F., 2004. Identification and distinct regulation of yeast TATA box-containing genes. Cell 116, 699–709.

Belem, M., Gibbs, B., Lee, B., 1999. Proposing sequences for peptides derived from whey fermentation with potential bioactive sites. J. Dairy Sci. 82, 486–493.

Bergkamp, R.J., Bootsman, T.C., Toschka, H.Y., Mooren, A.T., Kox, L., Verbakel, J.M., Geerse, R.H., Planta, R.J., 1993. Expression of an α-galactosidase gene under control of the homologous inulinase promoter in Kluyveromyces marxianus. Appl. Microbiol. Biotechnol. 40, 309–317.

Chakraborty, R., Baek, J.H., Bae, E.Y., Kim, W.-Y., Lee, S.Y., Kim, M.G., 2016. Comparison and contrast of plant, yeast, and mammalian ER stress and UPR. Applied Biological Chemistry 59, 337–347.

Cheng, C.-C., Yu, M.-C., Cheng, T.-C., Sheu, D.-C., Duan, K.-J., Tai, W.-L., 2006. Production of high-content galacto-oligosaccharide by enzyme catalysis and fermentation with Kluyveromyces marxianus. Biotechnol. Lett. 28, 793–797.

Curran, K.A., Karim, A.S., Gupta, A., Alper, H.S., 2013. Use of expression-enhancing terminators in Saccharomyces cerevisiae to increase mRNA half-life and improve gene expression control for metabolic engineering applications. Metab. Eng. 19, 88–97.

Diniz, R.H.S., Villada, J.C., Alvim, M.C.T., Vidigal, P.M.P., Vieira, N.M., Lamas-Maceiras, M., Cerdán, M.E., González-Siso, M.-I., Lahtvee, P.-J., da Silveira, W.B., 2017. Transcriptome analysis of the thermotolerant yeast Kluyveromyces marxianus CCT 7735 under ethanol stress. Appl. Microbiol. Biotechnol. 101, 6969–6980.

Erijman, A., Dantes, A., Bernheim, R., Shifman, J.M., Peleg, Y., 2011. Transfer-PCR (TPCR): a highway for DNA cloning and protein engineering. J. Struct. Biol. 175, 171–177.

Fonseca, G.G., de Carvalho, N.M.B., Gombert, A.K., 2013. Growth of the yeast Kluyveromyces marxianus CBS 6556 on different sugar combinations as sole carbon and energy source. Appl. Microbiol. Biotechnol. 97, 5055–5067.

Groeneveld, P., Stouthamer, A.H., Westerhoff, H.V., 2009. Super life–how and why 'cell selection'leads to the fastest-growing eukaryote. FEBS J. 276, 254–270.

Helden, J.v., Olmo, M.l.d., Pérez-Ortín, J.E., 2000. Statistical analysis of yeast genomic downstream sequences reveals putative polyadenylation signals. Nucleic Acids Res. 28, 1000–1010.

Ho, B., Baryshnikova, A., Brown, G.W., 2018. Unification of protein abundance datasets yields a quantitative Saccharomyces cerevisiae proteome. Cell systems 6, 192–205 e193.

Karim, A., Gerliani, N., Aïder, M., 2020. Kluyveromyces marxianus: an emerging yeast cell factory for applications in food and biotechnology. Int. J. Food Microbiol. 108818.

Lane, M.M., Morrissey, J.P., 2010. Kluyveromyces marxianus: a yeast emerging from its sister's shadow. Fungal Biology Reviews 24, 17–26.

Lang, X., Besada-Lombana, P.B., Li, M., Da Silva, N.A., Wheeldon, I., 2020. Developing a broad-range promoter set for metabolic engineering in the thermotolerant yeast Kluyveromyces marxianus. Metabolic engineering communications 11, e00145.

Lee, K.-S., Kim, J.-S., Heo, P., Yang, T.-J., Sung, Y.-J., Cheon, Y., Koo, H.M., Yu, B.J., Seo, J.-H., Jin, Y.-S., 2013. Characterization of Saccharomyces cerevisiae promoters for heterologous gene expression in Kluyveromyces marxianus. Appl. Microbiol. Biotechnol. 97, 2029–2041.

Lee, M.E., DeLoache, W.C., Cervantes, B., Dueber, J.E., 2015. A highly characterized yeast toolkit for modular, multipart assembly. ACS Synth. Biol. 4, 975–986.

Lertwattanasakul, N., Kosaka, T., Hosoyama, A., Suzuki, Y., Rodrussamee, N., Matsutani, M., Murata, M., Fujimoto, N., Tsuchikane, K., Limtong, S., 2015. Genetic basis of the highly efficient yeast Kluyveromyces marxianus: complete genome sequence and transcriptome analyses. Biotechnol. Biofuels 8, 47.

Lian, J., Mishra, S., Zhao, H., 2018. Recent advances in metabolic engineering of Saccharomyces cerevisiae: new tools and their applications. Metab. Eng. 50, 85–108.

Löbs, A.-K., Engel, R., Schwartz, C., Flores, A., Wheeldon, I., 2017. CRISPR–Cas9-enabled genetic disruptions for understanding ethanol and ethyl acetate biosynthesis in Kluyveromyces marxianus. Biotechnol. Biofuels 10, 164.

Löbs, A.-K., Schwartz, C., Thorwall, S., Wheeldon, I., 2018. Highly multiplexed CRISPRi repression of respiratory functions enhances mitochondrial localized ethyl acetate biosynthesis in Kluyveromyces marxianus. ACS Synth. Biol. 7, 2647–2655.

Mohammadi, S., Saberidokht, B., Subramaniam, S., Grama, A., 2015. Scope and limitations of yeast as a model organism for studying human tissue-specific pathways. BMC Syst. Biol. 9, 1–22.

Nielsen, J., Jewett, M.C., 2008. Impact of systems biology on metabolic engineering of Saccharomyces cerevisiae. FEMS Yeast Res. 8, 122–131.

Nonklang, S., Ano, A., Abdel-Banat, B.M., Saito, Y., Hoshida, H., Akada, R., 2009. Construction of flocculent Kluyveromyces marxianus strains suitable for high-temperature ethanol fermentation. Biosc. Biotech. Biochem. 0904101433-0904101433.

Ohler, U., Niemann, H., 2001. Identification and analysis of eukaryotic promoters: recent computational approaches. Trends Genet. 17, 56–60.
Paulo, J.A., O'Connell, J.D., Gaun, A., Gygi, S.P., 2015. Proteome-wide quantitative multiplexed profiling of protein expression: carbon-source dependency in Saccharomyces cerevisiae. Mol. Biol. Cell 26, 4063–4074.
Pecota, D.C., Rajgarhia, V., Da Silva, N.A., 2007. Sequential gene integration for the engineering of Kluyveromyces marxianus. J. Biotechnol. 127, 408–416.
Rajkumar, A.S., Varela, J.A., Juergens, H., Daran, J.-M.G., Morrissey, J.P., 2019. Biological parts for Kluyveromyces marxianus synthetic biology. Frontiers in bioengineering and biotechnology 7, 97.
Ramírez-Zavala, B., Mercado-Flores, Y., Hernández-Rodríguez, C., Villa-Tanaca, L., 2004. Purification and characterization of a lysine aminopeptidase from Kluyveromyces marxianus. FEMS Microbiol. Lett. 235, 369–375.
Rodrussamee, N., Lertwattanasakul, N., Hirata, K., Limtong, S., Kosaka, T., Yamada, M., 2011. Growth and ethanol fermentation ability on hexose and pentose sugars and glucose effect under various conditions in thermotolerant yeast Kluyveromyces marxianus. Appl. Microbiol. Biotechnol. 90, 1573.
Shabbir Hussain, M., Gambill, L., Smith, S., Blenner, M.A., 2016. Engineering promoter architecture in oleaginous yeast Yarrowia lipolytica. ACS Synth. Biol. 5, 213–223.
Yanase, S., Hasunuma, T., Yamada, R., Tanaka, T., Ogino, C., Fukuda, H., Kondo, A., 2010. Direct ethanol production from cellulosic materials at high temperature using the thermotolerant yeast Kluyveromyces marxianus displaying cellulolytic enzymes. Appl. Microbiol. Biotechnol. 88, 381–388.
Yang, C., Hu, S., Zhu, S., Wang, D., Gao, X., Hong, J., 2015. Characterizing yeast promoters used in Kluyveromyces marxianus. World J. Microbiol. Biotechnol. 31, 1641–1646.
Yuan, F., Li, K., Zhou, C., Zhou, H., Liu, H., Chai, H., Lu, F., Zhang, H., 2020. Identification of two novel highly inducible promoters from Bacillus licheniformis by screening transcriptomic data. Genomics 112, 1866–1871.
Zhang, J., Zhang, B., Wang, D., Gao, X., Sun, L., Hong, J., 2015. Rapid ethanol production at elevated temperatures by engineered thermotolerant Kluyveromyces marxianus via the NADP (H)-preferring xylose reductase-xylitol dehydrogenase pathway. Metab. Eng. 31, 140–152.
Zhou, H.-X., Xin, F.-H., Chi, Z., Liu, G.-L., Chi, Z.-M., 2014. Inulinase production by the yeast Kluyveromyces marxianus with the disrupted MIG1 gene and the over-expressed inulinase gene. Process Biochem. 49, 1867–1874.
Zhou, J., Zhu, P., Hu, X., Lu, H., Yu, Y., 2018. Improved secretory expression of lignocellulolytic enzymes in Kluyveromyces marxianus by promoter and signal sequence engineering. Biotechnol. Biofuels 11, 235.