Current Research in Biotechnology 6 (2023) 100147

Contents lists available at ScienceDirect

Current Research in Biotechnology

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

Strong expression of Cas9 under a new 3′-truncated *TEF*1α promoter enhances genome editing in *Yarrowia lipolytica*

Benjamin Ouellet [a,b], A.M. Abdel-Mawgoud [a,b,*]

[a] Department of Biochemistry, Microbiology and Bioinformatics, Faculty of Science and Engineering, Laval University, 1045, Ave. de la Médecine, Quebec, QC G1V 0A6, Canada

[b] Institute of Integrative Biology and Systems, Laval University, 1030, Ave. de la Médecine, Quebec, QC G1V 0A6, Canada

ARTICLE INFO

Keywords:
CRISPR-Cas9 Genome editing
3′truncation of promoters
5′-truncation of promoters, *TEF*1α
Yarrowia lipolytica
Transcription factor binding sites

ABSTRACT

The non-conventional yeast *Yarrowia lipolytica* is gaining interest in biotechnology as a workhorse for the production of proteins, lipids and other biomolecules. Site-specific genome editing is however limited in this yeast. Although, this was much improved by the recent adaptation of a CRISPR-Cas9 genome editing protocol for *Y. lipolytica* based on a tRNA-sgRNA fusion, yet, in the latter protocol, Cas9 is under the control of a synthetic hybrid promoter, pUAS1B8-*TEF*(136) that is associated with some drawbacks. This hybrid promoter contains tandem repeats are suggested to cause *in vivo* and *in vitro* inconveniences like polymerase slippage, random genetic rearrangements and cloning difficulties. Here we report a newly designed synthetic *TEF* promoter to drive Cas9 expression, p*TEF*(-41–406)-Kozak, which is a rationally 3′-truncated version of the already known 5′-truncated p*TEF*(406) promoter of *Y. lipolytica* fused to a synthetic Kozak sequence. Our comparison of the promoters' strength using hrGFP reporters and RT-qPCR showed that the synthetic p*TEF*(-41–406)-Kozak has an equivalent expression strength to that of p*TEF*(406), yet is at least 5 times stronger than the hybrid p*UAS*1B$_8$-*TEF*(136). The p*TEF*(-41–406) promoter mediated high expression of Cas9 and was not associated with any growth defects. Moreover, expression of Cas9 under p*TEF*(-41–406)-Kozak increased the gene integration efficiency by up to 40 % relative to that when Cas9 is expressed under p*UAS*1B$_8$-*TEF*(136). Both p*TEF*(-41–406) and p*UAS*1B$_8$-*TEF*(136) performed equally well as drivers of Cas9 expression with respect to gene deletion as demonstrated both on the genotypic and phenoypic levels. This is the first study conducting rational 3′-truncation in *TEF* promoter in *Y. lipolytica* based on *in silico* analysis of promoter sequence and structure. This approach of promoter engineering can be extended to the engineering of other yeast promoters to generate small-sized synthetic biology parts for convenient engineering of biological systems. This work provides a strong Cas9 expression cassette for more convenient and efficient CRISPR-Cas9-mediated genome editing in *Y. lipolytica* which will facilitate harnessing the full potential of this industrial strain.

Introduction

Yarrowia lipolytica is a non-conventional yeast that is gaining interest in biotechnology because of its ability to grow on cheap carbon sources, its quick cell growth, and its capacity to produce high yields of recombinant protein and lipids (Zhao, Gu et al. 2015, Abdel-Mawgoud, Markham et al. 2018, Dobrowolski, Drzymala et al. 2019, Ouellet et al. 2023). One of the factors, however, limiting the full exploitation of *Y. lipolytica* in biotechnology is their low malleability to genetic manipulations. For example, this yeast has a poor episomal system, which limits its ability to maintain plasmids. Moreover, *Y. lipolytica* is poor in homology directed repair (HDR), yet has competent non-homologous end-joining (NHEJ) as the main DNA repair system, making HDR-based site-specific genome editing of limited efficiency in this yeast (Fournier, Abbas et al. 1993, Matsuoka, Matsubara et al. 1993, Verbeke, Beopoulos et al. 2013, Larroude et al., 2018). HDR events assisted with donor DNA with homology arms as long as 500 bp take place at rates of 2 % and 56 % in NHEJ competent and incompetent cells, respectively (Verbeke, Beopoulos et al. 2013). Although disruption of NHEJ repair systems showed to favor HR, yet this comes at the cost of an increased stress-induced DNA damages induced by UV radiation (Lustig 1999, Kretzschmar, Otto et al. 2013, Schwartz, Frogue et al. 2017, Abdel-

* Corresponding author at: Department of Biochemistry, microbiology and bioinformatics, Faculty of Science and Engineering, Laval University, 1045, Ave. de la Médecine, Quebec, QC G1V 0A6, Canada.
E-mail address: ahmad.saleh@bcm.ulaval.ca (A.M. Abdel-Mawgoud).

https://doi.org/10.1016/j.crbiot.2023.100147
Received 3 July 2023; Received in revised form 9 September 2023; Accepted 28 September 2023
Available online 29 September 2023

Mawgoud and Stephanopoulos 2020).

The development of a CRISPR-Cas9 protocol to *Y. lipolytica* significantly enhanced HDR-based site-specific genome editing efficiency to rates up to 50 % and 73 % in NHEJ-competent and NHEJ-defective strains, respectively (Holkenbrink et al., 2018; Schwartz et al., 2016, 2017a,b).

Although it significantly improved genome editing efficiency in *Y. lipolytica*, yet much room was there for improvement, as about 60 % of genomic targets remained non-editable at all with such CRISPR-Cas9 protocol (Schwartz, Shabbir-Hussain et al. 2016, Abdel-Mawgoud and Stephanopoulos 2020). Further improvement of genome editing efficiency, especially for gene knock outs, was made by the work of Abdel-Mawgoud & Stephanopoulos (2020) where they optimized the sgRNA expression cassette whose design was modified such that to avoid potential secondary structure anomalies of the mature sgRNA impacting its function. This latter increased genome editing efficiencies of chromosomal gene knock in and out by at least two and 10 folds, respectively (Abdel-Mawgoud and Stephanopoulos 2020).

Another, yet less used strategy, to enhance genome editing efficiency by CRISPR-Cas9 is by boosting Cas9 expression. For instance, different strains of rice cells (calli) with different Cas9 expression strengths, induced by varying Cas9 promoters or by limiting Cas9 degradation using ubiquitin-associated domain, demonstrated a positive correlation between the level of Cas9 expression and mutation frequencies (Mikami et al., 2015a,b; Zheng et al., 2020), where a 6-fold increase of Cas9 correlated with at least 2.5 times higher genome editing efficiency. Similarly, using an intron-containing Cas9 under RPS5a promoter that was 3.75 times more expressed than the intron-less Cas9 resulted in 72 % mutation frequencies compared to none in *Arabidopsis* (Grützner et al., 2021). That Cas9 under this same strong promoter, p*RPS*5a-Cas9$_{intron}$, proved to be associated with high efficiency genome editing in other plants, e.g. *Catharanthus roseus* (100 %), *Nicotiana benthamiana* (100 %), and bacteria as *Agrobacterium* (50 % efficiency). In yeasts, strong promoters as p*TEF*1 paired with a Cas9$_{intron}$ in *Cryptococcus neoformans* (Huang et al., 2022), a bidirectional p*THX*1 in *Pichia pastoris* (Vogl et al., 2018; Weninger et al., 2016) and a p*TDH*3 in *Ogataea polymorpha* (Boisramé and Neuvéglise, 2022; Numamoto et al., 2017; Peng et al., 2015) significantly increased CRISPR-Cas9 genome editing efficiency.

Currently, no such investigation of the impact of modulation of Cas9 expression on CRISPR-Cas9 mediated genome editing efficiency was conducted in *Y. lipolytica*. In this study, we aimed at engineering a strong promoter that boosts Cas9 expression to further enhance the genome editing efficiency in *Y. lipolytica*.

The Cas9 promoter currently used in CRISPR-Cas9 protocols for *Y. lipolytica* is a hybrid promoter p*UAS*1B$_8$-*TEF*(136) previously reported as strong synthetic promoter (Abdel-Mawgoud and Stephanopoulos, 2020; Borsenberger et al., 2018; Holkenbrink et al., 2018; Larroude et al., 2020; Schwartz et al., 2016, 2017b). This p*UAS*1B$_8$-*TEF*(136) has a total size of 1047 bp and consists of eight tandem repeats of an upstream activating sequence (*UAS*) of *XPR*2 (Blanchin-Roland et al., 1994), UAS1B$_{XPR2}$ (105 bp) fused to a minimal core (-136 bp) of the strong Translation Elongation Factor 1-α (p*TEF*1α) promoter of *Y. lipolytica* (Blazeck et al., 2011; Shabbir Hussain et al., 2016; Trassaert et al., 2017). It has been reported that p*UAS1B*$_8$-*TEF*(136) hybrid promoter is associated with 30 % increase of gene expression relative to the complete native p*TEF*1α promoter (~1000 bp) as estimated in terms of fluorescence values of fused hrGFP reporters (Blazeck et al., 2011). p*UAS1B*$_8$-*TEF*(136) was used for Cas9 expression although the latter study reported a 4-fold stronger yet longer (~2,000 bp) version of that promoter, containing more *UAS* repeats and less 5′-truncation, p*UAS*1B$_{16}$-*TEF*(272) (Blazeck et al., 2011). The latter promoter was not used for Cas9 expression, probably because of technical difficulties encountered during cloning such a long promoter with that high content of repeats.

The level of expression and genome-editing efficiency of Cas9 expressed under synthetic hybrid p*UAS1B*$_8$-*TEF*(136) was compared to that under native p*TEF*1α promoter in *Y. lipolytica* (Borsenberger et al., 2018). This team reported that Cas9 was 2 to 3 folds more expressed under p*UAS1B*$_8$-*TEF*(136) compared to native p*TEF*1α promoter as assessed by Western blot, yet, the two promoters were equal in terms of genome editing efficiency (Borsenberger et al., 2018). In addition to a lack of statistical significance in Cas9 expression assessments in the latter study, Cas9 expression was assessed when expressed from plasmids (Borsenberger et al., 2018) which are known to be unstably maintained in *Y. lipolytica* (Fournier et al., 1993; Larroude et al., 2018; Matsuoka et al., 1993; Verbeke et al., 2013); this sheds doubts on the reliability of these results.

Although they are reported to amplify gene expression as demonstrated by the work of Blazeck et al. (2011), tandem *UAS*1B$_8$ repeats of p*UAS1B*$_8$-*TEF*(136) are hard to clone as they might cause polymerase slippage, or block polymerase chain reaction (Bzymek and Lovett, 2001; Hommelsheim et al., 2014; Larroude et al., 2018). This problem makes this promoter not in alignment with the criteria of synthetic biology parts which is to be easily manipulable and modulable. Moreover, *UAS* promoter activity is subject to variations depending on the environmental conditions, like pH and nitrogen level; this makes this promoters less constitutive and less robust (Shabbir Hussain et al., 2016).

With respect to the *TEF* element of the hybrid p*UAS1B*$_8$-*TEF*(136) promoter used to drive Cas9 expression (Schwartz, Shabbir-Hussain et al. 2016), it was one of the average-strength promoters among the series of native p*TEF*1α with variable 5′-truncations that were compared when used alone or fused to *UAS*1B$_8$, namely p*TEF*(136), p*TEF*(203), p*TEF*(272), p*TEF*(404), p*TEF*(504), p*TEF*(604), p*TEF*(804) and the complete *TEF*(1004) (Blazeck et al., 2011). The latter study showed that truncated p*TEF*(404) was equally strong as complete p*TEF*(1004) and that 5′-truncations downstream to the −404 bp were associated with significant reduction of p*TEF*1α promoter activity when used alone without UAS fusions (Blazeck et al., 2011). This entails that the region from −1 to −404 bp relative to the start codon of *TEF*1α might constitute the minimal core promoter of p*TEF*1α.

Since 5′ truncation was extensively explored on the p*TEF*1α leading to the core 404 bp core promoter, we thought that rational 3′ truncation of p*TEF*1α could further reduce the promoter size to obtain a similarly efficient, yet smaller promoter. The resulting promoter was intended to be used to drive protein expression including Cas9 in CRISPR-Cas9-mediated genome editing.

The impact of such 3′ truncations of p*TEF*1α was not investigated in yeasts in general; where *Saccharomyces cerevisiae* and *Y. lipolytica* are not exceptions. Such type of 3′ truncation was explored however in *CYR*1 (Hong et al., 2018) and *GPD1* (Ding et al., 2013) promoters in *S. cerevisiae* which amongst others had little to no effects on promoters' strength with truncations up to 60 pb. This inspired us to conduct rational 3′ truncation so that to reduce the size of p*TEF*1α while not jeopardizing its promoter activity.

In this study, we designed a novel synthetic promoter based on p*TEF*1α, having a 3′ truncations of 40 bp, such that it retains the 5' UTR region spanning from −41 to −406, thus named p*TEF*(-41–406). We compared the promoter activity of p*TEF*(-41–406) with those of p*TEF* (406) and p*UAS*1B$_8$-*TEF*(136) using fluorescence reporters and RT-qPCR. Moreover, we compared the performance of the three promoters on genome editing efficiency when used as drivers of Cas9 expression. The eliminated 40 bp region at the 3′-end of *Y. lipolytica*'s native p*TEF*1α showed to harbor less essential in transcription factor binding sites of native p*TEF*1α, whereas the retained region, from −41 to −406, harbors the essential promoter features as the TATA element, the Transcription Start Site (TSS), and the p*TEF* proximal motifs (Shabbir Hussain et al., 2016).

Material and methods

Strains and growth conditions

The respective single auxotroph and double auxotroph PO1g (*Mat*A, *leu*2-270, *ura*3-302::*URA*3, *xpr*2-322, *axp-2*) (Madzak et al., 2000) and PO1f (*Mat*A, *leu*2-270, *ura*3-302, *xpr*2-322, *axp-2*) (Madzak 2021) strains of *Yarrowia lipolytica* and *Escherichia coli* DH5*α* were used in this study.

E. coli DH5*α* was used for cloning and plasmid replication purposes and the strains were grown in test tubes containing Lennox Broth (LB; BioBasic, Canada) supplemented with carbenicillin 50 μg/mL (Fisher Scientific, Canada) for plasmid maintenance in a rotary drum at 37 °C.

Y. lipolytica were grown from −80 °C stocks in precultures composed of 3 mL of Yeast-Peptone-Dextrose broth (YPD; BioBasic, Canada) in 15 mL-test tubes placed in a rotary drum at 70 rpm and incubated at 28 °C. *Y. lipolytica* PO1g and PO1f transformants were selected on synthetic minimal solid media: Yeast Nitrogen Base without amino acids 1.7 g/L ($YNB_{w/o.aa}$; Bio Basic), glucose 20 g/L (BioBasic), ammonium chloride 5 g/L (BioBasic), uracil 0.076 g/L (BioBasic), agar 15 g/L (BioBasic) and phosphate buffer 25 mM, pH 6.8. Plates were incubated for at least 2 days at 28 °C until colonies were well developed. Minimal media were supplemented of leucine 0.3 g/L (BioBasic) or both leucine and uracil 0.076 g/L (BioBasic) when appropriate for PO1g and PO1f growth respectively.

Homologous recombination (URA^+) and/or Cas9 (LEU^+) plasmids were cured from positive transformants by growing isolated colonies in YPD broth supplemented (only for URA^+ plasmids) with 5-fluoroorotic acid (5-FOA; 1 mg/mL), and incubated at 30 °C for 1–2 days. Complete curing of plasmids was achieved after 2 to 3 subcultures of subclones on YPD supplemented with 5-FOA if necessary. Subcultures were then streaked on YPD agar and isolated colonies were further streaked on both $YNB_{w/o.aa}$-Ura^+-Leu^- and/or $YNB_{w/o.aa}$-Ura^--Leu^+ to verify loss of URA^+ and/or LEU^+ plasmids. Colony PCR was conducted on cured subclones to verify the genotype.

Phenotypic selection of mutants

Transformants carrying different gene knockouts, namely, *gsy*1, *ade*2, *can*1 and *lip*2, were phenotypically screened on specific solid media that result in visible colonial changes upon successful inactivation of respective genes. Colonies of *GSY*1 transformants were picked from YNB selective medium, spotted on solid YPD medium using sterile toothpicks, incubated for 48 h at 28 °C, and sprayed with Lugol's solution (1:1 mixture of 2 % KI (BioBasic) and 1 % I_2 (BioBasic)) based on the protocol of Larroude (2020). Positive clones (Gsy^-, *gsy*1) were distinguished from negative ones (Gsy^+, *GSY*1) by the pale and brown coloration of peripheries of their colonies, respectively. Colonies of *ADE*2 transformants were transferred from YNB selective medium supplemented with adenine (BioBasic) at 800 μg/mL, according to Verbeke et al. (2013) with some modifications, onto a solid YPD medium and incubated for at least 5 days at 28 °C. Positive clones ($Ade2^-$, *ade*2) were distinguished from negative ones ($Ade2^+$, *ADE*2) by their reddish and pale colonies, respectively. Colonies of *CAN*1 transformants were transferred from selective YNB medium onto a solid on synthetic minimal medium supplemented with 50 μg/mL canavanine (Sigma-Aldrich, Canada) (Schwartz, Cheng et al., 2019). Positive clones (Can^-, *can*1) were able to grow in the presence of canavanine, a toxic antimetabolite (Rosenthal 1977), whereas negative clones (Can^+, *CAN*1) die. Colonies of *LIP*2 transformants were transferred from YNB selective medium onto solid minimal synthetic medium containing tributyrin 1 % (v/v) (Fisher Scientific) and incubated for 48 h at 28 °C (Nga, Heslot et al., 1988). Tributyrin plates appears cloudy except in regions of lipase production around Lip^+ colonies where tributyrin is degraded and plate appears translucid. Since *LIP*2 mutation disrupts one of the main extracellular lipases, we can distinguish positive clones (Lip^-, *lip*2) from negative clones (Lip^+, *LIP*2) by the formation of cloudy and translucid circumferences around the colonies, respectively. For each phenotype, 20 colonies in each of the biological triplicates were screened.

Fluorescence-based assay for expression

Precultures of *Y. lipolytica* PO1f strains AS510, AS511 and AS610 (Table 1) harboring hrGFP under variants of p*TEF* promoters, namely, UAS1B_8-*TEF*(136), p*TEF*(406), and p*TEF*(-41–406) were prepared in 3 mL of YPD broth and incubated in test tubes placed in a rotary drum (70 rpm) at 28 °C. Precultures were used to inoculate fresh 5 mL of YPD broth in test tubes at a starting optical density at 600 nm (OD_{600}) of 0.1 and incubated in a rotary drum (70 rpm) at 28 °C. At time-points throughout the logarithmic and early stationary phases (5, 10.5, 24 and 29 h), cells were collected by centrifugation out of culture samples of 100 μL, washed twice in 200 μL Phosphate Buffer (PB) (25 mM, pH 6.8), and transferred onto a 96-well black plate with transparent bottom to measure both their OD_{600} (Multiskan, Thermo Fisher Scientific) and fluorescence at excitation/emission wavelengths of 485/527 nm

Table 1
List of *Y. lipolytica* strains and mutants.

Code of *Y. lipolytica* strain	Name	Genotype	Ref.
AS335	*Y. lipolytica* PO1f	*Mat*A, *xpr*2-322, *axp*-2, *leu*2-270, *ura*3-302	ATCC no. MYA-2613
AS336	*Y. lipolytica* PO1g	*Mat*A, *xpr*2-322, *axp*2, *leu*2-270, *ura*3-302::*URA*3:: pBR322	(Madzak et al., 2000)
AS510	*Y. lipolytica* PO1f, p*TEF*(-41–406)-Kozak-hrGFP	*Mat*A, *axp*-2, *xpr*2-322::p*TEF* (-41–406)-Kozak-hrGFP-*LIP2t*, *leu*2-270, *ura*3-302	This study
AS511	*Y. lipolytica* PO1f, p*UAS1B*$_8$-*TEF* (136)-hrGFP	*Mat*A, *axp*-2, *xpr*2-322:: p*UAS1B*$_8$-*TEF*(136)-hrGFP-*CYC*1*t*, *leu*2-270, *ura*3-303	This study
AS606	*Y. lipolytica* PO1g, Δ*lip2*	*Mat*A, *xpr*2-322, *axp*2, *leu*2-270, *ura*3-302::*URA*3, Δ*lip*2	This study
AS610	*Y. lipolytica* PO1f, p*TEF*(406)-Kozak-hrGFP	*Mat*A, *axp*-2, *xpr*2-322::p*TEF* (406)-Kozak-hrGFP-*LIP2t*, *leu*2-270, *ura*3-302	This study
AS611	*Y. lipolytica* PO1g, Δ*ade2*	*Mat*A, *xpr*2-322, *axp*2, *leu*2-270, *ura*3-302::*URA*3, Δ*ade*2	This study
AS613	*Y. lipolytica* PO1g, Δ*gsy1*	*Mat*A, *xpr*2-322, *axp*2, *leu*2-270, *ura*3-302::*URA*3, Δ*gsy*1	This study
AS615	*Y. lipolytica* PO1g, Δ*can1*	*Mat*A, *xpr*2-322, *axp*2, *leu*2-270, *ura*3-302::*URA*3, Δ*can*1	This study
AS639	*Y. lipolytica* PO1f, pAS235	*Mat*A, *xpr*2-322, *axp*-2, *leu*2-270, *ura*3-302, pAS235 (pCRISPRyl_*XPR2*: p*TEF* (-41–406)-Kozak-Cas9-*CYC*1*t*, *SCR*1′-tRNA-sgRNA$_{XPR2}$ wo 9nt, *CEN*1, *LEU2*, AmpR, *ColE*1)	This study
AS640	*Y. lipolytica* PO1f, pAS141	*Mat*A, *xpr*2-322, *axp*-2, *leu*2-270, *ura*3-302, pAS141 (pCRISPRyl_*XPR2*: p*UAS1B*$_8$-*TEF*(136)-Cas9-*CYC*1t, *SCR*1′-tRNA-sgRNA$_{XPR2}$ wo 9nt, *CEN*1, *LEU2*, AmpR, *ColE*1)	This study
AS643	*Y. lipolytica* PO1f, pBO001	*Mat*A, *xpr*2-322, *axp*-2, *leu*2-270, *ura*3-302, pBO001 (pCRISPRyl_*XPR2*::p*TEF* (406)-Kozak-Cas9-*CYC*1*t*, *SCR*1′-tRNA-sgRNA$_{XPR2}$ wo 9nt, *CEN*1, *LEU2*, AmpR, *ColE*1)	This study
AS667	*Y. lipolytica* PO1f, p*TEF*(406)-hrGFP	*Mat*A, *axp*-2, *xpr*2-322::p*TEF* (406)-hrGFP-*LIP2t*, *leu*2-270, *ura*3-302	This study
AS668	*Y. lipolytica* PO1f, p*TEF*(-41–406)-hrGFP	*Mat*A, *axp*-2, *xpr*2-322::p*TEF* (-41–406)-hrGFP-*LIP2t*, *leu*2-270, *ura*3-302	This study

(Fluoroskan Ascent, Thermo Fisher Scientific). Fluorescence values were subtracted from those of control cultures of PO1f strain not harboring any GFP to account for autofluorescence. Specific fluorescence was calculated by normalising the measured fluorescence per one unit of OD_{600}.

Plasmids and oligonucleotides

The plasmids and oligonucleotides used in this study are listed in Table 2 and Table 3, respectively.

DNA cloning and gene assembly

All plasmids were constructed using conventional cloning and were then transformed into *E. coli* DH5α (Maniatis et al., 1982). The CRISPR plasmids, pBO001 (p*TEF*(406)-Kozak-Cas9 with $gRNA_{XPR2}$) and pAS235 (p*TEF*(-41–406)-Kozak-Cas9 with $gRNA_{XPR2}$), were constructed from pAS141 (p*UAS1B8-TEF*(136)-Cas9 with $gRNA_{XPR2}$). The new promoters, p*TEF*(-41–406) and p*TEF*(406), were PCR-amplified using the primer pairs 481(fwd)/482(rev) and 481(fwd)/558(rev) then digested with *Tsp*MI and *Asc*I (New England Biolab) and ligated T4 DNA ligase (New England Biolab) into the *Tsp*MI and *Asc*I-linearized pAS141 backbone creating plasmids pAS235 and pBO001, respectively. Similarly, the plasmids pBO002 and pBO003 were constructed using the same steps yet using instead the pAS142 (p$UAS1B_8$-*TEF*(136)-Cas9 with $gRNA_{XPR2}$) plasmid backbone.

The CRISPR plasmids targeting *GSY*1, *ADE*2, *CAN*1 and *LIP*2 genes were constructed by ligation of respective gRNA adapters with compatible ends with the *Bsm*BI- (New England Biolab) linearized pAS164 (p*UAS1B8-TEF*(136)-Cas9 with no gRNA) and pAS242 (p*TEF* (-41–406)-Cas9 with no gRNA) plasmid backbones. The gRNA adapters of *GSY*1, *ADE*2, *LIP*2 and *CAN*1 were constructed by the thermal annealing of the oligonucleotide pairs 490/491, 492/493, 494/495 and 496/497, respectively. The pAS242 was constructed from pAS164. pAS164 was constructed by cloning the *Aat*II-*Tsp*MI fragment carrying the empty gRNA-LEU of pAS164 onto the *Aat*II-*Tsp*MI backbone (carrying the p*TEF*(-41–406)-Cas9) of pAS235. The pAS164 plasmid (having *Bsm*BI instead of *Avr*II for gRNA cloning) was constructed by Gibson

Table 2
List of plasmids and their harboring *E. coli* strains.

Code of plasmid [Code of *E. coli* Strain]	Name	Genotype	Ref.
pAS141 [AS0481]	pCRISPRyl_*XPR*2_p$UAS1B_8$-*TEF*(136)-Cas9	p$UAS1B_8$-*TEF*(136)-Cas9-*CYC*1*t*, *SCR*1′-tRNA-sgRNA$_{XPR2}$, *CEN*1, *LEU2*, AmpR, *ColE*1	(Abdel-Mawgoud and Stephanopoulos 2020)
pCRISPRyl [AS0405]	pCRISPRyl, No targeting sgRNA	pCRISPRyl: p$UAS1B_8$-*TEF*(136)-Cas9-*CYC*1*t*, *SCR*1′-tRNA-sgRNA$_{empty}$, *CEN*1, *LEU2*, AmpR, *ColE*1	Addgene #70007, (Schwartz, Shabbir-Hussain et al. 2016)
pAS157 [AS0500]	pCRISPRyl, No targeting sgRNA, *Bsm*BI deleted	pCRISPRyl: p$UAS1B_8$-*TEF*(136)-Cas9-*CYC*1*t*, *SCR*1′-tRNA-sgRNA, *CEN*1, *LEU2*, AmpR, *ColE*1, No targeting sgRNA), *Bsm*BI sites downstream to *LEU*2 deleted.	This study
pAS164 [AS0507]	pCRISPRyl-no gRNA- p$UAS1B_8$-*TEF*(136)-Cas9	p$UAS1B_8$-*TEF*(136)-Cas9-*CYC*1*t*, *SCR*1′-tRNA-sgRNA$_{empty}$, *CEN*1, *LEU2*, AmpR, *ColE*1	This study
pAS242 [AS0590]	pCRISPRyl-no gRNA-pTEF (-41–406)-Cas9	p*TEF*(-41–406)-Cas9-*CYC1t*, *SCR*1′-tRNA-sgRNA$_{empty}$, *CEN*1, *LEU2*, AmpR, *ColE*1	This study
pBO001 [AS0626]	pCRISPRyl_*XPR*2_p*TEF*(406)-Kozak-Cas9	p*TEF*(406)-Kozak-Cas9-*CYC*1*t*, *SCR*1′-tRNA-sgRNA$_{XPR2}$, *CEN*1, *LEU2*, AmpR, *ColE*1	This study
pAS235 [AS0583]	pCRISPRyl_*XPR*2_p*TEF*(-41–406)-Kozak-Cas9	p*TEF*(-41–406)-Kozak-Cas9-*CYC*1*t*, *SCR*1′-tRNA-sgRNA$_{XPR2}$, *CEN*1, *LEU2*, AmpR, *ColE*1	This study
pHR_XPR2_hrGFP [AS0365]	pHR_*XPR*2_p$UAS1B_8$-*TEF*(136)-hrGFP	0.8kb_*XPR*2_up-p$UAS1B_8$-*TEF*(136)-hrGFP-*CYC*1*t*-1kb_*XPR*2_down, *CEN*1, *URA*3, AmpR, *Co1E*1	Addgene #84614, (Schwartz, Shabbir-Hussain et al. 2017)
pAS160 [AS0503]	pHR_*XPR*2_p*TEF*(-41–406)-Kozak-hrGFP	1kb_*XPR*2_up-p*TEF*(-41–406)-Kozak-hrGFP-*LIP*2t-1kb_*XPR*2_down, *CEN*1, *URA*3, AmpR, *Co1E*1	This study
pAS190 [AS0538]	pHR_*XPR*2_p*TEF*(406)-Kozak-hrGFP	1kb_*XPR*2_up-p*TEF*(406)-Kozak-hrGFP-*LIP*2t-1kb_1 Kb *XPR*2_down, *CEN*1, *URA*3, AmpR, *Co1E*1	This study
pAS191 [AS0539]	pHR_*XPR*2_p*TEF*(406)-hrGFP	1kb_*XPR*2_up-p*TEF*(406)-hrGFP-*LIP*2t-1kb_1 Kb *XPR*2_down, *CEN*1, *URA*3, AmpR, *Co1E*1	This study
pAS192 [AS0540]	pHR_*XPR*2_p*TEF*(-41–406)-hrGFP	1kb_*XPR*2_up-p*TEF*(-41–406)-hrGFP-*LIP2*t-1kb_1 Kb *XPR*2_down, *CEN*1, *URA*3, AmpR, *Co1E*1	This study
pAS142 [AS0482]	pCRISPRyl_*AXP*1_p$UAS1B_8$-*TEF*(136)-Cas9	p$UAS1B_8$-*TEF*(136)-Cas9-*CYC*1t, *SCR*1′-tRNA-sgRNA$_{AXP1}$, *CEN*1, *LEU2*, AmpR, *ColE*1	(Abdel-Mawgoud and Stephanopoulos 2020)
pBO002 [AS0630]	pCRISPRyl_*AXP1*_p*TEF*(406)-Kozak-Cas9	p*TEF*(406)-Kozak-Cas9-*CYC*1t, *SCR*1′-tRNA-sgRNA$_{AXP1}$, *CEN*1, *LEU2*, AmpR, *ColE*1	This study
pBO003 [AS0631]	pCRISPRyl_*AXP*1_p*TEF*(-41–406)-Kozak-Cas9	p*TEF*(-41–406)-Kozak-Cas9-*CYC*1t, *SCR*1′-tRNA-sgRNA$_{AXP1}$, *CEN*1, *LEU2*, AmpR, *ColE*1	This study
pHR_AXP_hrGFP [AS0404]	pHR_*AXP*1_p$UAS1B_8$-*TEF*(136)-hrGFP	1kb_*AXP*1_up-p$UAS1B_8$-*TEF*(136)-hrGFP-*CYC1*t-1kb_*AXP*1_down, *CEN*1, *URA*3, AmpR, *Co1E*1	Addgene #84613, (Schwartz, Shabbir-Hussain et al. 2017)
pAS238 [AS0586]	pCRISPRyl_*GSY*1_p$UAS1B_8$-*TEF*(136)-Cas9	p$UAS1B_8$-*TEF*(136)-Cas9-*CYC*1t, *SCR1*′-tRNA-sgRNA$_{GSY1}$, *CEN*1, *LEU2*, AmpR, *ColE*1	This study
pAS246 [AS0594]	pCRISPRyl_*GSY*1_p*TEF*(-41–406)-Kozak-Cas9	p*TEF*(-41–406)-Kozak-Cas9-*CYC*1t, *SCR*1′-tRNA-sgRNA$_{GSY1}$, *CEN*1, *LEU2*, AmpR, *ColE*1	This study
pAS239 [AS0587]	pCRISPRyl_*ADE*2_p$UAS1B_8$-*TEF*(136)-Cas9	p$UAS1B_8$-*TEF*(136)-Cas9-*CYC*1t, *SCR*1′-tRNA-sgRNA$_{ADE2}$, *CEN*1, *LEU2*, AmpR, *ColE*1	This study
pAS244 [AS0592]	pCRISPRyl_*ADE*2_p*TEF*(-41–406)-Kozak-Cas9	p*TEF*(-41–406)-Kozak-Cas9-*CYC*1t, *SCR1*′-tRNA-sgRNA$_{ADE2}$, *CEN*1, *LEU2*, *AmpR*, *ColE*1	This study
pAS241 [AS0589]	pCRISPRyl_*CAN*1_p$UAS1B_8$-*TEF*(136)-Cas9	p$UAS1B_8$-*TEF*(136)-Cas9-*CYC*1t, *SCR*1′-tRNA-sgRNA$_{CAN1}$, *CEN*1, *LEU2*, AmpR, *ColE*1	This study
pAS248 [AS0596]	pCRISPRyl_*CAN*1_p*TEF*(-41–406)-Kozak-Cas9	p*TEF*(-41–406)-Kozak-Cas9-*CYC1*t, *SCR*1′-tRNA-sgRNA$_{CAN1}$, *CEN*1, *LEU2*, AmpR, *ColE*1	This study
pAS240 [AS0588]	pCRISPRyl_*LIP*2_p$UAS1B_8$-*TEF*(136)-Cas9	p$UAS1B_8$-*TEF*(136)-Cas9-*CYC*1t, *SCR*1′-tRNA-sgRNA$_{LIP2}$, *CEN*1, *LEU2*, AmpR, *ColE*1	This study
pAS250 [AS0598]	pCRISPRyl_*LIP*2_p*TEF*(-41–406)-Kozak-Cas9	p*TEF*(-41–406)-Kozak-Cas9-*CYC*1t, *SCR*1′-tRNA-sgRNA$_{LIP2}$, *CEN*1, *LEU2*, AmpR, *ColE*1	This study

Table 3
List of oligonucleotides.

No	Name	Sequence	Ref.
78	F-*XPR*2-up	ggttggtggtgggaagaac	(Schwartz, Shabbir-Hussain et al. 2017)
79	F-pHR-*CYC*1t	ctcgaaggctttaatttgccctagg	(Schwartz, Shabbir-Hussain et al. 2017)
80	R-*XPR*2-dn	ggtcctatgcatccctgaaac	(Schwartz, Shabbir-Hussain et al. 2017)
115	F-*AXP*1-up	cctccgaagaagcaaaagtg	(Abdel-Mawgoud and Stephanopoulos 2020)
116	R-*AXP*1-dn	ctctgggccgaatacaacac	(Abdel-Mawgoud and Stephanopoulos 2020)
182	pCRISPRyl_gRNA-*Bsm*BI-Flong	tcgattccgggtcggcgcaggagacggtccacgccgtctcggttttagagctagaaatag	This study
183	pCRISPRyl_gRNA-*Bsm*BI-R-Long	ctatttctagctctaaaaccgagacggcgtggaccgtctcctgcgccgacccggaatcga	This study
231	pCRISPRyl_gRNA-*Bsm*BI-del-F	ccgttgatttccgaacagatctggccctttgtcacagcttgtctgtaagcggatgccggg	This study
232	pCRISPRyl_gRNA-*Bsm*BI-del-R	cccggcatccgcttacagacaagctgtgacaaagggccagatctgttcggaaatcaacgg	This study
481	F-*TEF*-366-*Tsp*MI	atcccccgggagagaccgggttggcg	This study
482	R-*TEF*-366-*Asc*I	cttcggcgcgccttcgggtgtgagttgacaagg	This study
490	*GSY*1-up-50 bp	caaccccaactcttcctctttctctctcgcaacaaccgattccaacatgtaggcagaag	This study
491	*GSY*1-dn-50 bp	tccttctccctactatcaagtagtacattcaaacactcaacttctgcctacatgttggaatcg	This study
492	*ADE*2-up-50 bp	gtctagtttcacagaaaccaccctcacaatccccaatcccttcagcaatgtagttctagacgc	This study
493	*ADE*2-dn-50 bp	atttaacgtcttacagtgtgcatatagataatttttgtgcgtctagaactacattgctgaagg	This study
494	*CAN*1-up-50 bp	gacgtttcgaccttaacgaccctgccgtctccatccatccgaccacaatgtaagaagatcg	This study
495	*CAN*1-dn-50 bp	taagagtggtttgctccaggagagagcgtcgggaatccccgatcttcttacattgtggtcgga	This study
496	*LIP*2-up-50 bp	gccatagaaagccctcaattgatacccaagtaccagctcccctcactatgtaagctatttatcactc	This study
497	*LIP*2-dn-50 bp	attaaagtagatagttgaggtagaagttgtaaagagtgataaatagcttacatagtgagggg	This study
500	F-*ADE*2-up	tcaggatcggaacttcgcac	This study
501	R-*ADE*2-dn	tgcattgccacgacctgtta	This study
504	F-*GSY*1-up	cgaacacagaggggaaatacga	This study
505	R-*GSY*1-dn	actactcgttccactgctgc	This study
508	F-*LIP*2-up	tccgacacgacagcttttga	This study
509	R-*LIP*2-dn	ggttgatcggttccctacgg	This study
512	F-*CAN*1-up	accggcacctatccacacta	This study
513	R-*CAN*1-dn	gagtacatgcgcgtcgtttc	This study
558	R-*TEF*-406-*AscI*	acacggcgcgcctttgaatgattcttatactcagaaggaaatgc	This study
593	d(T)20	ttttttttttttttttttt	This study
656	F-*XPR*2-dn-pAS160	aaaggggatcccccgggttcgaagtaataagagcctcgacctg	This study
657	R-*XPR*2-dn_rev-pAS160	atcgacgcgtaaagctggcgcgccactcgtccaacggtatcac	This study
658	F-*TEF*-pAS160	ggcgcgccagctttacgcgtcgatagagaccgggttggcggc	This study
659	R-*TEF*-pAS160	ttgtggatccttcgggtgtgagttgacaagg	This study
660	F-hrGFP-pAS160	cacacccgaaggatccacaatggtgagcaagcagatc	This study
661	R-hrGFP-pAS160	cagacaccctaggttacacccactcgtgcag	This study
662	F-*LIP*2t-pAS160	ggtgtaacctagggtgtctgtggtatctaag	This study
663	R-*LIP*2t-pAS160	gactctctggtccggactgcagattaatttcgatttgtcttagaggaac	This study
664	F-*XPR*2-up-pAS160	caaatcgaaattaatctgcagtccggaccagagagtccagcttctg	This study
665	R-*XPR*2-up-pAS160	ctatgaccatgattacgccaagcttttaattaagaacacagtgtcaagacaag	This study
666	F-*TEF*(406)-pAS190	atatacgcgtcgatagagaccgggttggcggc	This study
667	R-*TEF*(406)-pAS190	caccattgtgtttgaatgattcttatactcagaaggaaatgcttaacgatttcg	This study
668	F-hrGFP-pAS190	atcattcaaacacaatggtgagcaagcagatc	This study
669	R-hrGFP-pAS190	acaccctaggttacacccactcgtgcag	This study
670	R-*TEF*(406)-pAS191	tgctcaccattttgaatgattcttatactcagaaggaaatgcttaacgatttcg	This study
671	F-hrGFP-pAS191	atcattcaaaatggtgagcaagcagatc	This study
672	R-*TEF*(-41–406)-pAS192	tgctcaccatttcgggtgtgagttgacaaggagag	This study
673	F-hrGFP-pAS192	cacacccgaaatggtgagcaagcagatc	This study

assembly (NEBuilder HiFi, New England Biolab) of dsDNA (60 bp) prepared by thermal annealing of complementary oligonucleotides 182/183 onto *Avr*II linearized pAS157. The pAS157 plasmid was constructed (Addgene #70007) by deletion of *Bsm*BI sites on the pCRISPRyl backbone downstream to *LEU*2. This was conducted by Gibson assembly of dsDNA (60 bp) prepared by thermal annealing of complementary oligonucleotides 231/232 onto *Bsm*BI linearized pCRISPRyl (Addgene #70007). The pAS160 (pHR-*XPR*2-1 KB-p*TEF*(-41–406)-Kozak-hrGFP-*lip2*t) was constructed by the 6 fragments Gibson assembly (NEBuilder HiFi, New England Biolab) of the *Bst*BI/*Hin*dIII linearized (Addgene #84614) and the PCR fragments of XPR2-dn, p*TEF*(-41–406), hrGFP, *LIP*2t, and *XPR*2-up amplified using the pair of oligonucleotides 656/657, 658/659, 660/661, 662/663 and 664/665. The derivative pHR plasmids integrating hrGFP at *XPR*2 site, namely pAS190, pAS191 and pAS192 harbored *pTEF*(406)-Kozak-hrGFP, *pTEF*(406)-hrGFP and *pTEF*(-41–406)-hrGFP. The pAS190, pAS191 and pAS192 were constructed by the ligation of *Mlu*I/*Avr*II digested *TEF*-hrGFP fragments, that were previously assembled by PCR-spliced by overlap extension (SOE), onto the *Mlu*I/*Avr*II digested backbone of pAS160. The SOE assembly of different bipartite *TEF*-hrGFP fragments were conducted using the oligonucleotide pairs 666/667 + 668/669, 666/670 + 671/669 and 666/672 + 673/669 for pAS190, pAS191 and pAS192, respectively.

The pAS141 and pAS142 derivative plasmids with variation in the promoter sequences with validated by sanger sequencing using the universal M13 (−20) primer.

Transformation of bacterial and yeast cells

Transformation of *E. coli* was conducted by transferring 100 ng of DNA to 25 μL of competent cells incubated on ice 30 min followed by heat shock at 42 °C for 45 s. Cell recovery was performed, after a 5-min of incubation on ice, at 37 °C in 20 volumes of LB broth for 45 min. Cell suspension is concentrated to a volume of 100 μL by centrifugation then plated on LB solid medium with carbenicillin 100 μg/mL.

Transformation of *Y. lipolytica* PO1g and PO1f was Transformation of *Y. lipolytica* PO1g and PO1f was carried out using a LiAc/ssDNA/PEG protocol (Abdel-Mawgoud and Stephanopoulos 2020). A volume equivalent to 1,000 μL·OD was collected from an overnight culture grown in YPD at an OD_{600} between 10 and 15. Cells were centrifuged at 22,000 xg for 1 min, washed twice with 50 μL of lithium acetate solution (100 mM, pH 6.0), then suspended in 100 μL of transformation mix (PEG-3350 45 % w/v, Sigma-Aldrich); lithium acetate 0.1 mM (Sigma-Aldrich); DL-dithiothreitol 0.1 mM, Fisher Scientific) and incubated on ice for 15 min. A volume of 2.5 μL of salmon sperm DNA (10 mg/mL,

Sigma-Aldrich, Canada) was added together with 1 μg of respective pCRISPRyl (Cas9/gRNA) and deletion allele constructed by Overlap Extension PCR of respective pair of oligonucleotides (Table 2, 3) for chromosomal deletions (gene knock out). For chromosomal insertion, transformation mixes were supplemented with the respective pCRISPRyl (Cas9/gRNA) and a pHR (homologous recombination template) plasmids (Table 2). Transformation mixes were incubated 30 min at 28 °C then thermally shocked at 39 °C for 30 min. Suspensions were plated on selective $YNB_{w/o.aa}$-Ura minimal media and incubated for at least 2 days at 28 °C until the development of colonies. As for *ade2* knock-out transformants, the transformation suspension was plated on a minimal synthetic medium supplemented with adenine (BioBasic) at 800 μg/mL to allow the transformants harboring the deletion to grow (Verbeke et al., 2013).

Colony-PCR for screening of positives clones

The *Y. lipolytica* transformants were screened using colony-PCR as described earlier (Abdel-Mawgoud and Stephanopoulos, 2020). A biomass of roughly 1 mm^3 was picked using micropipette tips and was then suspended to homogeneity in PCR tubes containing 10 μL of NaOH 20 mM and incubated at 98 °C, 15 min. A volume of the lysates not exceeding 10 % of the total PCR reaction volume was added to the respective PCR mix. Screening of correct clones was conducted with multiplex PCR using the sets of oligonucleotides 115(Fwd)/79(Fwd)/116(Rev) and 78(Fwd)/79(Fwd)/80 (Rev) (Table 3) for identification of integrations at *axp*1 and *xpr*2 loci, respectively. For the screening of deletions, the respective sets of primers 504(Fwd)/505(Rev), 500(Fwd)/501(Rev), 512(Fwd)/513(Rev) and 508(Fwd)/509(Rev) were used to identify knock-out transformants at the loci *GSY*1, *ADE*2, *CAN*1 and *LIP*2. All of which was done using the Phire Hot Start II DNA Polymerase (Thermo Fisher Scientific). Every transformation experiment was conducted in biological triplicates from each of which 20 colonies were screened by colony PCR.

*LIP*2 transformants were sequenced by Sanger sequencing as to identify the indel events. *LIP*2 was amplified using the 508(Fwd)/509 (Rev) primers and sequenced using the same primers.

mRNA quantification using RT-qPCR

Precultures of *Y. lipolytica* PO1f strains grown in YPD were used to inoculate 3 mL of fresh YPD broth at initial OD_{600} of 0.1 in test tubes that were placed in a rotary shaker (60 rpm) and incubated at 28 °C. After about 5 h of incubation where cells reached ~ 1 OD_{600}, cells were collected by centrifugation at 22,000 xg for 2 min and washed twice in phosphate buffer PB. Cell pellets were then suspended in 10 volumes of TRIzol reagent (Ambion) and RNA was extracted according to the manufacturer protocol. TRIzol recovered RNA was purified by ethanol precipitation and then column purified (Bio Basic). An amount of 1 μg of purified RNA was treated with RQ1 Dnase (Promega) in the presence of Murine Rnase inhibitor (New England Biolabs), then cDNA synthesis was conducted with M−MLV Reverse transcriptase (Promega) using d $(T)_{20}$ primer (# 593,Table 3) all according to manufacturer protocols. Preparations of cDNA were diluted to have 100 ng to be used as templates for qPCR using the standard program a qPCR thermocycler (Applied Biosystems 7500, Thermo Fisher Scientific) and using Ssoadvanced universal SYBR Green supermix (Bio-rad) with the primer sets 595(fwd)/606(rev), 602(fwd)/603(rev) and 609(fwd)/610(rev) to measure the expression of the chromosomally integrated *ACT*1 (housekeeping gene) and *hrGFP* (gene of interest), respectively. The qPCR amplification efficiencies of *ACT*1 and hrGFP were calculated using serial dilutions of gDNA containing hrGFP chromosomal integration, *xpr*2::hrGFP. The average fold changes of expression were calculated using the Pfaffl method (Pfaffl 2001).

In silico analysis of promoter architecture

The Transcription Start Site (TSS) of promoter was predicted using the Neural Network Promoter Prediction (NNPP) webtool (Reese 2001), accessed from https://www.fruitfly.org/seq_tools/ promoter.html, and using a minimum eukaryotic promoter prediction score of 0.98. The TATA box was predicted using the YAPP Eukaryotic Core Promoter Predictor webtool, accessed from http://www.bioinformatics.org/yapp/cgi-bin/yapp.cgi.

For prediction of potential Transcription Factor Binding Sites (TFBS) were screened with Patch 1.0 web application (https://gene-regulation.com/) relying on the TRANSFAC® Public 6.0 transcription factor database. Apart from restricting our search to TFBS of fungi, we used the default search parameters allowing a minimum length of TFBS of 4 bp, a maximum number of mismatches of 2, a mismatch penalty of 100 and a minimum score of 90. In addition, we searched for potential TFBS using Match 1.0 web application relying on the TRANSFAC® Public 6.0 transcription factor database. Apart from restricting our search to TFBS of fungi, we used the default search parameters employing matrixTFP60.lib as the Weight Matrix library using a cut-off that minimizes false positive matches.

Results

In silico analysis of the pTEF1α architecture

For engineering a minimal *TEF* promoter, we built on the previous 5′-truncated version of p*TEF*1*α* that was reported to harbor the minimal core promoter region corresponding to −1 to −406, p*TEF*(406), relative to the start codon of *TEF*1α (Blazeck et al., 2011). Building on that promoter, we wanted to investigate if 3′-truncations of p*TEF*(406) could further allow reduction of the promoter size without jeopardizing its activity.

We first looked for the transcription start site (TSS) using Neural Network Promoter Prediction (NNPP) web tool (Reese 2001). NNPP predicted the TSS to be at −81 thymine nucleotide at a prediction score of 0.98. We then searched for the TATA box using the Eukaryotic Core Promoter Predictor (YAPP) web application while restraining the position of TSS to −81. The TATA box was predicted (at a prediction score of 0.99) as 12-base motif, GGTATAAAAGAC, stretching from −102 to −113, i.e., upstream to the TSS position by 20 bp.

Afterwards, we identified potential transcription factor binding sites (TFBS) over the p*TEF*(406) sequence using the Patch 1.0 and Match 1.0 web applications. These tools resulted in the identification of 30 potential non-redundant TFBS with a score of 100, 14 of which are sense, 13 are antisense and 3 are bi-directional (Fig. 1). Of the predicted TFBS, 24 were upstream to the TSS and were outside the region of our focus, yet only 6 were overlapping or non-overlapping with TSS were downstream to it. The overlapping TFBS were 4 and their motif sequences were partially overlapping to TSS and to each other, whereas the remaining 3 were further downstream and were not overlapping at all. Given that we aimed at conducting 3′-truncations of p*TEF*(406) that are not jeopardizing its activity, we focused on the region encompassing the 3 non-overlapping motifs downstream to TSS that spanned an area of 40 bp and contained the last 3 identified TFBS, namely *GAL*4, *MED*8 and *GCN*4 the first of which is bi-directional (the antisense of which is part of the 4 overlapping TFBS) and the latter two are reverse oriented relative to the direction of transcription (Fig. 1).

Given the debatable orientation-dependency of TFBS functions (Lis and Walther 2016), we decided to assume a potential regulatory role for bi-directional, *GAL*4, or reverse oriented TFBS, *MED*8 and *GCN*4.

*GAL*4 is a TFBS that is composed of a bi-directional pentameric motif (CCGAA), and to which binds the transcription factor (TF) to activate *GAL* gene and its large regulon that is not necessarily related to galactose metabolism, yet to global adaptation to growth on galactose (Bhat and Murthy, 2001; Traven et al., 2006). It has been shown that, upon

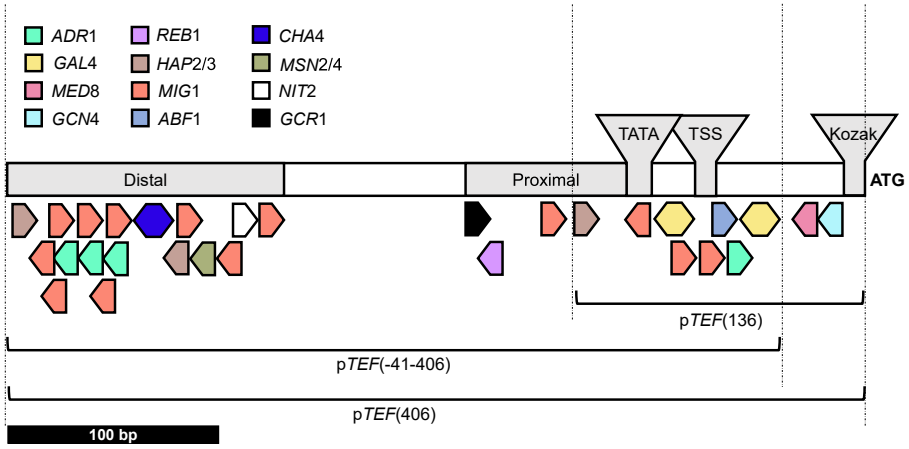

Fig. 1. Architecture of the essential elements and transcription factor binding sites predicted on the p*TEF*1α based promoters. The binding sites were obtained using the web-based tool NNPP and the web application Patch 1.0 of TRANSFAC® on the p*TEF*(406) from which derivates the p*TEF*(-41–406) and the p*UAS*1B_8-*TEF*(136). The structural features TATA, TSS are common among the three promoters, while the native Kozak differ in the p*TEF*(-41–406) and the proximal and distal regions in the p*TEF*(136). The proximal and distal regions account in total 12 different TF binding to 30 potential non-redundant TFBS. In the case of the p*TEF*(-41–406), 40 bp were retracted from starting from the ATG which removes the TFBS *GCN*4 and *MED*8.

galactose induction, disruption of the complex chromatin structure, namely nucleosomes positioned over the TATA boxes and transcription initiation sites, is carried out in a Gal4-dependent manner (Traven et al., 2006). For its potential positive galactose-dependent activating functions regulatory effect, we decided not to remove Gal4 binding site, so that it could be later used for artificial induction of overexpression by external supplementation of galactose.

*MED*8 is a TFBS that is composed of a heptameric motif (AGGAAAT) to which binds the Med8 TF. It has been reported that Med8 acts as coupling factor or mediator protein bridging both activating and repressing factors to the RNA polymerase II holoenzyme, therefore is involved in both positive and negative regulatory responses (Chaves et al., 1999).

*GCN*4 is a TFBS that is composed of a hexameric motif (TGATTC) to which binds the Gcn4 TF. Gcn4 is a positive transcriptional regulator of amino acid biosynthesis (Natarajan et al., 2001) yet a negative regulator of ribosomal proteins synthesis and protein synthesis in general (Joo et al., 2011, Mittal et al., 2017).

Given the potential negative regulatory functions of *MED*8 and *GCN*4 TFBS, we decided to knock them out from the 3′end of p*TEF*(406) to engineer a shorter version of the promoter that is either higher or at least not lower in activity than the parent promoter. The decided 3′-truncation region is 40 bp flanked by the *GAL*4 TFBS and the start codon at its 5′-end and 3′-end, respectively. The newly 3′-truncated promoter was denoted for as p*TEF*(-41–406).

Upstream to the core promoter region (TATA and TSS) of p*TEF* (-41–406), two regions could be identified, proximal and distal one. The proximal upstream region (-114 to −189 in p*TEF*(406) and −74 to −149 in p*TEF*(-41–406)) contained the TF *GCR*1, *REB*1, *MIG*1 and *HAP*2/3 TFBS. *REB*1 TFBS is composed of an undecameric motif while the others are pentameric motifs. Gcr1 is a key transcriptional activator of glycolytic genes which possesses two isoforms activating transcription differentially throughout growth phases (Cha et al., 2021). Reb1 is suggested to alter nucleosome position and thus favors the accessibility of polymerase II for gene transcription (Wang and Donze 2016). On the other hand, Hap2/3 both act as activators, especially for mitochondrial genes, and repressors like Mig1 in response to carbon sources (Hagerman and Willis 2002, Buschlen et al., 2003, Chen et al., 2005).

The distal regions is upstream to the proximal region and separated from it by an 85-bp region devoid of any predicted TFBS. This distal region is thought to be the enhancer region spanning the region from −275 to −406 in p*TEF*(406) and from −235 to −366 in p*TEF*(-41–406). The distal region comprises 6 different TFBS present in multiple copies and orientations, namely, *MIG*1, *HAP*2/3, *ADR*1, *CHA*4, *MSN*2/4 and *NIT*2. Those TFBS are targets of positive regulators activating expression in response to different cellular conditions as stress response and oxidative metabolism (Chiang et al., 1996, Holmberg and Schjerling 1996, Mo and Marzluf 2003, Rajvanshi et al., 2017).

Construction of chromosomal integration of hrGFP fusions under different TEF promoters

We constructed *Y. lipolytica* strains harboring hrGFP coding gene that is chromosomally integrated at *XPR*2 and is expressed under different promoter designs, namely, p*TEF*(-41–406)-Kozak-hrGFP, p*TEF*(406)-Kozak-hrGFP, p*TEF*(406)-hrGFP, p*TEF*(-41–406)-hrGFP, p*TEF*(406)-hrGFP$_{intron}$, and p*UAS*1B_8-*TEF*(136) (Fig. 2).

Both the designs p*TEF*(-41–406)-Kozak-hrGFP (pAS160, Table 2) and the Kozak-less p*TEF*(-41–406)-hrGFP (pAS192, Table 2), harbor our engineered 3′-truncated promoter, p*TEF*(-41–406), that is fused to hrGFP, yet in the former, a Kozak sequence, CACA, is fused in between p*TEF*(-41–406) and hrGFP coding sequence. As controls representing the native *TEF* promoter, the p*TEF*(406)-Kozak-hrGFP (pAS190, Table 2) and p*TEF*(406)-hrGFP (pAS191, Table 2) were similarly constructed. The 5′-truncated, p*TEF*(406), instead of the full *TEF* promoter, p*TEF* (1004), was used as Blazeck et al. (2011) showed them to be equally performing. The p*TEF*(406)-Kozak-hrGFP harboring Kozak in between the promoter and the coding sequence was constructed however to act as control to the our 3′-truncated version, p*TEF*(-41–406)-Kozak-hrGFP. The p*UAS*1B_8-*TEF*(136) was reported by Blazeck et al. (2011) and showed to be 30 % stronger than p*TEF*(406). The p*UAS*1B_8-*TEF*(136) construction was directly supplied from Addgene (pHR_*XPR*2_hrGFP, addgene# 84614, Table 2) and in which a fusion of 8 tandem Upstream Activating Sequence (*UAS*1*B*) units were fused to 5′-truncated p*TEF* (136) comprised of the native −1 to −136 upstream region of the coding sequence of *TEF*1α and that was directly fused to the hrGFP coding sequence.

Assay of the relative strength of 3′truncated pTEF

To investigate the relative strength of the 3′-truncated p*TEF* (-41–406), five *Y. lipolytica* PO1f strains harboring the p*TEF* variants fused to GFP reporter integrated at *xpr*2 site, *xpr*2-322::p*TEF*-hrGFP, were grown in YPD and compared in terms of relative fluorescence of their expressed hrGFP using microplate readers as well as the relative quantity of their mRNA transcripts using RT-qPCR (Fig. 3). For GFP-

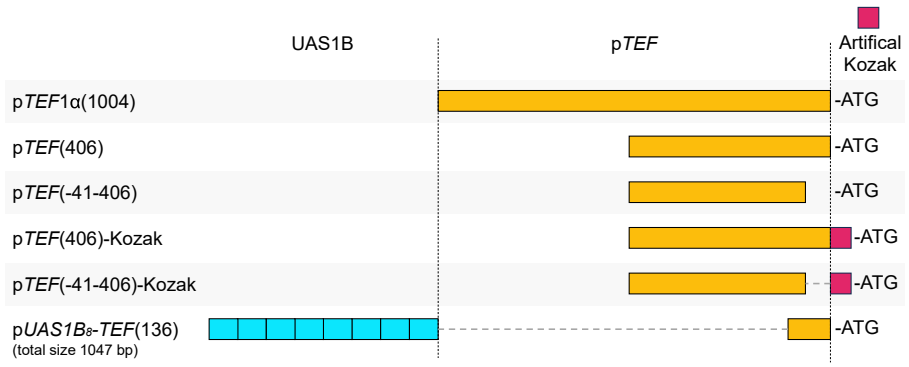

Fig. 2. Construction of the p*TEF*-based promoters expressing hrGFP. The different promoters p*TEF*(406), p*TEF*(-41–406), p*TEF*(406)-Kozak, p*TEF*(-41–406)-Kozak and p*UAS*1B$_8$-*TEF*(136) are shown. The lines show differences in length due to either truncations or missing features present on other promoters such as the fused Kozak and *UAS*1*B*.

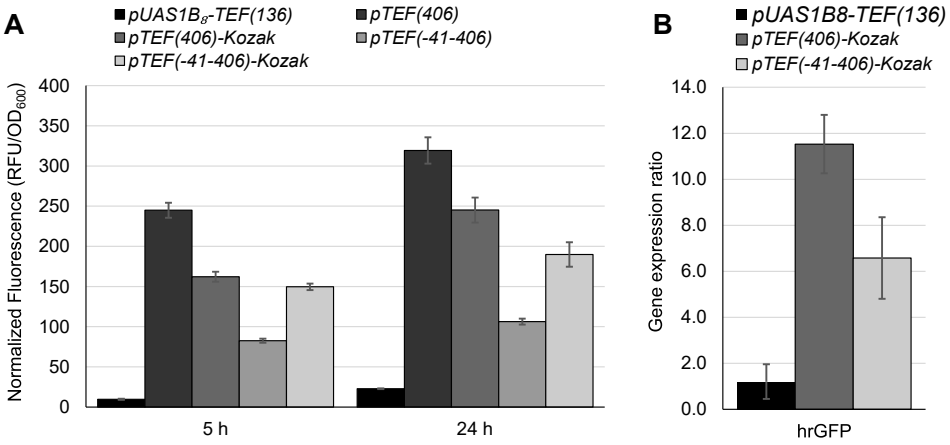

Fig. 3. Fluorescence-based assay of protein expression and RT-qPCR-based assay of gene transcription to determine the strength of expression of hrGFP under different promoters. (**A**) Fluorescence readings normalized to optical density at 600 of hrGFP expressed under p*UAS*1B$_8$-*TEF*(136), p*TEF*(406)-Kozak and p*TEF*(-41–406)-Kozak measured in washed cells collected after 5 and 24 h of incubation. (**B**) RT-qPCR measurements of transcription of chromosomally integrated hrGFP expressed under p*UAS*1B$_8$-*TEF*(136), p*TEF*(406)-Kozak and p*TEF*(-41–406)-Kozak promoters in cells collected at early logarithmic phase. Expression values are normalized to that of the reference p*UAS*1B$_8$-*TEF*(136) promoter. The mean and the standard deviation from 3 biological replicate data are shown. For RT-qPCR, technical triplicates were additionally done for every biological replicate.

fluorescence-based assay of promoter strength, GFP fluorescence was monitored in early and late growth phases (5 and 24 h) as signal plateaued and decreased with longer incubation. Maximum GFP fluorescence values, normalized to optical density, were observed at late logarithmic phases, after 24 h of incubation, which was 30 % higher that after 5 h (Fig. 3**A**).

Interestingly, GFP expressions, as measured in terms of GFP fluorescence levels, were 8 to 25-folds higher when expressed under Kozak or Kozak-less versions of p*TEF*(406) and p*TEF*(-41–406) compared to the p*UAS*1B$_8$-*TEF*(136) early on. In cells recovered at the late logarithmic phase of growth the expressions were 5 to 14 times higher than p*UAS*1B$_8$-*TEF*(136).

When comparing GFP expression under p*TEF*(406) and p*TEF*(-41–406), we observed that the Kozak or Kozak-less versions of p*TEF*(406) was associated with higher expression than the corresponding versions of p*TEF*(-41–406) by 1.2 and 3 folds, respectively. The p*TEF*(-41–406)-Kozak is thus leading to 15 % lower expression than the p*TEF*(406)-Kozak whereas the p*TEF*(-41–406) is less expressive by 66 % than the p*TEF*(406). Moreover, the fusion of Kozak sequence to 3′-ends of p*TEF*(-41–406) appeared to be beneficial as the p*TEF*(-41–406)-Kozak was only 40 % less strong than p*TEF*(406).

With regard to the impact of Kozak sequence on promoter strength compared to its absence, we observed that it has a positive effect (increased by 1.8 times, 180 %) on p*TEF*(-41–406) yet a negative effects (decreased by 30 %) on p*TEF*(406) strengths, when each is compared to itself without this Kozak sequence. This entails that the presence of Kozak at 3′-end of these promoters has opposite effects depending on its upstream environment. The partial recovery of promoter strength of the 3′-truncated p*TEF*(-41–406) relative to the 3′-intact one, p*TEF*(406), by the addition of a Kozak sequence at its 3′-end suggests that the fused Kozak sequence partially relieves the negative effect of 3′-truncation, which is an important element helping ribosomes localize the start codon for initiation of translation. On the other hand, the addition of an artificial Kozak sequence downstream to p*TEF*(406) is not necessary as it decreases its overall promoter strength. This might entail that the native 3′-intact p*TEF* have an intrinsic Kozak sequence (CAAA**ATG**) whose function is halted by the addition of an artificial Kozak sequence in tandem.

To further validate fluorescence-based protein expression assays of strengths of different promoter designed, we assayed GFP transcription levels from chromosomally integrated GFP under different promoter designs at early logarithmic phase of growth, i.e. after ~ 5 h, using RT-qPCR. In concordance with the GFP expression values obtained with the fluorescence-based assay (Fig. 3**A**), both p*TEF*(406)-Kozak and p*TEF*(-41–406)-Kozak were associated with at least 5 times higher transcription levels than that of p*UAS*1B$_8$-*TEF*(136) (Fig. 3**B**). Again, we observed the same expression profile where p*TEF*(406)-Kozak was associated with a higher expression than p*TEF*(-41–406)-Kozak by 70 %.

To conclude, the translational and transcriptional assays of of GFP expression under different promoter designs are corroborating and confirm that, although less expressive than the p*TEF*(406), the p*TEF* (-41–406)-Kozak is a much stronger and a more convenient promoter for protein expression than p*UAS1B*$_8$-*TEF*(136) currently in use for driving Cas9 expression.

Impact of Cas9 expression level on chromosomal insertion of genes

Under the goal to understand the impact of Cas9 expression level on the CRISPR-Cas9 mediated genome editing efficiency, we first compared the gene knock-in, i.e. gene integration, efficiencies when using plasmid-carried Cas9 expressed under different promoter designs, namely, p*UAS1B*$_8$-*TEF*(136), p*TEF*(406)-Kozak and p*TEF*(-41–406)-Kozak. The used CRISPR plasmids are those expressing both Cas9 and the improved gRNA versions developed by Abdel-Mawgoud and Stephanopoulos (2020) targeting gene integrations at *axp*1 or *xpr*2 loci, which are already mutated loci in the chromosomes of *Y. lipolytica* PO1f strain. To induce HDR-mediated hrGFP integration at these loci, locus-specific homologous recombination plasmids (pHR) were used as repair templates, where hrGFP was flanked with upstream and downstream homology arms appropriate to respective target loci (Table 2). Following transformation, genome editing efficiency was measured by colony PCR on transformants and reported as percentage of positive clones with correct gene integrations.

Results showed that, apart from gene integration efficiencies in which gene integration was already close to 100 % at *axp*1 locus, Cas9 expressed under the strong promoters, p*TEF*(406)-Kozak or p*TEF* (-41–406)-Kozak, were associated with about 35 % higher gene integration efficiencies at the challenging *xpr*2 locus compared to p*UAS1B*$_8$-*TEF*(136) (Fig. 4). This means that higher Cas9 expression under significantly stronger promoters, p*TEF*(406)-Kozak or p*TEF*(-41–406)-Kozak, is associated with higher genome editing efficiencies than that observed with Cas9 expressed under weaker promoters like the p*UAS1B*$_8$-*TEF*(136).

In addition, Cas9 expressed under our p*TEF*(-41–406)-Kozak was associated with gene integration efficiencies equivalent to that when expressed under the p*TEF*(406)-Kozak (Fig. 4) although the latter showed to be stronger promoter (by as much as 30 %) (Fig. 3). This means that slight differences in Cas9 expression does not necessarily reflect on differences on genome editing efficiencies.

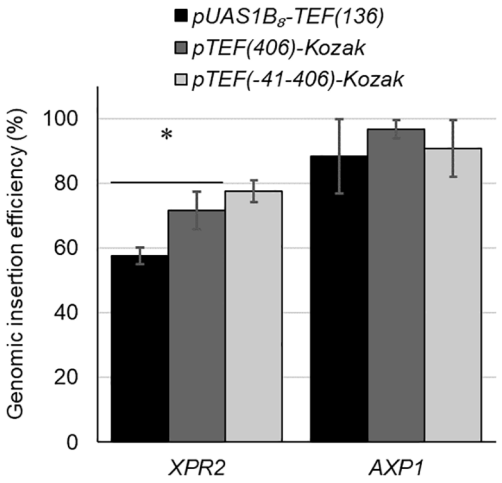

Fig. 4. Gene integration efficiency at chromosomal site-specific loci using Cas9 expressed under different promoter designs. Gene integrations at *axp*1 and *xpr*2 loci were conducted with pCRISPR-Cas9 plasmids carrying gRNA$_{AXP1}$ and gRNA$_{XPR2}$, respectively, together with locus-specific pHR-hrGFP plasmids having 1000 bp-long homology arms Gene integration efficiencies were assessed on transformants by identification of the percentage of positive clones using multiplex colony-PCR. The mean and the standard deviation are from data of 20 transformants each of 3 replicates are shown. Analysis of variance was carried out at $\alpha = 0.05$ (*).

Overall, p*TEF*(-41–406)-Kozak as driver for Cas9 expression clearly outperforms the p*UAS1B*$_8$-*TEF*(136) in genome editing efficiency. Considering the equally efficient gene integration obtained from p*TEF* (-41–406)-Kozak and p*TEF*(406)-Kozak, we therefore decided to pursue our study of the impact of Cas9 expression strength on genome editing efficiency using the relatively shorter p*TEF*(-41–406)-Kozak in comparison with p*UAS1B*$_8$-*TEF*(136).

Impact of Cas9 expression level on efficiency of chromosomal gene deletion

After having demonstrated the impact of Cas9 expression level on the efficiency of gene integrations at chromosomes, we wanted to examine also its impact on the efficiency of chromosomal gene deletion (knock out). In this experiment, we compared the efficiency of gene deletions with Cas9 expressed under our engineered 3′-truncated p*TEF*(-41–406)-Kozak with that expressed under the weakest promoter, p*UAS1B*$_8$-*TEF* (136). For each promoter, we examined the Cas9-mediated inactivation of target genes, namely, *GSY*1, *ADE*2, *CAN*1 and *LIP*2 using respective deletion templates (gene-specific 100-bp deletion allele) inducing a knock-out of the whole open reading frame between the start and stop codons.

Even when using repair templates to induce full gene knock out by HDR (resulting in typical deletions), DSB induced by CRISPR-Cas9 might also be repaired by NHEJ resulting in micro-indels (microinsertions and microdeletions) which are hardly detected by PCR, yet, both types of repairs cause gene inactivation and loss of gene functions. Therefore, we decided to estimate the efficiency of gene inactivation using specific phenotypic screens via estimation of the percentage of positive clones developing visible phenotypic changes on or around their colonies, as a result of loss of gene functions, relative to the total number of screened transformants. In addition to phenotypic screens, we also used PCR and agarose gel electrophoresis to estimate the rate of full gene knock out events (typical deletions) induced by used repair templates, as well as PCR and sequencing to detect microindel events.

Cas9 expressed under p*TEF*(-41–406)-Kozak or p*UAS1B*$_8$-*TEF*(136) promoters resulted in equivalent gene inactivation efficiencies as estimated in terms of phenotypic loss as well as in terms PCR-confirmed typical deletion (Fig. 5). For the four gene deletion targets, *GSY*1, *ADE*2, *CAN*1 and *LIP*2, we obtained a near complete loss of relevant phenotypes on screened transformants (Fig. 5**A, S1**). In contrast however, the rate of typical deletions (knock-out) varied from one gene target to another with a rate of typical deletions occurring at 76.2 %, 60.8 %, 39.3 and 14 % for *CAN*1, *LIP*2, *ADE*2 and *GSY*1, respectively (Fig. 5**B**). The discrepancy between the rate of phenotypic loss (~100 %) and the rate of HDR (from 14 to 76 %) indicates the remainder of genetic events inducing gene inactivation is most probably mediated by NHEJ. According to our results, the ratio of contribution of HDR to NHEJ in DNA repairs seems to be locus specific, where *CAN*1 and *GSY*1 loci are mostly repaired by HDR and NHEJ, respectively.

The above results incited us to further characterize and compare the genetic events caused by the NHEJ on the different genetic targets. To do so, we Sanger sequenced the PCR amplicons of target genes that were not associated with typical deletions. Consistent with previous reports, the events responsible for phenotypic loss are equally due to thymine insertions at DSB sites by NHEJ (44 %) and to Microhomology-Mediated End Joining (MMEJ) between the DNA strands upstream and downstream to the DSB causing microdeletions or between those and the salmon DNA (56 %) that is supplemented in transformation mixture (Fig. 5**C**) causing microinsertions.

Discussion

The engineering of promoters has been widely approached in

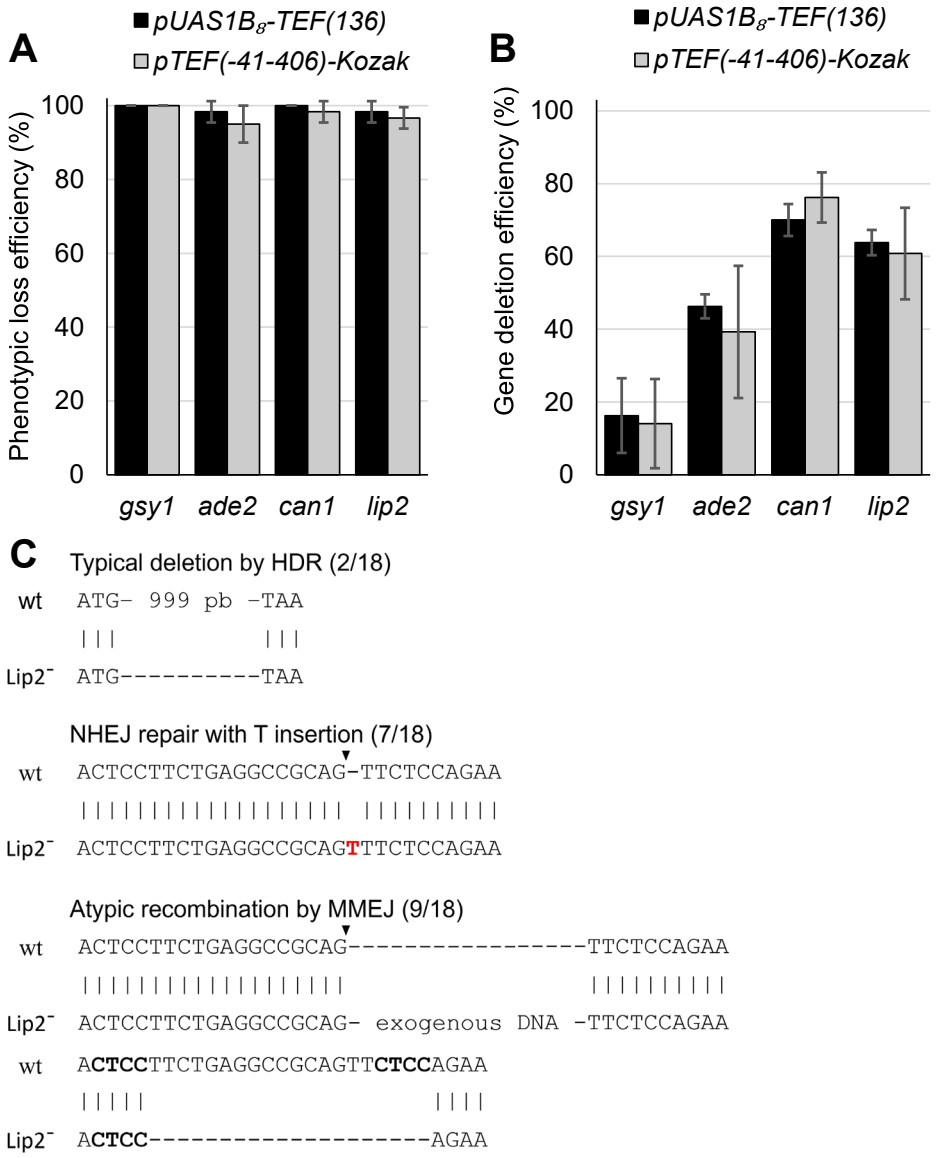

Fig. 5. Gene knock-out efficiency at chromosomal site-specific loci using Cas9 expressed under different promoter designs. (**A**) Phenotype losses (Gsy1$^-$, Ade2$^-$, Mfe1$^-$, Can1$^-$ and Lip2$^-$) were detected based on phenotypic assays detecting visible changes related to loss of gene function. The p*UAS1B*$_8$*-TEF*(136) and the p*TEF*(-41–406)-Kozak leads to the same frequency of phenotype loss which is approximately of 100 % among all phenotypic targets. (**B**) Gene deletion efficiencies were assessed on *GSY*1, *ADE*2, *CAN*1 and *LIP*2 with their respective deletion allele of 100 bp. Transformants done in biological replicates for each locus were screened using multiplex-PCR. For the three genetic targets, no differences in efficiency were observed when using either the p*UAS*1*B*$_8$*-TEF*(136) and the p*TEF*(-41–406)-Kozak. (**C**) Lip2$^-$ mutants were sequenced using the primers 508(fwd) and 509(rev). Sequencing revealed frameshift in non-HDR of *lip*2 mutant due to thymine insertion by NHEJ and to exogenic insertion and microdeletion by MMEJ.

Saccharomyces cerevisiae to deepen the fundamental knowledge on transcriptional elements controlling promoters, but also to widen the array of genetic tools for biotechnological purposes (Feng and Marchisio 2021). Among the work done on synthetic promoters having better transcriptional control and strength, the group of Blazeck et al. (2012) developed a library of promoters in *S. cerevisiae*, amongst them, a hybrid constituve p*UAS*$_3$*-GPD* showed to be one of the strongest promoter in the yeast. With the involvement of non-conventional yeasts, like *Y. lipolytica*, in areas of biotechnology, as in pharmaceuticals, biofuel and food industries, there is a need to extend these promoter engineering advances to this important yeast.

In this study, we aimed at engineering a new strong constitutive *TEF*-based promoter in replacement of the currently used p*UAS*1*B*$_8$*-TEF*(136) to drive high expression of Cas9 to improve the efficiency of CRISPR-Cas9 mediated genome editing in *Yarrowia lipolytica*. We rationally designed a new small-sized synthetic promoter, p*TEF*(136), in which we removed 40 bp from the 3′-end of *TEF* promoter as it contained transcription factor binding sites (TFBS) predicted to have potentially negative regulatory roles in transcription of downstream genes. This resulted in the generation of a more compact constitutive promoter retaining the core promoter regions, namely, the TATA element and the Transcription Start Site (TSS), essential downstream as well as proximal and distal upstream regions.

In contrast to the predominant 5′-truncation done to develop compact and expressive form of native promoters in different organisms (Wang et al., 1999, Blazeck et al., 2011, Jiang et al., 2018, Kanjo et al., 2019, Gao et al., 2021), we went with a 40 bp 3′ truncation with the rationale of removal of negatively regulating TFBS. A 3′-truncation of

*GPD*1 and *CYR*1 promoters was previously adopted in *S. cerevisiae* and led to positive effects, as stronger tolerance to osmotic stress, temperature, and ethanol (Ding et al., 2013; Hong et al., 2018) although the expression strength of such truncated promoter was nearly equivalent to that of the native promoter. Similarly, our 3′-truncated p*TEF*(-41–406) exhibited a lower of expression compared to the p*TEF*(406). The addition of an artificial Kozak sequence at the 3′-end of p*TEF*(-41–406), whose native equivalent had been removed with the 40-bp 3′-truncation, partially recovered (from 45 % to 80 %) the promoter strength relative to the p*TEF*(406). On the other hand, the additional Kozak fused to the p*TEF*(406) in the form of the p*TEF*(406)-Kozak lead to decrease in overall promoter strength relative to the p*TEF*(406). This is most probably due to the presence of two conflicting tandem Kozak sequences, the native and the artificial downstream to it, hindering the efficient initiation of mRNA translation. The reduction of p*TEF* promoter strength upon removal of both *MED*8 and *GCN*4 TFBS, that were part of 3-truncated region of p*TEF*, indicates their potential positive regulatory role in p*TEF* transcription.

Surprisingly, the p*UAS1B*$_8$*-TEF*(136) showed to be the least efficient promoter in terms of expression under our experimental conditions. According to our analysis of the structure of the promoter p*TEF*(406), the minimal core promoter p*TEF*(136) lacks the distal region and part of the proximal region, which might partly explain the reason why the p*UAS*1*B*$_8$*-TEF*(136) showed to be of significantly lower strength than p*TEF*(-41–406) and p*TEF*(406) that harbor intact distal and proximal regions. The absence and/or alteration of these regions might be the major cause of the lower strength of the p*UAS*1*B*$_8$*-TEF*(136) promoter, despite the addition of 8 tandem *UAS*1B activating sequences. According to the expression assays of the group of Blazeck et al. (2011), the expression level of p*UAS*1*B*$_8$*-TEF*(136) is similar to that of p*TEF*(406), yet, both were carried on plasmids. It is noteworth to mention that Blazeck et al. (2011) noted that, under such plasmid-mediated expression, variable and unstable episomal maintenance hindered efficient measurement of expression, which was also reported elsewhere in the literature (Jensen et al., 2014, Inoue et al., 2017) and by our preliminary results (data not shown). Furthermore, recent report measured promoters' strength in *Y. lipolytica* from both plasmids and integration cassettes under different conditions and showed that the p*UAS*1*B*$_8$*-TEF* (136) had equivalent or lower expression strength than the p*EXP*1 (Georgiadis et al., 2023), which was shown by Blazeck et al. (2011) to be of equal strength to p*TEF*(406). While considering that chromosomally integrated cassettes can be subject to variations in expression strength in different insertion sites, due to variable surrounding genetic environment, and in different culture media, our *pTEF*(-41-406)-Kozak was shown to be a stronger promoter under our experimental conditions.

Our results are concordant with previous work that positively correlated a higher CRISPR-Cas9 mediated genome editing efficiency with higher Cas9 expression with strong promoters (Mikami et al., 2015a, Mikami et al., 2015b, Peng et al., 2015, Weninger, Hatzl et al. 2016, Numamoto, Maekawa et al. 2017, Vogl, Kickenweiz et al. 2018, Zheng, Qi et al. 2020, Grützner, Martin et al. 2021, Boisramé and Neuvéglise 2022, Huang, Joshi et al. 2022); higher Cas9 expression under our p*TEF*(-41–406)-Kozak resulted in to higher genome editing efficiency compared to Cas9 under the p*UAS*1*B*$_8$*-TEF*(136) as seen with the chromosomal gene integration at *xpr*2 site. The rate of gene insertion at *axp*1 was already at 100 % when using Cas9 under p*UAS*1*B*$_8$*-TEF*(136) which limited the demonstration of any potential positive effects of p*TEF* (-41–406)-Kozak. Similarly, a nearly 100 % frequency of phenotype loss was observed in targeted gene deletions experiments using Cas9 expressed under any of the tested promoters, entailing a high rate of DNA cleavage, DSB, by Cas9. The implicated mechanisms of DNA repair, HDR, NHEJ or MMEJ, in gene deletion experiments showed to be site-specific and to vary probably according to the accessibility to different genomic locations which is dependent on genomic environment around the target site, e.g. chromatin structures, densities and complexities (Daer et al., 2017; Schwartz et al., 2017b). This might explain why maximal editing efficiency was observed with weak Cas9 expression under p*UAS*1*B*$_8$*-TEF*(136) inducing gene integrations at *axp*1, a chromosomal locus thought to be highly accessible. It is worthy to mention that the strength of Cas9 expression is not supposed to shift the equilibrium between different DNA repair mechanism, HDR or NHEJ. The proportion of whichever DNA repair mechanism is locus-specific as witnessed in *GSY*1 and *CAN*1 sites that were repaired by NHEJ and HDR, respectively. It may be doubted that increasing Cas9 expression might be associated with increased off-target indels in the genome. Although this might be of concern in clinical contexts like in CRISPR-mediated gene therapy, yet, it is of lower concern in industrial microbiology contexts. Accordingly, although off-target effects caused by overexpression of Cas9 can not be excluded, they are not expected to cause problems as long as desired functional mutants are obtained. Moreover, occurrence of off-target effects is expected to be minimal in our experimental design where Cas9 is expressed transiently from plasmids that are cured out of the cells, rather than be chromosmally integrated and constitutively expressed. Moreover, it has been previously reported that off target effects are more attributed to bad gRNA design or to continuous expression of Cas9 (Modrzejewski et al., 2020), rather than strong expression of Cas9.

Conclusions and perspective

This study report a new shorter 3′-truncated *TEF* promoter, the p*TEF* (-41–406)-Kozak that demonstrated an increased promoter activity to its predecessor 5′-truncated hybrid *TEF* that is the p*UAS*1*B*$_8$*-TEF*(136). The p*TEF*(-41–406)-Kozak also retains 60 % of the core p*TEF*(406), one of the already known strongest promoter in *Y. lipolytica*, yet in a more compact form. We showed that the p*TEF*(-41–406)-Kozak can be used as promoter for driving the expression of Cas9 for efficient CRISPR-mediated genome editing, although the current bottleneck seems to arise from a weaker HDR systems relative to a favored NHEJ DNA repair in *Y. lipolytica*.

As to overcome the drawbacks of NHEJ disruption to enhance HDR in *Y. lipolytica*, Schwartz et al. (2017a) have proposed an alternative solution to this problem with transient inhibition of Ku70/80 genes by transcriptional interference using CRISPRi. This strategy makes use of the advantage of better HDR frequency introduced by CRISPR while eliminating the drawbacks associated with the complete disruption of NHEJ. We think that a combinatory approach using strongly expressed Cas9 and transient inhibition of Ku70/80 could enable the full potential of genome editing in *Y. lipolytica* and in other non-model yeasts.

This study is the first to conduct rational 3′-truncation in *TEF* promoter. It can be extended for the engineering of other promoters in yeast under the ultimate goals to generate small-sized synthetic biology bioparts for convenient gene expression in biological systems.

Value, impact, significance

This study demonstrates the rational development of a new synthetic 3′-truncated *TEF*1α promoter that is one third the size of yet comparable in strength to the native *TEF*1α promoter. The 3′-truncated *TEF*1α, as driver to Cas9 expression, showed much superior CRIPR-Cas9 genome editing efficiency compared to the currently used synthetic hybrid promoters, p*UAS*1*B*$_8$*-TEF*(136). Higher strength of Cas9 expression correlates with higher genome editing efficiency in a site-specific manner indicating a possible implication of epigenetic factors.

Authors' contributions

BO conducted the experiments and data analysis, data processing, statistical analyses, and data visualization and wrote the draft of the manuscript. AMAM conceived the work plan, constructed pAS plasmid series, supervised the research work, analyzed data, wrote and edited the manuscript and provided funding and resources. All authors read

and approved the manuscript.

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.

Data availability

No data was used for the research described in the article.

Acknowledgement

This work was funded by NSERC-Discovery Grant (RGPIN-2022-05307) and Fonds de recherche du Québec–Nature et technologies (FRQNT)-Établissement de la Relève Professorale (Application no. 2022-NC-298442), to AMAM. BO was a MSc student under the supervision of AMAM and was funded by Fonds de recherche du Québec - Nature et technologies (FRQNT)-MSc scholarship.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crbiot.2023.100147.

References

Abdel-Mawgoud, A.M., Markham, K.A., Palmer, C.M., Liu, N., Stephanopoulos, G., Alper, H.S., 2018. Metabolic engineering in the host *Yarrowia lipolytica*. Metabolic Engineering 50, 192–208.

Abdel-Mawgoud, A.M., Stephanopoulos, G., 2020. Improving CRISPR/Cas9-mediated genome editing efficiency in *Yarrowia lipolytica* using direct tRNA-sgRNA fusions. Metabolic Engineering 62, 106–115.

Bhat, P.J., Murthy, T.V., 2001. Transcriptional control of the GAL/MEL regulon of yeast *Saccharomyces cerevisiae*: mechanism of galactose-mediated signal transduction. Molecular Microbiology 40 (5), 1059–1066.

Blanchin-Roland, S., Cordero Otero, R.R., Gaillardin, C., 1994. Two upstream activation sequences control the expression of the *XPR2* gene in the yeast *Yarrowia lipolytica*. Molecular Cell. Biology 14 (1), 327–338.

Blazeck, J., Liu, L., Redden, H., Alper, H., 2011. Tuning gene expression in *Yarrowia lipolytica* by a hybrid promoter approach. Applied and Environmental Microbiology 77 (22), 7905–7914.

Blazeck, J., Garg, R., Reed, B., Alper, H.S., 2012. Controlling promoter strength and regulation in *Saccharomyces cerevisiae* using synthetic hybrid promoters. Biotechnol and Bioeng 109 (11), 2884–2895.

Boisramé, A., Neuvéglise, C., 2022. Development of a vector set for high or inducible gene expression and protein secretion in the yeast genus *Blastobotrys*. J Fungi (basel) 8 (5).

Borsenberger, V., Onésime, D., Lestrade, D., Rigouin, C., Neuvéglise, C., Daboussi, F., Bordes, F., 2018. Multiple parameters drive the efficiency of CRISPR/Cas9-induced gene modifications in *Yarrowia lipolytica*. Journal of Molecular Biology 430 (21), 4293–4306.

Buschlen, S., Amillet, J.M., Guiard, B., Fournier, A., Marcireau, C., Bolotin-Fukuhara, M., 2003. The *S. Cerevisiae* HAP complex, a key regulator of mitochondrial function, coordinates nuclear and mitochondrial gene expression. Comparative and Functional Genomics 4 (1), 37–46.

Bzymek, M., Lovett, S.T., 2001. Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. Proceedings of the National Academy of Sciences of the United States of America 98 (15), 8319–8325.

Cha, S., Hong, C.P., Kang, H.A., Hahn, J.S., 2021. Differential activation mechanisms of two isoforms of Gcr1 transcription factor generated from spliced and un-spliced transcripts in *Saccharomyces cerevisiae*. Nucleic Acids Research 49 (2), 745–759.

Chaves, R.S., Herrero, P., Moreno, F., 1999. Med8, a subunit of the mediator CTD complex of RNA polymerase II, directly binds to regulatory elements of *SUC2* and *HXK2* genes. Biochemical and Biophysical Research Communications 254 (2), 345–350.

Chen, X.J., Wang, X., Kaufman, B.A., Butow, R.A., 2005. Aconitase couples metabolic regulation to mitochondrial DNA maintenance. Science 307 (5710), 714–717.

Chiang, Y.C., Komarnitsky, P., Chase, D., Denis, C.L., 1996. ADR1 activation domains contact the histone acetyltransferase GCN5 and the core transcriptional factor TFIIB. The Journal of Biological Chemistry 271 (50), 32359–32365.

Daer, R.M., Cutts, J.P., Brafman, D.A., Haynes, K.A., 2017. The impact of chromatin dynamics on Cas9-mediated genome editing in human cells. ACS Synthetic Biology 6 (3), 428–438.

Ding, W.T., Zhang, G.C., Liu, J.J., 2013. 3' Truncation of the *GPD1* promoter in *Saccharomyces cerevisiae* for improved ethanol yield and productivity. Applied and Environmental Microbiology 79 (10), 3273–3281.

Dobrowolski, A., Drzymala, K., Rzechonek, D.A., Mitula, P., Mironczuk, A.M., 2019. Lipid production from waste materials in seawater-based medium by the yeast *Yarrowia lipolytica*. Front in Microbiol 10, 9.

Feng, X., Marchisio, M.A., 2021. *Saccharomyces cerevisiae* promoter engineering before and during the synthetic biology era. Biol (basel) 10 (6).

Fournier, P., Abbas, A., Chasles, M., Kudla, B., Ogrydziak, D.M., Yaver, D., Xuan, J.W., Peito, A., Ribet, A.M., Feynerol, C., He, F., Gaillardin, C., 1993. Colocalization of centromeric and replicative functions on autonomously replicating sequences isolated from the yeast *Yarrowia lipolytica*. Proceedings of the National Academy of Sciences of the United States of America 90 (11), 4912–4916.

Gao, Z., van der Velden, Y.U., Fan, M., van der Linden, C.A., Vink, M., Herrera-Carrillo, E., Berkhout, B., 2021. Engineered miniature H1 promoters with dedicated RNA polymerase II or III activity. The Journal of Biological Chemistry 296, 100026.

Georgiadis, I., Tsiligkaki, C., Patavou, V., Orfanidou, M., Tsoureki, A., Andreadelli, A., Theodosiou, E., Makris, A.M., 2023. Identification and construction of strong promoters in *Yarrowia lipolytica* suitable for glycerol-based bioprocesses. Microorganisms 11 (5).

Grützner, R., Martin, P., Horn, C., Mortensen, S., Cram, E.J., Lee-Parsons, C.W.T., Stuttmann, J., Marillonnet, S., 2021. High-efficiency genome editing in plants mediated by a Cas9 gene containing multiple introns. Plant Commun 2 (2), 100135.

Hagerman, R.A., Willis, R.A., 2002. The yeast gene *COQ5* is differentially regulated by Mig1p, Rtg3p and Hap2p. Biochimica et Biophysica Acta 1578 (1–3), 51–58.

Holkenbrink, C., Dam, M.I., Kildegaard, K.R., Beder, J., Dahlin, J., Belda, D.D., Borodina, I., 2018. EasyCloneYALI: CRISPR/Cas9-based synthetic toolbox for engineering of the yeast *Yarrowia lipolytica*. Biotechnology Journal 13 (9), 8.

Holmberg, S., Schjerling, P., 1996. Cha4p of *Saccharomyces cerevisiae* activates transcription via serine/threonine response elements. Genetics 144 (2), 467–478.

Hommelsheim, C.M., Frantzeskakis, L., Huang, M., Ülker, B., 2014. PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications. Science Reports 4, 5052.

Hong, K.Q., Hou, X.Y., Hao, A.L., Wang, P.F., Fu, X.M., Lv, A., Dong, J., 2018. Truncation of *CYR1* promoter in industrial ethanol yeasts for improved ethanol yield in high temperature condition. Process Biochemistry 65, 37–45.

Huang, M.Y., Joshi, M.B., Boucher, M.J., Lee, S., Loza, L.C., Gaylord, E.A., Doering, T.L., Madhani, H.D., 2022. Short homology-directed repair using optimized Cas9 in the pathogen *Cryptococcus neoformans* enables rapid gene deletion and tagging. Genetics 220 (1).

Inoue, F., Kircher, M., Martin, B., Cooper, G.M., Witten, D.M., McManus, M.T., Ahituv, N., Shendure, J., 2017. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. Genome Research 27 (1), 38–52.

Jensen, N.B., Strucko, T., Kildegaard, K.R., David, F., Maury, J., Mortensen, U.H., Forster, J., Nielsen, J., Borodina, I., 2014. EasyClone: method for iterative chromosomal integration of multiple genes in *Saccharomyces cerevisiae*. FEMS Yeast Research 14 (2), 238–248.

Jiang, P., Zhang, K., Ding, Z., He, Q., Li, W., Zhu, S., Cheng, W., Li, K., 2018. Characterization of a strong and constitutive promoter from the Arabidopsis serine carboxypeptidase-like gene AtSCPL30 as a potential tool for crop transgenic breeding. BMC Biotechnology 18 (1), 59.

Joo, Y.J., Kim, J.H., Kang, U.B., Yu, M.H., Kim, J., 2011. Gcn4p-mediated transcriptional repression of ribosomal protein genes under amino-acid starvation. The EMBO Journal 30 (5), 859–872.

Kanjo, K., Surin, S.I., Gupta, T., Dhanasingh, M., Singh, B., Saini, G.K., 2019. Truncated, Strong Inducible Promoter P*mcl1* from *Metarhizium Anisopliae*. 3 Biotech 9(3), 75.

Kretzschmar, A., Otto, C., Holz, M., Werner, S., Hübner, L., Barth, G., 2013. Increased homologous integration frequency in *Yarrowia lipolytica* strains defective in non-homologous end-joining. Current Genetics 59 (1–2), 63–72.

Larroude, M., Rossignol, T., Nicaud, J.M., Ledesma-Amaro, R., 2018. Synthetic biology tools for engineering *Yarrowia lipolytica*. Biotechnology Advances 36 (8), 2150–2164.

Larroude, M., Trabelsi, H., Nicaud, J.M., Rossignol, T., 2020. A set of *Yarrowia lipolytica* CRISPR/Cas9 vectors for exploiting wild-type strain diversity. Biotechnology Letters 42 (5), 773–785.

Lis, M., Walther, D., 2016. The orientation of transcription factor binding site motifs in gene promoter regions: does it matter? BMC Genomics 17, 185.

Lustig, A.J., 1999. The Kudos of non-homologous end-joining. Nature Genetics 23 (2), 130–131.

Madzak, C., 2021. *Yarrowia lipolytica* strains and their biotechnological applications: How natural biodiversity and metabolic engineering could contribute to cell factories improvement. J Fungi (basel) 7 (7).

Madzak, C., Treton, B., Blanchin-Roland, S., 2000. Strong hybrid promoters and integrative expression/secretion vectors for quasi-constitutive expression of heterologous proteins in the yeast *Yarrowia lipolytica*. Journal of Molecular Microbiology and Biotechnology 2 (2), 207–216.

Maniatis, T., Fritsch, E.F., Sambrook, J., 1982. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory.

Matsuoka, M., Matsubara, M., Daidoh, H., Imanaka, T., Uchida, K., Aiba, S., 1993. Analysis of regions essential for the function of chromosomal replicator sequences from *Yarrowia lipolytica*. Molecular & General Genetics: MGG 237 (3), 327–333.

Mikami, M., Toki, S., Endo, M., 2015a. Comparison of CRISPR/Cas9 expression constructs for efficient targeted mutagenesis in rice. Plant Molecular Biology 88 (6), 561–572.

Mikami, M., Toki, S., Endo, M., 2015b. Parameters affecting frequency of CRISPR/Cas9 mediated targeted mutagenesis in rice. Plant Cell Reports 34 (10), 1807–1815.
Mittal, N., Guimaraes, J.C., Gross, T., Schmidt, A., Vina-Vilaseca, A., Nedialkova, D.D., Aeschimann, F., Leidel, S.A., Spang, A., Zavolan, M., 2017. The Gcn4 transcription factor reduces protein synthesis capacity and extends yeast lifespan. Nature Communications 8 (1), 457.
Mo, X., Marzluf, G.A., 2003. Cooperative action of the NIT2 and NIT4 transcription factors upon gene expression in *Neurospora crassa*. Current Genetics 42 (5), 260–267.
Modrzejewski, D., Hartung, F., Lehnert, H., Sprink, T., Kohl, C., Keilwagen, J., Wilhelm, R., 2020. Which factors affect the occurrence of off-target effects caused by the use of CRISPR/Cas: A Systematic review in plants. Frontiers Plant Science 11.
Natarajan, K., Meyer, M.R., Jackson, B.M., Slade, D., Roberts, C., Hinnebusch, A.G., Marton, M.J., 2001. Transcriptional profiling shows that Gcn4p is a master regulator of gene expression during amino acid starvation in yeast. Molecular Cell. Biology 21 (13), 4347–4368.
Nga, B.H., Heslot, H., Gaillardin, C.M., Fournier, P., Chan, K., Chan, Y.N., Lim, E.W., Nai, P.C., 1988. Use of nystatin for selection of tributyrin non-utilizing mutants in *Yarrowia lipolytica*. J of Biotechnol 7 (1), 83–86.
Numamoto, M., Maekawa, H., Kaneko, Y., 2017. Efficient genome editing by CRISPR/Cas9 with a tRNA-sgRNA fusion in the methylotrophic yeast *Ogataea polymorpha*. J of Biosci and Bioeng 124 (5), 487–492.
Ouellet, B., Morneau, Z., Abdel-Mawgoud, A.M., 2023. Nile red-based lipid fluorometry protocol and its use for statistical optimization of lipids in oleaginous yeasts. Appl. Microbiol. Biotechnol. https://doi.org/10.1007/s00253-023-12786-9.
Peng, B., Williams, T.C., Henry, M., Nielsen, L.K., Vickers, C.E., 2015. Controlling heterologous gene expression in yeast cell factories on different carbon substrates and across the diauxic shift: a comparison of yeast promoter activities. Microbial Cell Factories 14, 91.
Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Research 29 (9), e45.
Rajvanshi, P.K., Arya, M., Rajasekharan, R., 2017. The stress-regulatory transcription factors Msn2 and Msn4 regulate fatty acid oxidation in budding yeast. The Journal of Biological Chemistry 292 (45), 18628–18643.
Reese, M.G., 2001. Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. Computers & Chemistry 26 (1), 51–56.
Rosenthal, G.A., 1977. The biological effects and mode of action of L-canavanine, a structural analogue of L-arginine. The Quarterly Review of Biology 52 (2), 155–178.
Schwartz, C., Shabbir-Hussain, M., Blenner, M., Wheeldon, I., 2016. Synthetic RNA polymerase III promoters facilitate high-efficiency CRISPR-Cas9-mediated genome editing in *Yarrowia lipolytica*. ACS Synthetic Biology 5 (4), 356–359.
Schwartz, C., Frogue, K., Ramesh, A., Misa, J., Wheeldon, I., 2017a. CRISPRi repression of nonhomologous end-joining for enhanced genome engineering via homologous recombination in *Yarrowia lipolytica*. Biotechnology and Bioengineering 114 (12), 2896–2906.
Schwartz, C., Shabbir-Hussain, M., Frogue, K., Blenner, M., Wheeldon, I., 2017b. Standardized markerless gene integration for pathway engineering in *Yarrowia lipolytica*. ACS Synthetic Biology 6 (3), 402–409.
Schwartz, C., Cheng, J.F., Evans, R., Schwartz, C.A., Wagner, J.M., Anglin, S., Beitz, A., Pan, W., Lonardi, S., Blenner, M., Alper, H.S., Yoshikuni, Y., Wheeldon, I., 2019. Validating genome-wide CRISPR-Cas9 function improves screening in the oleaginous yeast *Yarrowia lipolytica*. Metabolic Engineering 55, 102–110.
Shabbir Hussain, M., Gambill, L., Smith, S., Blenner, M.A., 2016. Engineering promoter architecture in oleaginous yeast *Yarrowia lipolytica*. ACS Synthetic Biology 5 (3), 213–223.
Trassaert, M., Vandermies, M., Carly, F., Denies, O., Thomas, S., Fickers, P., Nicaud, J.-M., 2017. New inducible promoter for gene expression and synthetic biology in *Yarrowia lipolytica*. Microbial Cell Factories 16 (1), 141.
Traven, A., Jelicic, B., Sopta, M., 2006. Yeast Gal4: a transcriptional paradigm revisited. EMBO Reports 7 (5), 496–499.
Verbeke, J., Beopoulos, A., Nicaud, J.M., 2013. Efficient homologous recombination with short length flanking fragments in Ku70 deficient *Yarrowia lipolytica* strains. Biotechnology Letters 35 (4), 571–576.
Vogl, T., Kickenweiz, T., Pitzer, J., Sturmberger, L., Weninger, A., Biggs, B.W., Köhler, E. M., Baumschlager, A., Fischer, J.E., Hyden, P., Wagner, M., Baumann, M., Borth, N., Geier, M., Ajikumar, P.K., Glieder, A., 2018. Engineered bidirectional promoters enable rapid multi-gene co-expression optimization. Nature Communications 9 (1), 3589.
Wang, Q., Donze, D., 2016. Transcription factor Reb1 is required for proper transcriptional start site usage at the divergently transcribed *TFC6-ESC2* locus in *Saccharomyces cerevisiae*. Gene 594 (1), 108–116.
Wang, D., Fischer, H., Zhang, L., Fan, P., Ding, R.X., Dong, J., 1999. Efficient CFTR expression from AAV vectors packaged with promoters–the second generation. Gene Therapy 6 (4), 667–675.
Weninger, A., Hatzl, A.M., Schmid, C., Vogl, T., Glieder, A., 2016. Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*. J of Biotechnol 235, 139–149.
Zhao, C., Gu, D.Q., Nambou, K., Wei, L.J., Chen, J., Imanaka, T., Hua, Q., 2015. Metabolome analysis and pathway abundance profiling of *Yarrowia lipolytica* cultivated on different carbon sources. J of Biotechnol 206, 42–51.
Zheng, X., Qi, C., Yang, L., Quan, Q., Liu, B., Zhong, Z., Tang, X., Fan, T., Zhou, J., Zhang, Y., 2020. The improvement of CRISPR-Cas9 system with ubiquitin-associated domain fusion for efficient plant genome editing. Frontiers in Plant Science 11, 621.