FEMS Yeast Research, 19, 2019, foy105

doi: 10.1093/femsyr/foy105
Advance Access Publication Date: 26 September 2018
Research Article

RESEARCH ARTICLE

Engineering the architecture of erythritol-inducible promoters for regulated and enhanced gene expression in *Yarrowia lipolytica*

Young-Kyoung Park[1], Paulina Korpys[1,2], Monika Kubiak[1,2], Ewelina Celińska[2], Paul Soudier[1], Pauline Trébulle[1], Macarena Larroude[1], Tristan Rossignol[1,†] and Jean-Marc Nicaud[1,*,‡]

[1]Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France and [2]Department of Biotechnology and Food Microbiology, Poznan University of Life Sciences, ul. Wojska, Polskiego 48, 60-627 Poznan, Poland

*Corresponding author: Micalis Institute, INRA-AgroParisTech, UMR1319, Team BIMLip: Integrative Metabolism of Microbial Lipids, domaine de Vilvert, 78352 Jouy-en-Josas, France. Tel: +33 1 74 07 18 20; jean-marc.nicaud@inra.fr

One sentence summary: This study identified cis-regulatory modules (CRMs) for the *EYK1* and *EYD1* promoters in *Yarrowia lipolytica*, which allowed the development of erythritol-inducible hybrid promoters with practical applications in metabolic engineering and synthetic biology.

†Tristan Rossignol, http://orcid.org/0000-0003-0718-0684
‡Jean-Marc Nicaud, http://orcid.org/0000-0002-6679-972X

ABSTRACT

The non-conventional model yeast *Yarrowia lipolytica* is of increasing interest as a cell factory for producing recombinant proteins or biomolecules with biotechnological or pharmaceutical applications. To further develop the yeast's efficiency and construct inducible promoters, it is crucial to better understand and engineer promoter architecture. Four conserved cis-regulatory modules (CRMs) were identified via phylogenetic footprinting within the promoter regions of *EYD1* and *EYK1*, two genes that have recently been shown to be involved in erythritol catabolism. Using CRM mutagenesis and hybrid promoter construction, we identified four upstream activation sequences (UASs) that are involved in promoter induction by erythritol. Using RedStarII fluorescence as a reporter, the strength of the promoters and the degree of erythritol-based inducibility were determined in two genetic backgrounds: the *EYK1* wild type and the *eyk1*Δ mutant. We successfully developed inducible promoters with variable strengths, which ranged from 0.1 SFU/h to 457.5 SFU/h. Erythritol-based induction increased 2.2 to 32.3 fold in the *EYK1* + wild type and 2.9 to 896.1 fold in the *eyk1*Δ mutant. This set of erythritol-inducible hybrid promoters could allow the modulation and fine-tuning of gene expression levels. These promoters have direct applications in protein production, metabolic engineering and synthetic biology.

Keywords: *Yarrowia lipolytica*; promoter; inducible; erythritol; Golden Gate; gene expression; synthetic biology

INTRODUCTION

Yarrowia lipolytica is an oleaginous yeast species that serves as a non-conventional model organism in research on lipid turnover and bio-oil production (Beopoulos *et al.* 2008, 2009), dimorphic transition and fungal differentiation (Martinez-Vazquez *et al.* 2013), and secretory protein synthesis (Matoba *et al.* 1988; Matoba and Ogrydziak 1989; Boisramé *et al.* 1998; Pignède *et al.* 2000; Nicaud *et al.* 2002). *Y. lipolytica* is also the focus of increasing

Received: 6 June 2018; **Accepted:** 21 September 2018

interest because it can serve as an industrial workhorse in a number of processes (Bankar, Kumar and Zinjarde 2009; Coelho, Amaral and Belo 2010; Groenewald *et al.* 2014). Indeed, *Y. lipolytica* has been used as a biocatalyst in the high-level production of citric acid (Rywińska, Rymowicz and Marcinkiewicz 2010; Holz *et al.* 2011; Rywińska *et al.* 2011), erythritol (Rymowicz, Rywińska and Marcinkiewicz 2009; Carly *et al.* 2017a), aroma compounds (Pagot *et al.* 1998; Gomes, Teixeira and Belo 2010; Celińska, Olkowicz and Grajek 2015), and a number of proteins of diverse origins (Nicaud *et al.* 2002; Madzak 2015, Dulermo et al. 2017).

Given the growing number of research areas in which *Y. lipolytica* has been found to be a model organism of choice, the need for efficient molecular tools dedicated to this species has concomitantly grown. The systematic examination of a specific metabolic phenomenon requires the construction and testing of several genetic variants to obtain useful, well-supported conclusions. Thus, high-throughput techniques that allow broad-scale genetic manipulation and the testing of extensive clone libraries are continuously being developed and adopted. Recently, genetic engineering tools used to manipulate the *Y. lipolytica* genome have greatly grown in number thanks to CRISPR-Cas9 technology (Schwartz *et al.* 2016, 2017; Wong *et al.* 2017) and modular cloning techniques (Leplat, Nicaud and Rossignol 2015; Celińska *et al.* 2017; Larroude *et al.* 2017). Simultaneously, high-throughput screening techniques for evaluating traits of interest have been developed; they include droplet-based microfluidic screening and micro bioreactor culturing (Bordes *et al.* 2007; Leplat, Nicaud and Rossignol 2015; Weizhu *et al.* 2015; Back *et al.* 2016; Beneyton *et al.* 2017).

When carrying out the heterologous overexpression of a given protein or metabolically engineering a pathway of interest, it is crucial to carefully examine and select the regulatory elements driving the expression of the genes to be manipulated. Promoter sequences play a major role: transcription is initiated by harnessing the appropriate transcription factors and polymerase. Thus, not surprisingly, the selection and optimization of promoter sequences is one of the most frequently adopted strategies in the fine-tuning of gene expression. In *Y. lipolytica*, the promoter that natively regulates expression of the *XPR2* gene, which encodes an alkaline extracellular protease, was the first to be examined and remains the most extensively studied (Blanchin-Roland, Cordero Otero and Gaillardin 1994; Madzak *et al.* 1999). This regulatory sequence has been subject to great scrutiny, and its characteristics appear to render it unfit for applications related to industrial protein production or basic research, as it requires very specific conditions for full induction. Nevertheless, the knowledge gained during past studies has allowed researchers to design and develop a strong, hybrid, synthetic promoter that is semi-constitutive (Blanchin-Roland, Cordero Otero and Gaillardin 1994; Madzak, Treton and Blanchin-Roland 2000). It is composed of upstream activation sequences (UASs) and involves a minimal promoter of the *LEU2* gene. It has been incorporated in commercially available YLEX vectors (Yeastern Biotech Co.; Taiwan) and has successfully been used in a large number of applications. In addition to the XPR2-based promoter and its derivatives, several other promoter sequences have been analyzed and described, most notably in a comprehensive study by Müller *et al.* (1998). The functional dissection of pXPR2 allowed the identification of one of its UASs ($UAS1B_{XPR2}$). The hybrid hp4d promoter contains four direct repeats of the 109-bp $UAS1B_{XPR2}$ sequence, which is found upstream from the minimal LEU2 promoter (mLEU2) (Madzak, Treton and Blanchin-Roland 2000). Shabbir Hussain *et al.* (2016) investigated promoter strength by shuffling the constitutive

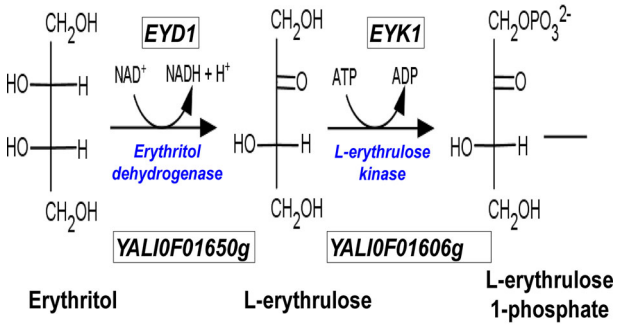

Figure 1. Pathways of erythritol catabolism in *Y. lipolytica*. Erythritol is converted into erythrulose by the erythritol dehydrogenase encoded by *EYD1* (YALI0F01650g). The erythrulose then becomes erythrulose-phosphate via a phosphorylation reaction catalyzed by the erythrulose kinase encoded by *EYK1* (YALI0F1606g) (Carly *et al.* 2017b, 2018).

elements (UAS, proximal promoter, TATA box and core promoter) of various fungal gene promoters (TEF, POX2, LEU2 and PAT1) in *Y. lipolytica*.

In synthetic biology, gene expression must be fine-tuned to ensure optimal flows in related pathways or to avoid metabolic burdens. Cis-regulatory modules (CRMs) are non-coding DNA elements that help regulate gene expression via the binding of transcription factors to motifs in CRM sequences, thus facilitating cell adaptation to internal conditions and the exterior environment. Predicting CRMs is thus a key part of understanding the complex processes underlying cell regulation; it is also necessary if researchers wish to design efficient cellular factories, notably by engineering new promoters with context-specific expression. As indicated in a review by Aerts (Aerts 2012), many computational strategies have been developed throughout the years to identify CRMs. One such strategy—phylogenetic footprinting—exploits the fact that regulatory modules have been evolutionarily conserved among related species. Motifs identified in the promoters of orthologous genes can be tested for functionality, and the corresponding UASs can then be used to construct hybrid promoters.

Recently, the catabolic pathway of erythritol was identified (Fig. 1). It involves the conversion of erythritol into erythrulose, catalyzed by the erythritol dehydrogenase encoded by *EYD1* (YALI0F01650g) (Carly *et al.* 2018), and then the phosphorylation of erythrulose into erythrulose-phosphate, catalyzed by the erythrulose kinase encoded by *EYK1* (YALI0F01606g) (Carly *et al.* 2017b).

Expression of both genes has been shown to be induced by erythritol; the *EYD1* gene displayed 46-fold higher expression on erythritol medium than on glucose medium, a pattern that is similar to the 41-fold increase observed for *EYK1* (Carly *et al.* 2017b, 2018; Carly and Fickers 2018). Consequently, both genes might contain CRMs that respond to erythritol or erythrulose. Two CRMs were identified within the *EYK1* promoter region using sequence conservation among members of the *Yarrowia* clade, which led to the identification of a UAS1-eyk1 motif that responds to erythritol, thus allowing the development of the first erythritol-induced hybrid promoters (Trassaert *et al.* 2017).

However, to engineer complex pathways, a large set of promoters with different strengths and expression profiles is needed. Differential expression in the exponential phase (such as that seen with the constitutive pTEF1) or in the late exponential phase (such as that seen with the promoter hp4d

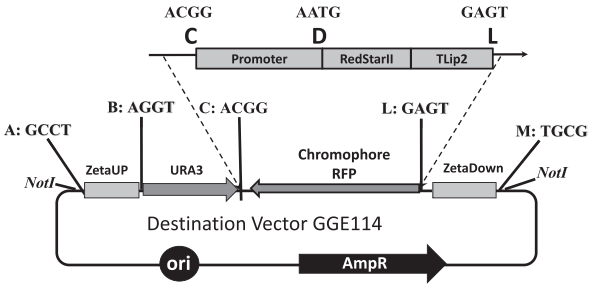

Figure 2. Schematic representation of the Golden Gate assembly technique used to study the promoters. The GG biobricks containing the promoter (overhangs **C** and **D**) were assembled alongside the fragment carrying RedStarII and the Lip2 terminator (overhangs **D** and **L**) and incorporated into the destination vector GGE114. The assembled vector contained the zeta region for expression cassette integration, the *URA3* marker for *Y. lipolytica* selection, and RedStarII as a reporter gene. The chromophore red fluorescent protein RFP was eliminated upon successful cloning of the biobricks. The expression cassette was released via *Not*I digestion.

that contains UAS1B-xpr2) as well as inducible expression could be used to switch on expression at a defined time or to switch off expression upon inducer removal or depletion.

In this study, we identified UASs for *EYK1* and *EYD1* and constructed a set of inducible promoter biobricks useful in Golden Gate assembly (GGAS) in *Y. lipolytica*; gene expression can be regulated by adapting or creating promoters with different behaviors (e.g. with different strengths, expression profiles and degrees of inducibility) with a view to fine-tuning gene expression in *Y. lipolytica*. Here, we constructed expression cassettes using Golden Gate assembly that carried various promoters upstream of a reporter fluorescent protein (RedStarII), which was used to characterize the new promoters.

MATERIALS AND METHODS

Plasmid construction by Golden Gate assembly

Most of the promoter amplicons were cloned using donor vectors (pCR Blunt II TOPO vectors; Thermo Fisher Scientific, Villebon sur Yvette, France), a process that was verified via *Bsa*I digestion and sequencing. Some of the promoters were synthesized and cloned in a donor vector (pUC57) from GeneScript Biotech (New Jersey, US) (see Table 1 and Table S1, Supporting Information). All the primers used to amplify the promoters were designed to have the upstream overhang **'ACGG'** and the downstream overhang **'AATG'** (see Table 2, Fig. 2), which were utilized as part of the Golden Gate assembly process. Other Golden Gate assembly building blocks (destination vector, RedStarII, and Lip2 terminator) were prepared by purifying plasmids from our own GGE collection (Golden Gate *E. coli* collection). The destination vector GGE114, pSB1A3-ZetaUP-URA3-RFP-ZetaDOWN (Table 1) contains the following components: zeta UP, *URA3*ex, RFP (red fluorescent protein, which can be used to generate a red *E. coli* colony) and zeta DOWN, as described in Fig. 2. The promoter names, primer pairs and templates used in PCR are described in Table S1 (Supporting Information). The Golden Gate reaction conditions have been described elsewhere (Celińska *et al.* 2017). The reaction mixture contained a predetermined equimolar amount of each Golden Gate biobrick and of the destination vector (50 pmoles of ends); 1 μL of T4 DNA ligase buffer (NEB); 5 U of *Bsa*I, 200 U of T4; and up to 10 μL of ddH_2O. The following thermal profile was applied: 37°C for 5 min, 16°C for 5 min for 60 cycles, 55°C for 5 min, 80°C for 5 min and 15°C ∞. The reaction mixture was then used for *E. coli DH5α* transformation (Sambrook and Russell 2001). White colonies were screened for the presence of the complete assembly. Afterwards, PCR and restriction enzyme digestion of the plasmids were conducted for verification purposes. All the biobricks were verified by sequencing before the Golden Gate assembly reaction.

Strains, growth media and culture conditions

The *E. coli* and *Y. lipolytica* strains used in the study are described in Table 1. The *EYK1* wild-type (WT) strain, JMY1212 (MatA *ura3-302 xpr2-322*, *LEU2*, zeta platform, derived from Po1d, wild-type for *EYK1*), was used as the basis for characterizing promoters in this study. The *eyk1*Δ strain, JMY7126, which displays a deletion of *EYK1*, was used to examine the inducible expression of promoters in a strain that cannot use erythritol as a carbon source. In this genetic background, erythritol is used as an inducer rather than as a carbon source. Rich medium (YPD) and minimal glucose medium (YNB) were prepared as described below. The YPD medium contained 10 g/L of yeast extract (Difco, Paris, France), 10 g/L of Peptone (Difco, Paris, France) and 10 g/L of glucose (Sigma Aldrich, Saint-Quentin Fallavier, France). The YNB medium contained 1.7 g/L of yeast nitrogen base without amino acids and ammonium sulfate (YNBww; Difco, Paris, France), 10 g/L of glucose (Sigma), 5.0 g/L of NH_4Cl and 50 mM phosphate buffer (pH 6.8). To meet auxotrophic requirements, uracil (0.1 g/L), lysine (0.8 g/L) and leucine (0.1 g/L) were added to the culture medium when necessary. Solid media were created by adding 1.5% agar.

Construction of Y. lipolytica strains

The *eyk1*Δ strain JMY7126 was derived from the *EYK1* WT strain JMY1212, via successive gene deletion (*LYS5* and *EYK1*) and marker rescue. The PUT plasmids (Promoter-*URA3*ex marker-Terminator) were constructed for gene disruption as described in Fickers *et al.* (2003) and Vandermies *et al.* (2017) for *LYS5* and *EYK1*, respectively. The disruption cassettes were prepared by digesting PUT plasmids and used for the transformation of the *Y. lipolytica* strains. Transformants were selected on YNB-leucine or YNB-leucine-lysine medium, depending on genotype. The replicative plasmids (JME547, JME4265) harboring the Cre recombinase gene were used for excising the *URA3*ex marker. Strains from previous promoter studies are described in Table S1 (Supporting Information). The plasmids used in promoter analysis (assembled as described above) were digested by *Not*I, which allowed the expression cassette to be released prior to JMY1212 and JMY7126 transformation. Transformation employed 100 ng of DNA and the lithium acetate method (Le Dall, Nicaud and Gaillardin 1994); transformants were then selected using YNB or YNB-lysine medium, depending on genotype. Florescence tests were carried out for 12 transformants from each construct category, and a representative clone was selected (Table 1).

Microplate growth and florescence analysis

Yarrowia lipolytica pre-cultures were grown overnight in YNBD. They were then centrifuged, washed with an equal volume of YNB medium without a carbon source, and resuspended in 1 mL of the same medium. Microplates (96 well) containing 200 μL of the appropriate medium (final volume) were inoculated with washed cells at an OD_{600nm} of 0.1. YNB medium supplemented with glucose (10 g/L) or erythritol (10 g/L) was used

Table 1. List of strains and plasmids.

Strain	Genotype or description	Reference
E. coli		
DH5α	Φ80*lacZ*Δm15 Δ(*lacZYA-argF*) U169 *recA1 endA1 hsdR17* (r_k^-, m_k^+) *phoA supE44 thi-1 gyrA96 relA1* λ^-	Promega
pUC57	GeneScript Biotech donor vector	GeneScript Biotech
GGE114	pSB1A3-ZetaUP-URA3-RFP-ZetaDOWN	(Celińska *et al.* 2017)
GGE077	pCR4Blunt-TOPO-G1-RedStarII	(Celińska *et al.* 2017)
GGE020	pCR4Blunt-TOPO-T1-3Lip2	(Celińska *et al.* 2017)
GGE085	pCR4Blunt-TOPO-pTEF1	(Celińska *et al.* 2017)
JME547	php4d-Cre_Hyg	(Fickers *et al.* 2003)
JME3267	PUT of LYS5	This study
JME4056	PUT of EYK1 (RIE124)	(Vandermies *et al.* 2017)
JME4265	pTEF-EYK1_hp4d-Cre (RIE132)	(Vandermies *et al.* 2017)
GGE238	pCR4Blunt-TOPO-pEYK1	This study
GGE0130	pCR4Blunt-TOPO-pEYK1-2AB	This study
GGE0104	pCR4Blunt-TOPO-pEYK1-3AB	This study
GGE0132	pCR4Blunt-TOPO-pEYK1-4AB	This study
GGE250	pCR4Blunt-TOPO-pEYK1-5AB	This study
GGE140	pCR4Blunt-TOPO-pEYD1AB	This study
GGE172	pCR4Blunt-TOPO-pEYD1A*B	This study
GGE174	pCR4Blunt-TOPO-pEYD1AB*	This study
JME4417	pUC57-EYK1-4AB-coreTEF	This study
JME4418	pUC57-EYK1-4AB-R1-coreTEF	This study
JME4419	pUC57-EYK1-4AB-R2-coreTEF	This study
JME4420	pUC57-EYK1/EYD1A-coreEYK1	This study
JME4421	pUC57-EYK1/EYD1A-coreTEF	This study
JME4422	pUC57-EYK1/EYD1B-coreEYK1	This study
JME4423	pUC57-EYK1/EYD1B-coreTEF	This study
Y. lipolytica		
JMY195 (Po1d)	*MATA ura3-302 leu2-270 xpr2-322*	(Barth and Gaillardin 1996)
JMY2900	Po1d, Ura^+ Leu^+	(Barth and Gaillardin 1996)
JMY1212	Po1d *lip2*Δ *lip7*Δ *lip8*Δ *LEU2*-ZETA	(Emond *et al.* 2010)
JMY5207	JMY1212 *lys5::URA3 ex*	(Soudier *et al.* unpublished)
JMY7121	JMY1212 *lys5*Δ	(Soudier *et al.* unpublished)
JMY7123	JMY1212 *lys5*Δ *eyk1::URA3 ex*	(Soudier *et al.* unpublished)
JMY7126	JMY1212 *lys5*Δ *eyk1*Δ	(Soudier *et al.* unpublished)

for the growth and florescence analysis. The *eyk1*Δ strain was grown in YNB-lysine medium containing glucose (2.5 g/L) as the carbon source and erythritol (2.5 g/L) as the inducer, as described previously (Trassaert *et al.* 2017). The strains were maintained at 28°C and 110 rpm in a Synergy microplate reader (Biotek, Colmar, France) in accordance with the manufacturer's instructions. OD_{600nm} and red fluorescence were measured every 30 min for 120 h. Red fluorescence was analyzed at the following wavelength settings: excitation at 558 nm and emission at 586 nm. Fluorescence was expressed as mean specific fluorescence value per hour (SFU/h, mean value of SFU per hour). RedStarII fluorescence was expressed in specific fluorescence units per hour. For the RedStarII measurements, no intrinsic fluorescence was detected. Cultures were performed at least in duplicate.

Sequence analysis

The genome sequences of *Yarrowia* species were assembled and annotated by Cécile Neuvéglise, Hugo Devillers and their colleagues (to be published). Homologs of *EYD1* in *Yarrowia* species were identified using BLAST at the private GRYC website (Genome Resources for Yeast Chromosomes; http://gryc.inra.fr) was searched using the *EYD1* gene as a template, as described previously (Carly *et al.* 2018). Promoter regions were retrieved using the download functionality developed by H. Devillers. Multiple alignment of the nucleotide sequences of the *EYK1* and *EYD1* gene promoters among the *Yarrowia* clade (*Y. lipolytica* [YALI], *Y. phangngensis* [YAPH], *Y. yakushimensis* [YAYA], *Y. alimentaria* [YAAL] and *Y. galli* [YAGA]) was then performed using the program Clustal Omega (Larkin *et al.* 2007), which is available at http://www.ebi.ac.uk/Tools/msa/clustalo/. The alignment results highlighted the CRM motifs that have been conserved through evolution and that are thus more likely to have a regulatory function. The conserved motifs were named Box A and Box B. To test their ability to function as UASs, the region containing these motifs plus the 5 to 17 bases on either side of the motifs were selected.

RESULTS

Identification of CRMs within EYK1 and EYD1 promoters

The catabolic pathway of erythritol involves *EYD1* and *EYK1* (Fig. 1), which has been shown to be inducible by erythritol (Carly *et al.* 2017b,2018). We previously reported that the

Table 2. List of primers.

Primer	Sequence	Use
P1 TEF FW	**GGTCTC**T***ACGG***GGGTTGGCGGCG	Amplification for building block construction
P1 TEF RV	**GGTCTC**T***CATT***CTTCGGGTGTGAGTTAC	
P1 EYK FW	**GGTCTC**T***ACGG***CCCATCGATGGAAACCTTAATAGGAGACTACTTCC	
P1 EYK RV	**GGTCTC**T***CATT***GGATCCAGTAGATGTGTAAGTG	
P1 EYD FW	GGGG**GGTCTC**T***ACGG***CCCATCGATGGAAACCTTAATAGGAGACTACTTCC	
P1 EYD RV	CCC**GGTCTC**T***CATT***TGTGTATGTGTGTGTGTGTGTGTG	
EYD UAS1 MluI Fw	CCTTAATAGGAGACTACTTCCGACGCGTAATTAGG	Addition of the MluI site for EYD1 UAS mutation
EYD UAS1 MluI RV	CCTAATTACGCGTCGGAAGTAGTCTCCTATTAAGG	
EYD UAS2 MluI Fw	GAACTCGATACGCGTGCCGTACTCTGGAAA	
EYD UAS2 MluI RV	TTTCCAGAGTACGGCACGCGTATCGAGTTC	
ZetaUp-internal-FW	TATCTTCTGACGCATTGACCAC	Verification of Golden Gate assembly process
URA3-internal-FW	CATCCAGAGAAGCACACAGG	
URA3-internal-RV	CAACTAACTCGTAACTATTACC	
Redstar-internal-FW	AAGACGGTGGCGTTGTTACT	
RedStar-internal-RV	GACTTGCTTCTTGGCCTTGT	
Tlip2-internal-FW	TGCGTTCCTCTAAGACAAATC	
Tlip2-internal-RV	GATTTGTCTTAGAGGAACGCATA	
ZetaDown-internal-RV	GGTAACGCCGATTCTCTCTG	

The bold underlined bases correspond to the *Bsa*1 site; the overhang is in italics.

300-bp EYK1 promoter is not induced on glucose and glycerol media but is induced by erythritol (Trassaert *et al.* 2017). When sequence conservation within the *Yarrowia* clade was examined, two CRMs were identified within the EYK1 promoter region. They were named UAS1-eyk1 (Box A), which had the consensus sequence [**CGGNANCNNNANNNGGAAAGCCG**], and UAS2-eyk1 (Box B), which had the minimal consensus sequence [**CNTGCATNATCCGANGAC**]; both are located upstream from the *SpeI* restriction site (Fig. 3A). In a previous study, thanks to the mutagenesis of the two CRMs (i.e. performed via the introduction of a *MluI* restriction site) and the construction of hybrid promoters, researchers identified a UAS1-eyk1 motif that responded to erythritol, thus allowing the development of the first erythritol-inducible hybrid promoters (Trassaert *et al.* 2017). In the latter study, YFP was used a reporter; however, we have observed that *Y. lipolytica* displays a high degree of auto fluorescence, which depends on growth phase and media composition (Trassaert *et al.* 2017 and unpublished results). Therefore, we now use RedStarII as a reporter.

To identify the regulatory element (i.e. UAS) within the EYD1 promoter region, we analyzed the intergenic region between YALI0F01650g (*EYD1*) and the upstream gene YALI0F01672g, using a similar CRM search. Since this intergenic region was longer than 5500 bp (i.e. 5591 bp; Fig. 4), we analyzed the upstream region using the 800-bp nucleic acid sequence found upstream from *EYD1*. BLAST analysis of the EYD1 promoter did not yield evidence of any conserved motif within the *Y. lipolytica* genome (data not shown). Therefore, we examined how the promoter region of the *EYD1* gene in *Y. lipolytica* compared with that of other species in the *Yarrowia* clade (Fig. 4). This alignment process highlighted the existence of three putative conserved elements within the region 300 bp upstream; these elements were a putative TATA box (Box TATA; GATATAWA) and two CRMs. The first box, which had the main signature (ANTTNNNTTTCCNNATNNGG), was named CRM1-eyd1 (Box A). The second box, which had the main signature (CGGNNCTNNATTGAGAANNC), was named CRM2-eyd1 (Box B) and had a variable number of CA repeats just before the ATG. Like the EYK1 promoter, the EYD1 promoter also had two CRMs, which may also represent motifs required for erythritol and/or erythrulose regulation.

Promoter biobrick construction

Each promoter biobrick was designed and constructed to be compatible with *Y. lipolytica* GGAS, previously described by Celińska *et al.* (2017). First, the presence of internal *BsaI* sites within the promoter sequence was analyzed. Depending on the number of *BsaI* sites, either the sites were eliminated by PCR mutagenesis or promoters were purchased from GeneScript Biotech in the form of synthetic DNA fragments or plasmids. Second, we added *BsaI* sites at both ends of the promoter using PCR and specific overhangs, namely the upstream overhang **C** (ACGG) and the downstream overhang **D** (AATG). Third, we purified the PCR products by gel extraction and cloned them into a TOPO vector (Table 1).

Construction of expression cassettes by Golden Gate assembly for promoter analysis

The assemblies we designed contained different promoter variants; the ORF encoding fluorescent protein RedStarII; and the Lip2 terminator, which were all incorporated using the *BsaI* sites C and D as well as the L overhang (Fig. 2). The three corresponding fragments were assembled with the destination vector GGE114 by adding equimolar concentrations of each fragment type and carrying out a digestion/ligation PCR, as described above. *Escherichia coli* was transformed using the GGAS reaction, and white colonies were selected on LB ampicillin plates. Four positive transformants were screened by colony PCR using the primer pair URA3-internal-FW/RedStar-internal-RV

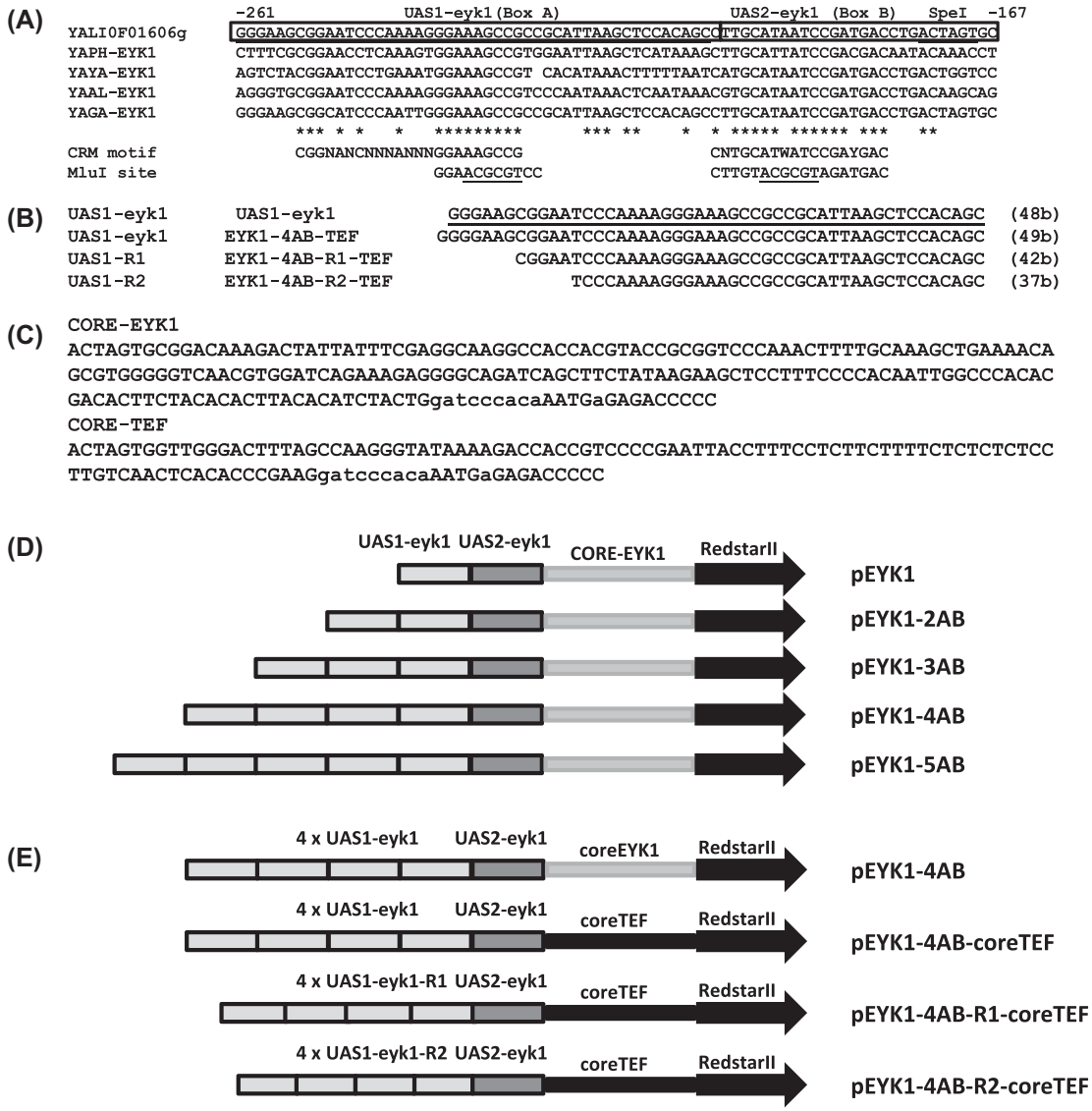

Figure 3. Multiple alignment of the EYK1 promoter containing the two CRMs and a schematic representation of the EYK1 hybrid promoters. **A**, Alignment of the *Y. lipolytica* EYK1 promoter region −261 to −167 containing the two CRMs that act as regulatory elements and control the expression of the *EYK1* gene in response to erythritol (the UAS1-eyk1 and UAS2-eyk1 motifs). The CRMs are indicated with asterisks, and the corresponding CRM consensus sequences are provided. The genomic sequences are from *Y. lipolytica* W29 (YALI; YALI0F01606g), *Y. phangngensis* (YAPH), *Y. yakushimensis* (YAYA), *Y. alimentaria* (YAAL) and *Y. galli* (YAGA). The *SpeI* site is underlined. The underlined *MluI* sites used in the mutation of the CRMs are shown in panel A. The region containing the UAS1-eyk1 motif used for tandem repeat construction is boxed. **B**, Sequence of the UAS1-eyk1 motif (48 bp) used for EYK3AB hybrid promoter construction (Trassaert *et al.* 2017) and the reduced regions used in this study. **C**, EYK and TEF core sequences. D, Schematic representation of the hybrid promoters that contained varying numbers of copies of UAS1-eyk1, which controlled the expression of RedStarII. **E**, Schematic representation of the hybrid EYK-4AB promoter and the hybrid EYK-coreTEF promoters with the reduced UAS1-eyk1 motif that controlled the expression of RedStarII: pEYK1-4AB; pEYK1-4AB-coreTEF (4 tandem copies of 49-bp UAS1-eyk1 + UAS2-eyk1 + coreTEF); pEYK1-4AB-R1-coreTEF (4 tandem copies of 42-bp UAS1-eyk1 + UAS2-eyk1 + coreTEF); and pEYK-4AB-R2-coreTEF (4 tandem copies of 37-bp UAS1-eyk1 + UAS2-eyk1 + coreTEF). The sequences are provided in additional file 1: Table S1 (Supporting Information). N (any base), W (A or T) and Y (C or T) are defined according to IUPAC nucleotide codes (1970).

(Table 2). Plasmids were extracted and verified by PCR using the primer pair URA3-internal-FW/ZetaDOWN-internal-RV and *NotI* digestion. The resulting plasmids are depicted in Table S1 (Supporting Information). The expression cassettes digested by *NotI* were used to generate the *Y. lipolytica EYK1* WT strain (JMY1212) and *eyk1*Δ strain (JMY7126). Ura + transformants were selected on YNB medium supplemented with lysine when necessary. Twelve transformants were selected for further analysis. A representative clone was conserved for comparative studies (Table S1, Supporting Information).

Tandem repeats of $UAS1_{EYK1}$ increase promoter strength in both the EYK1 wild type (JMY1212) and the eyk1Δ strain (JMY7126)

We showed that promoter strength was increased with the hybrid promoter pEYK300A3B, which was composed of three repeats of the 48-bp UAS1-eyk1 (Trassaert *et al.* 2017). Four new hybrid promoters were generated by fusing two, three, four and five UAS1-eyk1 tandem elements taken from the EYK1 promoter, which were named EYK1-2AB, EYK1-3AB, EYK1-4AB and EYK1-5AB, respectively (Fig. 3D). The expression levels and strength

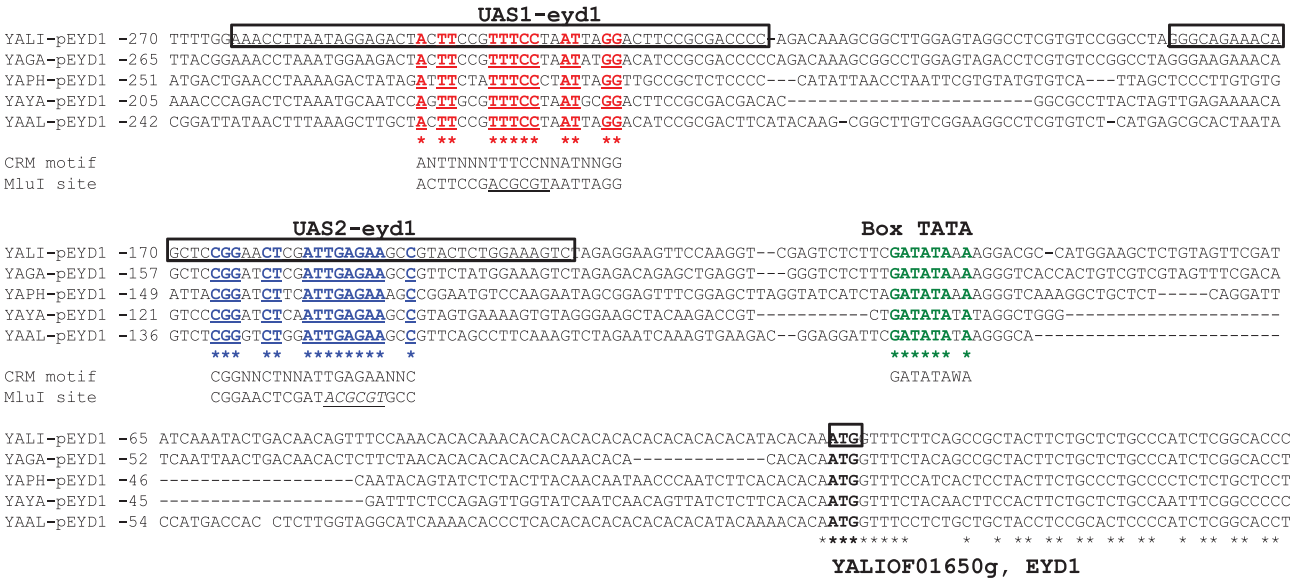

Figure 4. Multiple alignment of the *EYD1* promoter. The alignment of the region between YALI0F01650g (*EYD1*) and the upstream gene YALI0F01672g in *Y. lipolytica* and strains from the *Yarrowia* clade highlights the putative conserved cis-regulatory modules (CRMs) that represent putative regulatory elements for the expression and regulation of the *EYD1* gene by erythritol and erythrulose. The genomic sequences are from *Y. lipolytica* W29 (YALI-pEYD1), *Y. phangngensis* (YAPH-pEYD1), *Y. yakushimensis* (YAYA-pEYD1), *Y. alimentaria* (YAAL-pEYD1) and *Y. galli* (YAGA-pEYD1). The sequences are provided in additional file 1: Table S3 (Supporting Information). The region containing the UAS1-eyd1 and UAS2-eyd1 motifs used for tandem repeat construction is boxed. The nucleic acids that have been conserved in the five species are indicated by a star. The start codon of *EYD1* is indicated as a boxed ATG. The *MluI* sites used in the mutation of the CRMs are shown. In the CRM sequences, N represents any nucleotide.

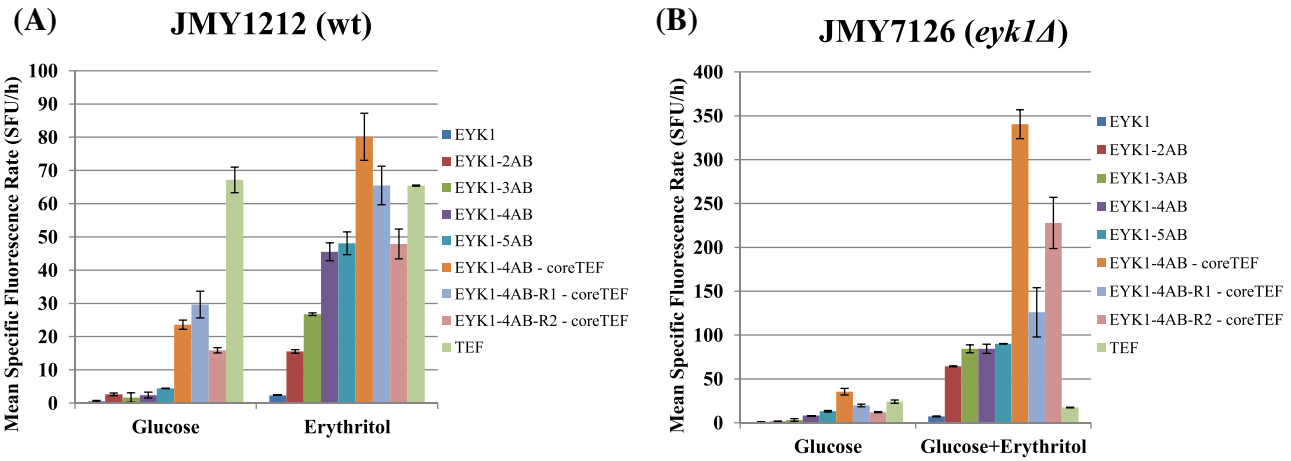

Figure 5. Hybrid EYK1 promoter expression and strength depending on the medium and strain *EYK1* wild type (JMY1212) and *eyk1*Δ mutant (JMY7126). **A**, Results for the *EYK1* wild type, which could use erythritol for growth and **B**, Results for the *eyk1*Δ mutant, which could not metabolize erythritol. Promoter strength was determined by quantifying RedStarII expression and comparing the mean rate of specific fluorescence (SFU/h) obtained when the *EYK1* wild type was grown on erythritol medium or the *eyk1*Δ mutant was grown on glucose + erythritol medium vs. when they were grown on glucose alone.

of the hybrid EYK1 promoters were determined by quantifying RedStarII expression: we determined the mean specific fluorescence rate (SFU/h) of the *EYK1* WT (JMY1212) grown on erythritol and of the *eyk1*Δ strain (JMY7126) grown on glucose + erythritol (results were compared to glucose-only medium; Fig. 5 and Table 3).

In the *EYK1* WT (JMY1212), activity increased slightly concomitantly with UAS1-eyk1 copy number and ranged from 0.54 to 4.42 SFU/h on the glucose medium (Table 3). The SFU rate increased significantly more on the erythritol medium, from 2.28 SFU/h for EYK1 (one copy) to 48.12 SFU/h for EYK1-5AB (five copies). Relative induction also increased, from 4.3 fold to 19.0 fold. Optimal levels were observed for EYK1-4AB. Under these growth conditions, EYK1 displayed low expression levels (0.54 SFU/h) compared to the TEF promoter (67.16 SFU/h). When erythritol was used as an inducer, TEF promoter strength (65.42 SFU/h) was equivalent to that on glucose medium; the strength of EYK1-4AB was comparable—48.12 SFU/h. Thus, when an inducer was present, the EYK hybrid promoter displayed similar activity to the TEF promoter and also had the significant advantage of being inducible.

In *eyk1*Δ strain (JMY7126), activity also increased concomitantly with UAS1-eyk1 copy number, ranging from 0.76 to 13.15 SFU/h on glucose medium (Table 3). The SFU rate increased significantly more on erythritol medium, from 7.13 for EYK1 (one copy) to 90.15 for EYK1-5AB (five copies). Relative

Table 3. Promoter expression and induction levels in the *EYK1* wild type (WT) and the *eyk1*Δ mutant.

	EYK1 WT (JMY1212)			*eyk1*Δ mutant (JMY7126)		
Promoter	Glucose[a]	Erythritol[a]	Fold change[b]	Glucose[a]	Glucose + Erythritol[a]	Fold change[b]
TEF	67.16 ± 3.87	65.42 ± 0.17	1.0	24.11 ± 1.88	17.45 ± 0.39	0.7
EYK1	0.54 ± 0.23	2.28 ± 0.04	4.3	0.76 ± 0.13	7.13 ± 0.51	9.4
EYK1-2AB	2.63 ± 0.38	15.55 ± 0.55	5.9	1.41 ± 0.57	64.48 ± 0.49	45.8
EYK1-3AB	1.68 ± 1.44	26.76 ± 0.38	15.9	3.23 ± 1.39	84.41 ± 4.55	26.1
EYK1-4AB	2.39 ± 0.88	45.50 ± 2.70	19.0	8.18 ± 0.07	84.29 ± 5.21	10.3
EYK1-5AB	4.42 ± 0.09	48.12 ± 3.43	10.9	13.15 ± 0.81	90.15 ± 0.30	6.9
EYK1-4AB-coreTEF	23.57 ± 1.37	80.14 ± 7.06	3.4	35.53 ± 3.73	340.52 ± 16.45	9.6
EYK1–4AB-R1-coreTEF	29.62 ± 4.01	65.50 ± 5.80	2.2	19.72 ± 1.54	125.94 ± 28.09	6.4
EYK1–4AB-R2-coreTEF	15.88 ± 0.76	47.89 ± 4.49	3.0	12.06 ± 0.68	227.84 ± 29.20	18.9

[a]Expressed in SFU/h as described in the materials and methods.
[b]Calculated by comparing the results on erythritol to those on glucose.

induction also increased, from 9.4 fold to 45.8 fold. Optimal levels were observed for EYK1-2AB. On glucose medium, EYK1 displayed low expression levels (0.76 SFU/h) compared to the TEF promoter (24.11 SFU/h). When erythritol was used as an inducer, the TEF promoter displayed slightly reduced strength (17.45 SFU/h), while EYK1-5AB remained strong (90.15 SFU/h). Under such growth conditions and for this strain background (deletion of *EYK1* gene), the performance of the EYK1 hybrid promoter surpassed that of the TEF promoter, as the former was 5.16-fold stronger.

Reduction of the UAS1-eyk1 region

Promoter strength also depends on the core promoter used (Shabbir Hussain *et al.* 2016). We tested hybrid promoters with a TEF core and examined the effect of reducing the size of the UAS1-eyk1 motif (Fig. 3E). We constructed synthetic promoters with different UAS sizes: UAS1-4AB-TEF (four copies of a 69-bp UAS1-eyk1), UAS1-4AB-R1-TEF (four copies of a 62-bp UAS1-eyk1r1) and UAS1—AB-R2-TEF (four copies of a 57-bp UAS1-eyk1r2). When erythritol was used as an inducer, the strength of the EYK1-4AB-coreTEF promoter increased 1.65 fold (80.14 SFU/h vs. 45.50 SFU/h for EYK1-4AB) in the *EYK1* WT (JMY1212) and, more surprisingly, that of the EYK1-4AB-coreTEF promoter increased 4.04 fold (340.52 SFU/h vs. 84.29 SFU/h for EYK1-4AB) in the *eyk1*Δ strain (JMY7126) (Fig. 5 and Table 3). Although we observed an increase in expression levels, induction levels declined (were just 9.6 fold). This result indicates that promoter strength declines when the size of the UAS1-eyk1 motif shrinks, which shows that CRM1 eyk1 extends all the way to the conserved CGG sequence, yielding a general consensus sequence of [**CGGNANCNNNANNGGAAAGCCG**].

Both $UAS1_{EYD1}$ and $UAS2_{EYD1}$ give rise to an inducible promoter in both the EYK1 wild type (JMY1212) and the eyk1Δ strain (JMY7126)

Two putative regulatory elements for the expression and regulation of the *EYD1* gene were found by comparing the upstream DNA sequences of *EYD1* homologs in the *Yarrowia* clade (Fig. 4). The two conserved motifs, CRMa and CRMb, were mutated by introducing a *MluI* site (Fig. 6A). The motif A [ACTTCCGTTTCCTAATTAGG] was replaced by [ACTTCCG*ACGCGT*AATTAGG] and was named A*. The motif B [CGGAACTCGATTGAGAAGCC] was replaced by [CGGAACTCGAT*ACGCGT*GCC] and was named B*. This process yielded the EYD1A*B and EYD1AB* promoters, respectively. Promoter strength and induction levels were compared with those of the EYK1 and EYD1 promoters using the *EYK1* WT (JMY1212) and the *eyk1*Δ mutant (JMY7126) (Fig. 6A, B and Table 4).

In the *EYK1* WT (JMY1212), on glucose medium, the RedstarII expression levels allowed by pEYD1 (0.85 SFU/h) were similar to those allowed by pEYK1 (0.54 SFU/h). The former promoter was also induced by erythritol (11.5 SFU/h, as compared to 2.28 SFU/h for pEYK1) (Table 4 and Fig. 7). The mutation of Box A (EYD1A*B) completely abolished the expression of RedStarII on glucose medium. However, RedStarII continued to be slightly expressed on erythritol (0.16 SFU/h), indicating that CRMa is important for expression and induction. In contrast, the mutation of Box B (EYD1AB*) resulted in just a 2-fold reduction of RedStarII expression on glucose medium (0.43 SFU/h). RedStarII expression levels were higher on erythritol (2.57 SFU/h), indicating that CRMb is less important for expression and induction (Table 4).

In the *eyk1*Δ mutant (JMY7126), unexpected patterns of expression and relative induction were observed on glucose + erythritol medium (Table 4 and Fig. 7). All three promoters, including the mutated ones, showed low levels of expression on the glucose medium (0.5 SFU/h) but higher levels of expression on the glucose + erythritol medium (194.50 to 457.51 SFU/h); a tremendous induction was observed, ranging from 357.6 to 896.1 SFU/h. These results indicate that both CRMa and CRMb are important for expression and induction under these growth conditions and in this genetic background.

Both EYD1 UASA and UASB respond to erythritol

CRMa and CRMb appeared to be involved in *EYD1* expression. To determine their respective role in erythritol-based expression and induction, four hybrid promoters were designed. We used UAS1-eyd1 containing CRMa and UAS2-eyd1 containing CRMb (Fig. 6A). Two hybrid promoters EYK1/EYD1 were designed; they incorporated either four tandem repeats of UAS1-eyd1 or four tandem repeats of UAS2-eyd1 in the place of UAS-eyk1, which gave rise to EYK1/EYD1A-coreEYK1 and EYK1/EYD1B-coreEYK1, respectively (Fig. 6C). Two additional hybrid promoters were designed using a TEF core, which gave rise to EYK1/EYD1A-coreTEF and EYK1/EYD1B-coreTEF (Fig. 6C). These expression cassettes were introduced into the *EYK1* WT (JMY1212) and the *eyk1*Δ mutant (JMY7126) (Table S1, Supporting Information). In *EYK1* WT (JMY1212), UAS1-eyd1 allowed efficient expression of RedStarII in erythritol medium (66.94 SFU/h, with a 6.8-fold change

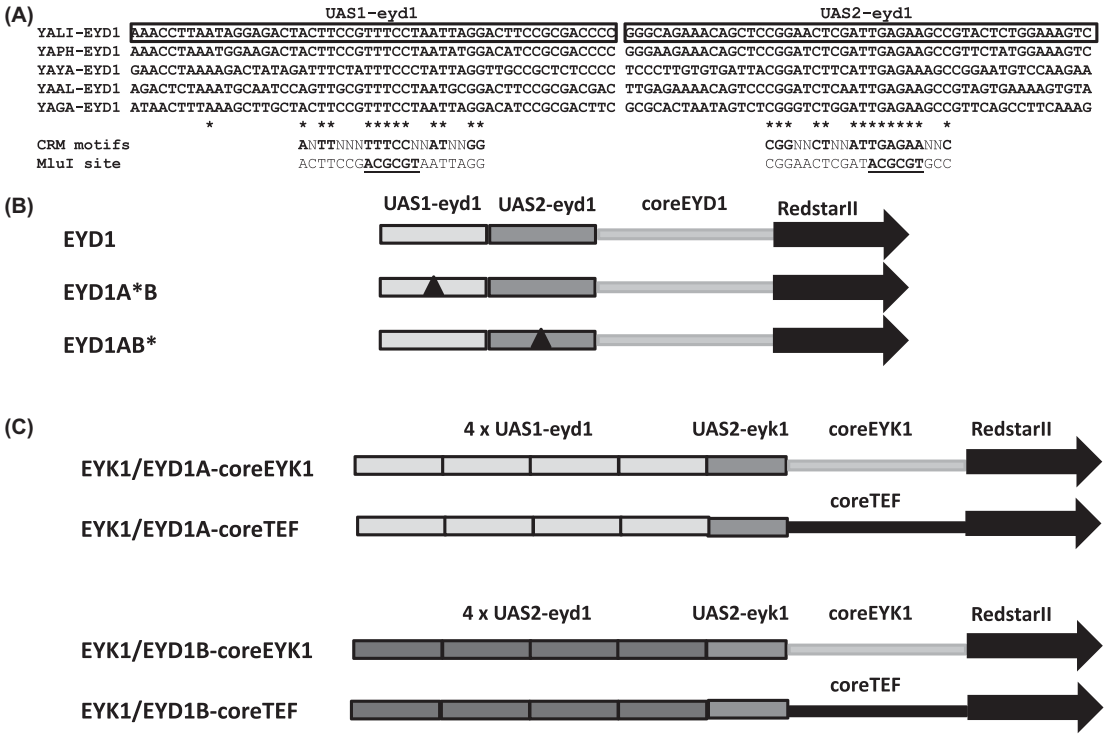

Figure 6. Multiple alignment of the *EYD1* UAS and a schematic representation of the mutated and hybrid promoters used in this study. **A**, Multiple alignment of the UAS1-eyd1 and UAS2-eyd1 motifs of the *EYD1* promoter for *Y. lipolytica* and strains in the *Yarrowia* clade. The CRMs are indicated with asterisks, and the corresponding CRM consensus sequences are provided. The region containing the UAS1-eyd1 and UAS2-eyd1 motifs used in tandem repeat construction is boxed. **B**, Schematic representation of the wild-type EYD1 promoter (EYD1) and the mutated EYD1 promoters within the UAS1-eyd1 (EYD1A*B) and the UAS2-eyd1 (EYD1AB*) motifs that controlled the expression of RedStarII. The *MluI* site used in the mutation of the CRM is shown in panel A. **C**, Schematic representation of the hybrid EYD1/EYK1 promoters containing either the EYK1 or the TEF core promoter; EYK1/EYD1A-EYK1, four tandem repeats of UAS1-eyd1 + UAS2-eyk1 + coreEYK1; EYK1/EYD1A-TEF, four tandem repeats of UAS1-eyd1 + UAS2-eyk1 + coreTEF; EYK1/EYD2-EYK1, four tandem repeats of UAS2-eyd1 + UAS2-eyk1 + coreEYK1; or EYK1/EYD1B-TEF, four tandem repeats of UAS2-eyd1 + UAS2-eyk1 + coreTEF.

Table 4. Strength of different promoters in the *EYK1* wild type (WT) and the *eyk1*Δ mutant.

	EYK1 WT (JMY1212)			*eyk1*Δ mutant (JMY7126)		
Promoter	Glucose[a]	Erythritol[a]	Fold change[b]	Glucose[a]	Glucose + Erythritol[a]	Fold change[b]
EYD1AB	0.85 ± 0.54	11.50 ± 0.25	13.4	0.67 ± 1.52	457.51 ± 11.37	682.5
EYD1A*B	–[c]	0.16 ± 0.32	–	0.54 ± 0.88	194.50 ± 11.50	357.6
EYD1AB*	0.43 ± 1.09	2.57 ± 0.66	5.9	0.27 ± 0.15	245.27 ± 14.56	896.1

[a]Expressed in SFU/h as described in the materials and methods.
[b]Calculated by comparing the results on erythritol to those on glucose.
[c]No fluorescence was detected.

between glucose and erythritol media). In contrast, in both media, low expression levels were observed for the promoters containing the four tandem repeats of UAS2-eyd1 (Fig. 7C and Table 5). In the *eyk1*Δ mutant (JMY7126), both UAS1-eyd1 and UAS2-eyd1 allowed expression of RedStarII in erythritol medium (91.15 SFU/h and 52.57 SFU/h, respectively). This result confirmed that both UAS1 and UAS2 are involved in erythritol induction (Fig. 7D and Table 5). In both strains, exchanging the EYK1 core with the TEF core had a drastic effect on erythritol induction (Table 4) but did not modify expression levels significantly. This result shows that, in this study, the use of a more efficient core promoter did not contribute to the development of inducible promoters.

Promoter expression depend on glucose and erythritol concentration in eyk1Δ

The best expression levels and greatest fold change were obtained with the EYK1, EYK3AB and EYD1 promoters in the *eyk1*Δ mutant. In a previous study (Trassaert *et al.* 2017), expression of EYK1 in the wild-type *EYK1* strain was shown to be modulated by erythritol and erythrulose concentrations in a glycerol medium. To examine how glucose and erythritol concentrations affected promotor expression patterns, RedStarII expression in the *eyk1*Δ mutant was characterized during strain growth on media with two concentrations of glucose, 0.25% and 0.50%, and three concentrations of erythritol, 0%, 0.25% and 0.50% (Figure S1, Supporting Information).

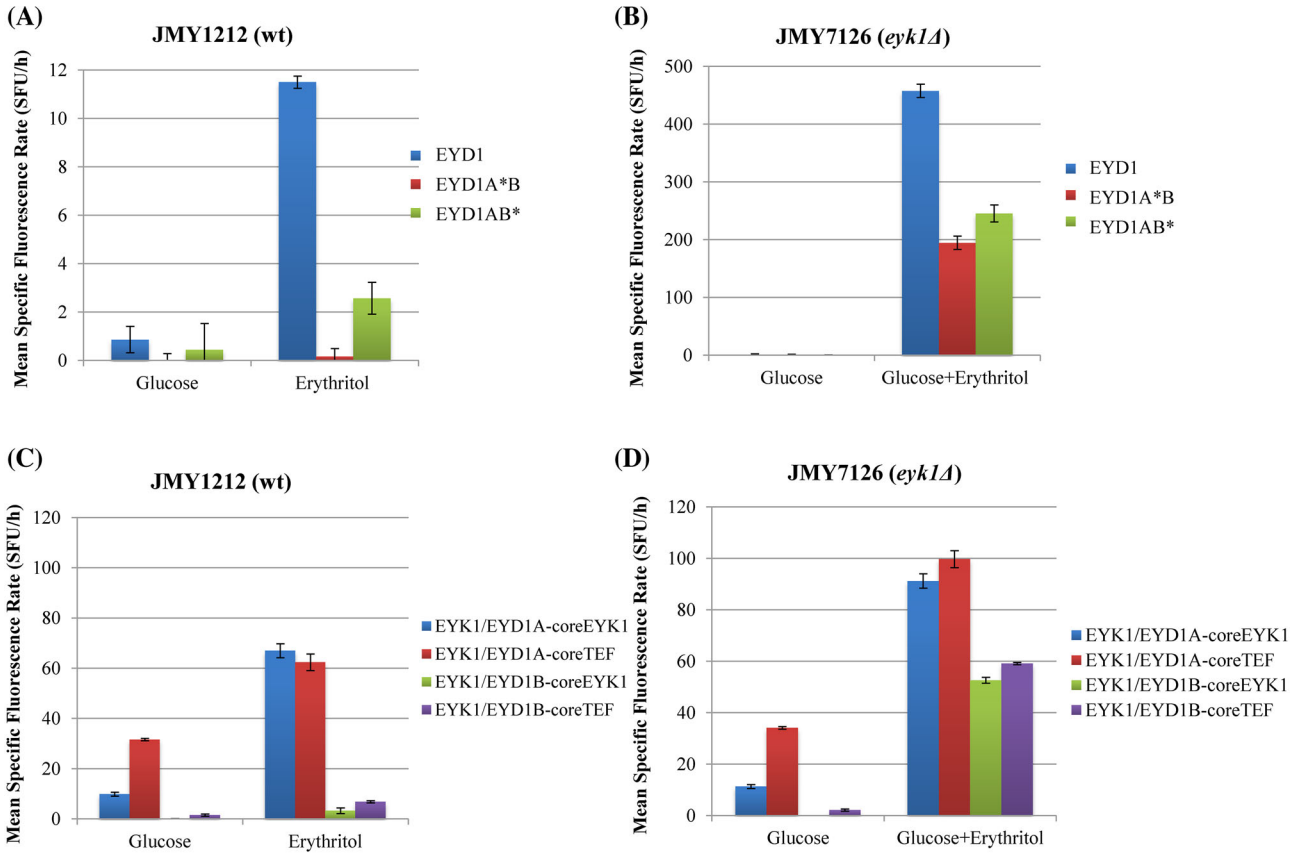

Figure 7. Hybrid EYD1 promoter expression and strength depending on medium and strain *EYK1* wild type (JMY1212) and *eyk1*Δ mutant (JMY7126). **A** and **C**, Results for the *EYK1* wild type, which could use erythritol for growth, and **B** and **D**, results for the *eyk1*Δ mutant, which could not metabolize erythritol. Promoter strength was determined by quantifying RedStarII expression and comparing the mean rate of specific fluorescence (SFU/h) obtained when the *EYK1* wild type was grown on erythritol medium or the *eyk1*Δ mutant was grown on glucose + erythritol medium vs. when they were grown on glucose alone.

Table 5. Promoter strength in the *EYK1* wild type (WT) and the *eyk1*Δ mutant depending on the *EYD1* upstreaming activating sequence (UAS) and core promoter.

	EYK1 WT (JMY1212)			*eyk1*Δ mutant (JMY7126)		
Promoter	Glucose[a]	Erythritol[a]	Fold change[b]	Glucose[a]	Glucose + Erythritol [a]	Fold change[b]
EYK1/EYD1A-coreEYK1	9.85 ± 0.78	66.94 ± 2.81	6.8	11.24 ± 1.82	91.15 ± 8.46	8.1
EYK1/EYD1A-coreTEF	31.57 ± 0.50	62.38 ± 3.30	2.0	34.06 ± 1.65	99.70 ± 17.14	2.9
EYK1/EYD1B-coreEYK1	0.10 ± 0.00	3.23 ± 1.62	32.3	–[c]	52.57 ± 0.76	
EYK1/EYD1B-coreTEF	1.51 ± 0.42	6.81 ± 0.42	4.5	2.11 ± 0.29	59.03 ± 6.00	28.0

[a]Expressed in SFU/h as described in the Materials and Methods.
[b]Calculated by comparing the results on erythritol to those on glucose.
[c]No fluorescence was detected.

On the 0.25% glucose medium containing no erythritol, the promoters were not induced; in contrast, when erythritol was present, there was dose dependent induction (Figure S1 A, C and E, Supporting Information). When the medium contained 0.25% erythritol, fluorescence at 120 h was 816 FU, 7956 FU and 6142 FU for EYK1, EYK3AB and EYD1, respectively. When the medium contained 0.50% erythritol, it reached 1378 FU, 15,018 FU and 11,883 FU, respectively. These results indicate that, in the 0.25% glucose medium, the higher erythritol concentration led to an approximately two-fold increase in fluorescence.

Similar results were observed on the 0.50% glucose medium, although the promoters responded differently. For EYK1 and EYD1, fluorescence at 120 h was lower, regardless of erythritol concentration (0.25% vs. 0.50% erythritol: 542 FU vs. 397 FU for EYK1 and 4512 FU vs. 5212 FU for pEYD1). In contrast, pEYK3AB was less affected by the increase in glucose concentration (0.25% erythritol—11 009 FU and 0.50% erythritol—13 394 FU).

The promoters' rate of fluorescence also varied depending on glucose and erythritol concentrations, making it possible to identify different growth phases (Fig. S1, Supporting Information). On the 0.25% glucose medium with 0.25% erythritol, EYK1, EYK3AB and EYD1 displayed constant fluorescence rates (13.99 FU/h, 136.26 FU/h and 102.92 FU/h, respectively) that lasted for 60 h, 52 h and 34 h, respectively. Duration was greater when the medium contained 0.50% erythritol: 100 h, 84 h and 100 h, respectively. In contrast, on the 0.50%

glucose medium, the fluorescence rate was drastically reduced for EYK1 at both erythritol concentrations (4.17 FU/h and 4.31 FU/h for 0.25% and 0.50% erythritol, respectively), while EYK3AB and EYD1 showed less pronounced differences during phases 2 and 3. These results demonstrate that promoter strength and expression can be modulated by varying glucose and erythritol concentrations.

DISCUSSION

UASs are essential for transcription in yeasts. They must be upstream from the TATA box and transcription start site, but they can be located at variable distances (Buratowski *et al.* 1988). Most often, promoters are studied and regulatory elements are identified by deleting promoters and measuring expression of reporter genes, as exemplified by the research in which the regulatory motifs of *XPR2*, *TEF1* and *POX2* promoters in *Y. lipolytica* were determined (Madzak, Treton and Blanchin-Roland 2000; Blazeck *et al.* 2011; Blazeck *et al.* 2013; Shabbir Hussain *et al.* 2016).

As the number of available genomes increases and the costs of sequencing decrease, researchers can more frequently employ strategies such as phylogenetic footprinting, which is a powerful tool for identifying CRMs with regulatory functions of interest. Recently, genes involved in the catabolism of erythritol were identified in *Y. lipolytica*, namely *EYD1*, which codes for erythritol dehydrogenase, and *EYK1*, which codes for erythrulose kinase. Using the N-terminus sequence of the erythritol dehydrogenase found in *Lipomyces starkeyi*, a BLAST search identified the coding gene ODQ69334.1 in the *L. starkeyi* genome, whose sequence was recently made available. A subsequent BLAST search of the *Y. lipolytica* genome using this gene revealed the *EYD1* gene, which is encoded by YALI0F01650g. Carly and Fickers confirmed that *EYD1* encodes erythritol dehydrogenase (Carly *et al.* 2018). However, *Y. lipolytica* genome mining did not lead to the identification of a gene coding for erythrulose kinase. Instead, this gene was discovered by screening a mutant library for strains unable to grow on erythritol. Sequencing of the mutagenesis cassette insertion site led to the identification of the *EYK1* gene, which is encoded by YALY0F01606g. Carly *et al.* confirmed that this gene encodes erythrulose kinase (Carly *et al.* 2017b). It has been shown that both genes are induced by erythritol (Carly *et al.* 2017b, 2018).

In this study, we employed phylogenetic footprinting within the *Yarrowia* clade to explore the CRMs of the *EYD1* and *EYK1* genes. We used the sequences of *Y. lipolytica* W29, *Y. phangngensis*, *Y. yakushimensis*, *Y. alimentaria* and *Y. galli*. This analysis detected two CRMs, -CRMA-eyd1 and CRMB-eyd1, that occurred within 300 bp of the *EYD1* promoter and two CRMs, CRMA-eyk1 and CRMB-eyk1, that occurred within 300 bp of the *EYK1* promoter; both pairs of CRMs may respond to erythritol. A restriction site was introduced into the most conserved region of the CRMs, leading to a mutation that functionally inactivated the CRMs, abolishing or reducing the response to erythritol. Consequently, the phylogenetic footprinting technique is a very powerful approach for rapidly identifying putative UASs and upstream regulatory sequences. However, it does not reveal the extent of the UASs. Here, when designing hybrid promoters, we defined the UAS as the region containing the CRM plus 5–17 bases to either side.

Thanks to our mutation test, we discovered that both UAS1-eyd1 and UAS2-eyd1 are important for effective expression and induction, regardless of genetic background. Between the conserved motifs A and B of the EYD1 promoter, motif A seemed to be more involved in erythritol-based induction. Trassaert *et al.* (2017) obtained similar results after introducing a mutation into the conserved motifs A (pEYK300aB) and B (pEYK300Ab) of the inducible EYK1 pEYK300 promoter. When grown in minimal YNB medium containing 1% erythritol, the strain carrying the pEYK300A*B-YFP cassette with the mutated motif A displayed a decreased level of YFP expression compared to that of the unmutated pEYK300 (683 and 3536 SFU after 60 h, respectively). In contrast, when motif B was mutated, induction levels were higher under the same conditions (8389 and 3536 SFU after 60 h, respectively).

Expression levels have been found to be dependent on UAS copy number, which have ranged from four tandem copies of UAS1B-xpr2 (Madzak, Treton and Blanchin-Roland 2000) to as many as 32 copies of UAS1B-xpr2 (Blazeck *et al.* 2011; Blazeck, Garg and Alper. 2012). However, this relationship was not observed for the EYK1 and EYD1 hybrid promoters examined in this study. Indeed, we found that an increased number of UAS1-eyk1 copies increased promoter strength when the *EYK1* wild type (JMY1212) was grown on glucose or erythritol (Fig. 6 and Table 3) and that four tandem repeats seemed optimal. Similar results were obtained for the *eyk1*Δ mutant (JMY7126); however, in that strain, optimal expression was reached with three tandem repeats. This result may reflect the titration of the transcription factor: the higher erythritol concentration may result in greater induction, leading to a saturation of expression.

For the hybrid promoter in which the core promoter was exchanged (i.e. EYK1-4AB-coreTEF vs. EYK1-4AB), expression levels were higher, while induction levels were lower. Indeed, when the strong core TEF hybrid promoter was used, expression increased 10 fold and 2 fold, respectively, in the *EYK1* WT (JMY1212) and *eyk1*Δ mutant (JMY7126) grown on glucose. When erythritol was used as an inducer, hybrid promoter strength increased less than when glucose medium was used (two fold in the *EYK1* WT [JMY1212], five fold in the *eyk1*Δ mutant [JMY7126]). It seems that while the core TEF is able to act similarly to the core elements of erythritol-inducible promoter, the strength of its inducible response is less than that of the native EYD1 promoter. The hybrid promoter could be further improved by exchanging the core promoters or by employing a combination of TATA boxes from other inducible promoters (Redden and Alper 2015, Shabbir Hussain *et al.* 2016). Some hybrid promoters of EYK1 and EYD1 promoters used in the *eyk1*Δ mutant (JMY7126) were functionally strong upon induction. For example, the response associated with EYK1/EYD1B-core EYK1 and EYK1/EYD1B-coreTEF displayed a 52-fold and 28-fold increase, respectively (Table 5).

These studies demonstrate that EYK1-4AB provided the best expression levels and the greatest relative induction in the *EYK1* wild type, while EYK1-2AB yielded more optimal expression in the *eyk1*Δ mutant. The EYD1 promoter is a very tight promoter with very low expression levels on glucose media. Its strength is tremendous: ten-fold that of the strong pTEF promoter, with nearly 500-fold greater induction in the *eyk1*Δ strain. Consequently, in the *eyk1*Δ strain, the strength and expression of the EYK1, EYK3AB and EYD1 promoters can be modulated by varying glucose and erythritol concentrations, which generates additional possibilities for promoter fine-tuning.

In this article, we have demonstrated how CRMs can be identified and used to design a broad range of hybrid promoters with applications in metabolic engineering and synthetic biology. These new promoters that respond to erythritol could be

very useful in metabolic engineering, fundamental research and protein expression, as is the case for the Gal1 promoter in *S. cerevisiae*. This may be especially true for the strain containing the deletion in the *EYK1* gene, which allows erythritol to be used as an inducer. This trait is advantageous because erythritol is a cost-effective inducer in the industry. Several industrially relevant proteins such as the Brazzein (a sweetener) and *Candida antarctica* lipase B (CalB) have been successfully expressed using erythritol-inducible hybrid promoters in *Y. lipolytica* (unpublished results). The development of synthetic expression systems will help further improve the production capacity of *Y. lipolytica* in industrial processes.

SUPPLEMENTARY DATA

Supplementary data are available at FEMSYR online.

ACKNOWLEDGEMENTS

We would also like to thank Jessica Pearce for her language editing services.

FUNDING

The characterization of the genome sequence of *Y. galli* was funded by the project CALIN (Carburants Alternatifs et Systèmes d'Injection, grant N° 25 331). The characterizations of the genomes of *Y. yakushimensis, Y. alimentaria* and *Y. phangngensis* were funded by INRA as part of the AIP-Bioressources 2011 program (YALIP project awarded to C. Neuvéglise). Paulina Korpys and Monika Kubiak received internship grants (371/WRiB/2018 and 370/WRiB/2018) given to PULS students, which were co-funded by the EU. Young-Kyoung Park received a PhD scholarship from the Kwanjeong Educational Foundation (KEF). Pauline Trébulle received a PhD scholarship from IDEX Paris-Saclay (ANR-11-IDEX-0003-02).

Conflict of interest. None declared.

REFERENCES

IUPAC-IUB Commission on Biochemical Nomenclature (CBN). Abbreviations and symbols for nucleic acids, polynucleotides and their constituents. *Biochem J* 1970;**120**:449–54.

Aerts S. Computational strategies for the genome-wide identification of cis-regulatory elements and transcriptional targets. *Curr Top Dev Biol* 2012;**98**:121–45.

Back A, Rossignol T, Krier F *et al.* High-throughput fermentation screening for the yeast *Yarrowia lipolytica* with real-time monitoring of biomass and lipid production. *Microb Cell Fact* 2016;**15**:147.

Bankar AV, Kumar AR, Zinjarde SS. Environmental and industrial applications of *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* 2009;**84**:847–65.

Barth G, Gaillardin C. Yarrowia lipolytica. In: Wolf Klaus (ed). Nonconventional Yeasts in Biotechnology: A Handbook. Berlin Heidelberg: Springer, 1996, 313–88.

Beneyton T, Thomas S, Griffiths AD *et al.* Droplet-based microfluidic high-throughput screening of heterologous enzymes secreted by the yeast *Yarrowia lipolytica*. *Microb Cell Fact* 2017;**16**:18.

Beopoulos A, Cescut J, Haddouche R *et al. Yarrowia lipolytica* as a model for bio-oil production. *Prog Lipid Res* 2009;**48**:375–87. doi: https://doi.org/10.1016/j.plipres.2009.08.005.

Beopoulos A, Mrozova Z, Thevenieau F *et al.* Control of lipid accumulation in the yeast *Yarrowia lipolytica*. *Appl Environ Microbiol* 2008;**74**:7779–89.

Blanchin-Roland S, Cordero Otero RR, Gaillardin C. Two upstream activation sequences control the expression of the XPR2 gene in the yeast *Yarrowia lipolytica*. *Mol Cell Biol* 1994;**14**:327–38.

Blazeck J, Garg BR, Alper HS. Controlling promoter strength and regulation in *Saccharomyces cerevisiae* using synthetic hybrid promoters. *Biotechnol Bioeng* 2012;**109**:2884–95.

Blazeck J, Liu L, Redden H *et al.* Tuning gene expression in *Yarrowia lipolytica* by a hybrid promoter approach. *Appl Environ Microbiol* 2011;**77**:7905–14.

Blazeck J, Reed B, Garg R *et al.* Generalizing a hybrid synthetic promoter approach in *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* 2013;**97**:3037–52.

Boisramé A, Kabani M, Beckerich JM *et al.* Interaction of Kar2p and Sls1p is required for efficient co-translational translocation of secreted proteins in the yeast *Yarrowia lipolytica*. *J Biol Chem* 1998;**273**:30903–8.

Bordes F, Fudalej F, Dossat V *et al.* A new recombinant protein expression system for high-throughput screening in the yeast *Yarrowia lipolytica*. *J Microbiol Methods* 2007;**70**:493–502. doi: https://doi.org/10.1016/j.mimet.2007.06.008.

Buratowski S, Steven H, Phillip AS *et al.* Function of a yeast TATA element-binding protein in a mammalian transcription system. *Nature* 1988;**334**:37–42.

Carly F, Vandermies M, Telek S *et al.* Enhancing erythritol productivity in *Yarrowia lipolytica* using metabolic engineering. *Metab Eng* 2017a;**42**:19–24. doi: https://doi.org/10.1016/j.ymben.2017.05.002.

Carly F, Gamboa-Melendez H, Vandermies M *et al.* Identification and characterization of EYK1, a key gene for erythritol catabolism in *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* 2017b;**101**:6587–96. doi: 10.1007 /s00253-017-8361-y. Epub 2017 Jun 12.

Carly F, Steels S, Telek S *et al.* Identification and characterization of EYD1, encoding an erythritol dehydrogenase in Yarrowia lipolytica and its application to bioconvert erythritol into erythrulose. *Bioresour Technol* 2018;**247**:963–9. doi: https://doi.org/10.1016/j.biortech.2017.09.168

Carly F, Fickers P. Erythritol production by yeasts: a snapshot of current knowledge. *Yeast* 2018;**35**:455–63.

Celińska E, Ledesma-Amaro R, Larroude M *et al.* Golden gate assembly system dedicated to complex pathway manipulation in Yarrowia lipolytica. *Microb Biotechnol* 2017;**10**:450–5.

Celińska E, Mariola O, Włodzimierz G. L-Phenylalanine catabolism and 2-phenylethanol synthesis in *Yarrowia lipolytica* -mapping molecular identities through whole-proteome quantitative mass spectrometry analysis. *FEMS Yeast Res* 2015;**15**, doi: https://doi.org/10.1093/femsyr/fov041.

Coelho M, Amaral P, Belo I. *Yarrowia Lipolytica: An Industrial Workhorse*. Spain: Formatex Research Center, 2010; 2.

Dulermo R, Brunel F, Dulermo T *et al.* Using a vector pool containing variable-strength promoters to optimize protein production in *Yarrowia lipolytica*. *Microb Cell Fact* 2017;**16**:31.

Emond S, Montanier C, Nicaud JM *et al.* New efficient recombinant expression system to engineer candida Antarctica Lipase B. *Appl Environ Microbiol* 2010;**76**:2684–7.

Fickers P, Le Dall MT, Gaillardin C *et al.* New disruption cassettes for rapid gene disruption and marker rescue in the

yeast *Yarrowia lipolytica*. *J Microbiol Methods* 2003;**55**:727–37. doi: https://doi.org/10.1016/j.mimet.2003.07.003.

Gomes N, Teixeira JA, Belo I. The use of methyl ricinoleate in lactone production by *Yarrowia lipolytica*: aspects of bioprocess operation that influence the overall performance. *Biocatal Biotransform* 2010;**28**:227–34.

Groenewald M, Boekhout T, Neuvéglise C *et al*. *Yarrowia lipolytica*: safety assessment of an oleaginous yeast with a great industrial potential. *Crit Rev Microbiol* 2014;**40**:187–206.

Holz M, Otto C, Kretzschmar A *et al*. Overexpression of alpha-ketoglutarate dehydrogenase in *Yarrowia lipolytica* and its effect on production of organic acids. *Appl Microbiol Biotechnol* 2011;**89**:1519–26.

Larkin MA, Blackshields G, Brown NP *et al*. Clustal W and Clustal X version 2.0. *Bioinformatics* 2007;**23**:2947–8.

Larroude M, Celińska E, Back A *et al*. A synthetic biology approach to transform *Yarrowia lipolytica* into a competitive biotechnological producer of β-carotene. *Biotechnol Bioeng* 2018;**115**:464–72.

Le Dall MT, Nicaud JM, Gaillardin C. Multiple-copy integration in the yeast *Yarrowia lipolytica*. *Curr Genet* 1994;**26**:38–44.

Leplat C, Nicaud JM, Rossignol T. High-throughput transformation method for *Yarrowia lipolytica* mutant library screening. *FEMS Yeast Res* 2015;**15**, doi: 10.1093/femsyr/fov052.

Madzak C. *Yarrowia lipolytica*: recent achievements in heterologous protein expression and pathway engineering. *Appl Microbiol Biotechnol*. 2015;**99**:4559–77.

Madzak C, Treton B, Blanchin-Roland S. Strong hybrid promoters and integrative expression/secretion vectors for quasi-constitutive expression of heterologous proteins in the yeast *Yarrowia lipolytica*. *J Mol Microbiol Biotechnol* 2000;**2**:207–16.

Madzak C, Blanchin-Roland S, Cordero Otero RR *et al*. Functional analysis of upstream regulating regions from the *Yarrowia lipolytica* XPR2 promoter. *Microbiol* 1999;**145**:75–87.

Martinez-Vazquez A, Gonzalez-Hernandez A, Domínguez A *et al*. Identification of the transcription factor Znc1p, which regulates the Yeast-to-Hypha transition in the dimorphic yeast *Yarrowia lipolytica*. *PLoS One* 2013;**8**:e66790.

Matoba S, Ogrydziak DM. A novel location for dipeptidyl aminopeptidase processing sites in the alkaline extracellular protease of *Yarrowia lipolytica*. *J Biol Chem* 1989;**264**:6037–43.

Matoba S, Fukayama J, Wing RA *et al*. Intracellular precursors and secretion of alkaline extracellular protease of Yarrowia lipolytica. *Mol Cell Biol* 1988;**8**:4904–16.

Müller S, Sandal T, Kamp-Hansen P *et al*. Comparison of expression systems in the yeasts *Saccharomyces cerevisiae*, *Hansenula polymorpha*, *Klyveromyces lactis*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica*. Cloning of two novel promoters *fromYarrowia lipolytica*. *Yeast* 1998;**14**:1267–83. doi: 10.1002/(SICI)1097-0061(1998100)14:14<1267::AID-YEA327>3.0.CO;2-2.

Nicaud JM, Madzak C, Broek P *et al*. Protein expression and secretion in the yeast *Yarrowia lipolytica*. *FEMS Yeast Res* 2002;**2**:371–9.

Pagot Y, Le Clainche A, Nicaud JM *et al*. Peroxisomal β-oxidation activities and γ-decalactone production by the yeast *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* 1998;**49**:295–300.

Pignède G, Wang H, Fudalej F *et al*. Characterization of an extracellular lipase encoded by LIP2 in *Yarrowia lipolytica*. *J Bacteriol* 2000;**182**:2802–10.

Redden H, Alper HS. The development and characterization of synthetic minimal yeast promoters. *Nat Commun* 2015;**6**:7810.

Rymowicz W, Rywińska A, Marcinkiewicz M. High-yield production of erythritol from raw glycerol in fed-batch cultures of *Yarrowia lipolytica*. *Biotechnol Lett* 2009;**31**:377–80.

Rywińska A, Juszczyk P, Wojtatowicz M *et al*. Chemostat study of citric acid production from glycerol by *Yarrowia lipolytica*. *J Biotechnol* 2011;**152**:54–7. doi: https://doi.org/10.1016/j.jbiotec.2011.01.007.

Rywińska A, Rymowicz W, Marcinkiewicz M. Valorization of raw glycerol for citric acid production by *Yarrowia lipolytica* yeast. *Electron J Biotechnol* 2010;**13**. DOI: 10.2225/vol13-issue4-fulltext-1.

Schwartz CM, Shabbir Hussain M, Blenner M *et al*. Synthetic RNA polymerase III promoters facilitate high-efficiency CRISPR-Cas9-Mediated genome editing in *Yarrowia lipolytica*. *ACS Synth Biol* 2016;**5**:356–9.

Sambrook JF, Russell DW. *Molecular Cloning: A Laboratory Manual (3-Volume Set)*. New York:Cold Spring Harbor Laboratory Press. 2001.

Schwartz C, Shabbir-Hussain M, Frogue K *et al*. Standardized markerless gene integration for pathway engineering in *Yarrowia lipolytica*. *ACS Synth Biol* 2017;**6**:402–9.

Shabbir Hussain M, Gambill L, Smith S *et al*. Engineering promoter architecture in oleaginous yeast *Yarrowia lipolytica*. *ACS Synth Biol* 2016;**5**:213–23.

Trassaert M, Vandermies M, Carly F *et al*. New inducible promoter for gene expression and synthetic biology in *Yarrowia lipolytica*. *Microb Cell Fact* 2017;**16**:141.

Vandermies M, Denies O, Nicaud JM *et al*. EYK1 encoding erythrulose kinase as a catabolic selectable marker for genome editing in the non-conventional yeast *Yarrowia lipolytica*. *J Microbiol Methods* 2017;**139**:161–4.

Weizhu Z, Du G, Chen J *et al*. A high-throughput screening procedure for enhancing α-ketoglutaric acid production in *Yarrowia lipolytica* by random mutagenesis. *Process Biochem* 2015;**50**:1516–22. doi: https://doi.org/10.1016/j.procbio.2015.06.011.

Wong L, Engel J, Jin E *et al*. YaliBricks, a versatile genetic toolkit for streamlined and rapid pathway engineering in *Yarrowia lipolytica*. *Metab Eng Commun* 2017;**5**:68–77.