UNIVERSITY OF TARTU
Faculty of Science and Technology
Institute of Technology

Paula Bayón Buján

Characterization of non-native promoters in *Saccharomyces cerevisiae*

Bachelor’s Thesis (12 ECTS)

Curriculum Science and Technology

Supervisor:
Researcher, PhD Steven A. van der Hoek

Tartu 2023

Characterization of non-native promoters in Saccharomyces cerevisiae

Abstract.

Yeast has been a valuable model organism for both industry and research. However, until recently, the use of native promoters in scale-up production has led to strain instability and subsequently, loss of productivity. Previous studies have found TEF promoters to give stable expression levels across culture conditions in addition to being a gene very conserved among yeast strains. Therefore, we investigated the strength of TEF promoters from different organisms by analyzing their expression of GFP on different carbon sources and developed a synthetic biology toolkit containing different strength non-native promoters for *Saccharomyces cerevisiae* to avoid several copies of the same promoter in the genome. *Kluyveromyces marxianus* promoter sequence showed promising results, surpassing our positive control. *Cyberlindnera jardinii* and *Candida parapsilosis* TEF2 exhibited medium strength, and *Candida parapsilosis* TEF1, *Scheffersomyces stipites,* and *Yarrowla lipolytica* promoters demonstrated weaker strength but were still functional in the yeast, performing better than the negative control. These results show that the investigated promoters could be used to engineer cell factories in research and industry.

Keywords:

Biotechnology, Modular Cloning, EasyClone-MarkerFree, non-native promoter, Saccharomyces cerevisiae

CERCS:

T490 Biotechnology

Saccharomyces cerevisiae mittenatiivsete promootorite iseloomustus

Lühikokkuvõte:

Pärm on olnud väärtuslik mudelorganism nii tööstuse kui ka teadustöö jaoks. Kuid viimasel ajal on looduslike promootorite kasutamine suuremahulises tootmises põhjustanud tüve ebastabiilsust ja seejärel tootlikkuse vähenemist. Varasemad uuringud on leidnud, et TEF-i promootorid annavad stabiilse ekspressioonitaseme kõigis kultiveerimistingimustes, lisaks sellele on nad pärmitüvede sees konserveerunud geenid. Seetõttu uurisime erinevate organismide TEF-i promootorite tugevust, analüüsides nende GFP ekspressiooni erinevatel süsinikuallikatel, ja töötasime välja sünteetilise bioloogia tööriistakomplekti, mis sisaldas erineva tugevusega *Saccharomyces cerevisiae* mittenatiivseid promootoreid, et vältida sama promootori mitut koopiat genoomis. *Kluyveromyces marxianus* promootorjärjestus näitas paljutõotavaid tulemusi, ületades meie positiivse kontrolli. *Cyberlindnera jardinii* ja *Candida parapsilosis* TEF2 olid keskmise tugevusega ning *Candida parapsilosis* TEF1, *Scheffersomyces stipites* ja *Yarrowla lipolytica* promootorid näitasid nõrgemat tugevust, aga endiselt toimisis pärmis paremini kui negatiivne kontroll. Need tulemused näitavad, et uuritud promootoreid saab kasutada rakutehaste kavandamiseks teadusuuringutes ja tööstuses.

Võtmesõnad:

Biotehnoloogia, moodulkloonimine, EasyClone-MarkerFree, mittenatiivne promootor, Saccharomyces cerevisiae

CERCS:

T490 Biotehnoloogia

TABLE OF CONTENTS

TERMS, ABBREVIATIONS AND NOTATIONS

CRISPR- Clustered regularly interspaced short palindromic repeats

DSBs- double strand breaks

FACS- fluorescence-activated cell sorting

GRAS- generally recognized as safe

gRNA- guide RNA

HDR- homology-directed repair

LB- Lysogeny broth medium
mRNA- messenger RNA

MoClo- Modular cloning system

TEF- Transcription elongation factor 1-alpha

yeGFP- Yeast enhanced fluorescent protein

YPD- Yeast extract peptone dextrose medium

INTRODUCTION.

Many of the substances and materials of industrial interest are still currently produced by chemical synthesis. However, both environmental concerns and industrial needs strongly push toward sustainable bio-production. The production of chemicals from renewable biomass using microbial cell factories has the potential to change our heavy dependence on fossil resources. Typically, heterologous genes encoding enzymes in biosynthetic pathways are introduced into the microbial host to enable the production of a compound of interest. Subsequently, genetic engineering methods are applied to restructure metabolic networks to overcome these limits. During this process native promoters are used to alter the gene expression of enzymes, but their repeated use leads to strain instability in industrial settings through homologous recombination. Therefore, researchers have focused on the development of synthetic promoters to reduce the risk associated with the repetitive use of native promoters. However, synthetic promoters still contain minimized native DNA sequences that may also lead to genetic rearrangements, are not characterized across multiple carbon sources, and only a handful of synthetic promoters have sufficient transcriptional strength for gene expression.

Therefore, the aim of this study is to characterize non-native promoters that do not share sequence similarity with native yeast promoters and evaluate their performance in different substrates (glucose, sucrose, galactose, and ethanol) so that they could be used in the industry environment. As the promoter sequence of the translation elongation factor 1-alpha (TEF) gene is known for its stable and strong expression levels across substrates in *Saccharomyces cerevisiae*, TEF promoters from other yeast species were selected as non-native promoters with potential for strong and stable expression in *S. cerevisiae*.

To accomplish this, we use a combination of the Modular Cloning and EasyClone-MarkerFree methods, as they offer an efficient and streamlined way of assembling heterologous promoter sequences into the genome of *Saccharomyces cerevisiae*, the industrial host for the production of fuels and chemicals.

The non-native promoters were integrated into the genome to drive the expression of yeast enhanced green fluorescent protein (yeGFP) under their control. In turn, this allowed us to use yeGFP as a read-out for their strength across various substrates to characterize their function.

Further research and optimization of non-native promoters can contribute to the development of robust and efficient cell factories for industrial biotechnology, helping to reduce greenhouse gas emissions and promote sustainable production processes.

1 LITERATURE REVIEW

1.1 Importance of industrial biotechnology and limitations

Industrial biotechnology, which involves the production of fuels and chemicals from bio-based renewable raw materials, has emerged as a response to the increasing need for sustainable and efficient processes due to the limitations of chemical synthesis and natural producers (Borodina & Nielsen, 2014). Chemical routes can be energy-intensive, and environmentally detrimental, while natural producers, like plants, tend to produce their compounds of interest only in amounts needed for their own survival and growth without overproducing them. To overcome these limitations, scientists have turned to the promising concept of cell factories, by engineering organisms to maximize and optimize their biochemical pathways (Cho et al., 2022).

The development of cell factories involves multiple steps, including pathway design, host selection, and optimization of enzymatic activity, and may require further modifications to the metabolism of the host organism. Successful cell factory development requires careful consideration of these factors to achieve efficient production of the desired end product (Li & Borodina, 2015). The total cost of building a cell factory is typically in the range of 50+ million US dollars, and it takes more than 30 person-years (Choi et al., 2019)(Nielsen & Keasling, 2016). Therefore, it needs substantial investment, especially when moving from the proof-of-concept stage, where only small amounts of the desired product are generated, to a final strain capable of meeting certain requirements in titer, rate, and yield, and this final strain must be obtained to make an economically viable process (Nielsen, 2019).

Biobased production processes that rely on renewable feedstocks, such as biomass-derived sugars or agricultural waste, can indeed help reduce reliance on fossil fuels and hence lower greenhouse gas emissions, energy consumption, and waste, as well as improve the efficiency of production (Akbarian et al., 2022). For example, the use of genetically engineered organisms to produce sustainable alternatives, known as green oils, through fermentation processes can help reduce deforestation for palm oil production. (Sheldon, 2016)

Native promoters and terminators are often used in genetic engineering due to their well-characterized regulatory properties and compatibility with the host organism's transcriptional machinery. However, because these native promoters are naturally occurring in the host organism's genome, their repeated use in metabolic engineering can lead to homologous recombination with their endogenous regulatory elements or other overexpressed genes in the heterologous pathway during cell division or genetic recombination events. Homologous recombination is an evolutionarily conserved process by which DNA strands exchange genetic information through the alignment and exchange of similar sequences, leading to the deletion or rearrangement of the inserted DNA. These events become more likely over time as the organism is cultivated across generations, leading to strain instability and reduced productivity of the engineered cell factory, causing increased costs in commercial bioprocesses (Rugbjerg et al., 2018).

1.2 Saccharomyces cerevisiae

S. cerevisiae is a valuable model organism for both industry and research. Its versatility and robustness, coupled with the wide availability of genome engineering tools, make it an exceptional organism for genetic engineering and optimization to produce a wide range of products, including biofuels, enzymes, pharmaceuticals, and chemicals (Nevoigt, 2008). As a result, it is extensively utilized in the production of bioethanol, the largest-volume fermentation product and it is also used to produce several pharmaceuticals such as human insulin, hepatitis vaccines, and human papillomavirus vaccines(Nielsen, 2013).

S. cerevisiae's tolerance to high temperatures, low pH, and high ethanol concentrations makes it an ideal candidate for large-scale fermentation processes (Parapouli et al., 2020), where maintaining productivity and viability despite adverse conditions is crucial for efficient and cost-effective production. Economically, its use in industrial biotechnology processes is often attractive due to its relatively low production costs(Stewart, 2014). Moreover, the great availability of online databases containing its genome information(Cherry et al., 2012), along with the organism's genetic tractability(Parapouli et al., 2020) and being generally recognized as safe (GRAS)(Sewalt et al., 2016), were major factors when using this organism in this study.

1.3 Synthetic promoters

As *S. cerevisiae* is such an important organism in industry, scientists have developed synthetic promoters to combat strain instability. Synthetic promoters are artificially designed DNA sequences that have gained considerable interest in synthetic biology and biotechnology. They offer flexibility in controlling gene expression levels and dynamics, as they can be designed to have specific response characteristics, such as inducibility or tunability, which may not be naturally present in native promoters. This can be particularly useful in optimizing biosynthetic pathways that may require precise regulation of multiple genes (Redden et al., 2015).

However, there are often limited numbers of well-characterized synthetic promoters available (Köbbing et al., 2020), and many biosynthetic pathways can involve a large number of enzymes(Weeks & Chang, 2011). Developing enough synthetic promoters for each enzyme in a pathway can be time-consuming and challenging. Additionally, synthetic promoters may not always perform as expected (Jin et al., 2019), and despite being designed as artificial sequences, many currently used still share sequence similarity with the host organism's genome (Redden & Alper, 2015). This can potentially lead to homologous recombination events as well, resulting in the strain instability previously mentioned. Furthermore, synthetic promoters are often only characterized on glucose or galactose, with no other substrates investigated, while industrial processes use a variety of substrates (Redden & Alper, 2015) (Patel et al., 2022).

1.4 Transcription Elongation Factor

Previous studies have found the transcription elongation factor 1-alpha (TEF) promoter to give stable and strong expression levels across culture conditions (Peng et al., 2015; Schirmaier & Philippsen, 1984). TEF genes are expressed across eukaryotes as their function is to bring tRNAs to the ribosomes (Steiner & Philippsen, 1994). As the TEF gene is ubiquitous among yeast species and its function is required across various cultivation conditions, TEF promoters from other yeast species could potentially drive strong heterologous gene expression in *S. cerevisiae*. In fact, the nourseothricin and G418 resistance cassette used in *S. cerevisiae* makes use of the TEF promoter of *Eremothecium gossypii* (Matanović et al., 2022). Therefore, we decided to investigate TEF promoters from different yeast species, to evaluate non-native promoters for their gene expression potential in *S. cerevisiae*.

1.5 Modular cloning

The direct insert of a promoter into the genome of an organism does not ensure its functionality, it needs to be positioned correctly to drive gene expression. An expression cassette ensures the entry and the efficiency of the integration, as it also contains regulatory elements that aid the translation of the mRNA. The construction of these cassettes carrying the promoter sequence is achieved using the modular cloning system (MoClo) which employs Golden Gate, allowing rapid and efficient assembly of DNA fragments into larger constructs. The MoClo system is based on libraries containing standardized genetic elements, such as promoters, coding sequences, and terminators (Figure 1)(Weber et al., 2011). The Golden Gate method is based on Type IIs restriction enzymes, which cut outside of their recognition site, allowing the user to determine their own customized four base pair overhangs that can be standardized for each genetic element. These genetic elements are then first designed as level 1 modules that can be easily assembled in a single restriction-ligation step. In this system, BbsI assembly is used to insert DNA fragments into a universal entry vector, resulting in ‘part’ plasmids, or level 1 plasmids that are sequenced before subsequent assembly steps. The individual ‘parts’ can then be released from the plasmid using the recognition site of a second Type IIs restriction enzyme, namely BsaI. Each part has defined flanking overhangs. For instance, genes could all be designated part 3 with the 5’ overhang TATG and 3’ overhang ATCC. These overhangs then recombine with complementary overhangs on respectively digested part 2 (promoters) and part 4 (terminators) genetic elements. The complementary overhangs are finally ligated together into a backbone plasmid to form a full expression cassette. To minimize background, antibiotic selection is changed at each round of assembly, such as kanamycin and ampicillin in our case (Lee et al., 2015).

This technique was used in this study to assemble non-native promoters into expression cassettes that will express GFP under the regulation of these same promoters.

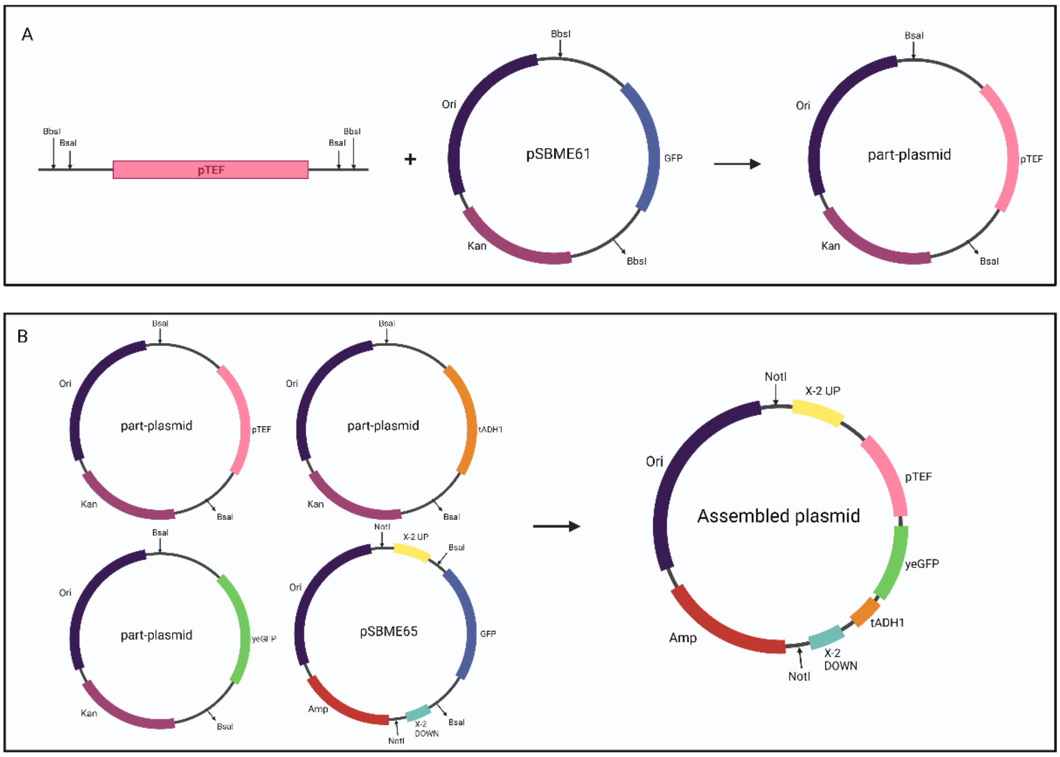


Figure 1. Assembly of Expression Plasmids. (A) The Biobricks encoding promoters are generated by PCR amplification, then assembled using BbsI into the integration vector, creating part plasmid type 2. (B) One part plasmid of each type is digested by BsaI producing unique upstream and downstream overhangs and then assembled into a cassette plasmid.

1.6 EasyClone-MarkerFree

Plasmids have advantages in terms of easy construction, but they do not provide stable gene expression as it does not engage with the yeast DNA integration machinery (Jensen et al., 2014). Therefore, an extra step is needed to target specific genomic loci, introduce DNA with the homology to the target site, and engage the cellular machinery to promote the integration.

The EasyClone method is a genetic engineering technique that allows for targeted integration of up to three DNA fragments into a host organism's genome in a single transformation step (Figure 2).In the EasyClone-MarkerFree method, clustered regularly interspaced short palindromic repeats (CRISPR)-Cas technology is used to introduce double-strand breaks at the defined integration sites. This causes the integrated DNA fragments to be efficiently integrated into the genome without the need for selection markers on the integrated DNA, bypassing the step of marker removal (Otto et al., 2021).

CRISPR-Cas was first described as an adaptive immune system in bacteria and archaea against invading nucleic acids derived from bacteriophages or plasmids. This prokaryotic system was adapted in molecular biology, and it has been used for genome editing, allowing the efficient integration of endogenous genes in various species and cell types.

CRISPR-Cas9 requires guide RNA (gRNA) and CRISPR-associated endonuclease (Cas9). Cas9, the Cas9 protein is then recruited to the chromosomal location by the gRNA, binds to the target DNA and creates a double-strand DNA break at the target site which must be followed by a protospacer adjacent motif (PAM)(Ma et al., 2014). Along with the Cas9-induced DNA break, a synthetic DNA template containing the desired genetic modification is introduced into the cell. A single step of homologous recombination is used to introduce the desired genetic modification directly into the genome without the need for any additional selection markers. This is achieved by designing the donor DNA molecule with homology arms that are specific to the target genome. The cell's DNA repair machinery uses the synthetic DNA template as a guide to repair the broken DNA strands, allowing the desired genetic modification to be incorporated into the repaired DNA strands by homologous recombination (Ahmed et al., 2021).

Therefore, the use of the EasyClone-MarkerFree method causes very efficient integration of expression cassettes, meaning no selection marker is introduced in the genome of the edited cells. The gRNA helper vectors, expressing gRNA molecules that recruit Cas9 to the chromosomal locations, can be easily removed from the strain by growth on a non-selective medium, after which the strain is ready for the next modification. This shortens the time and simplifies the design process (Otto et al., 2021).

This method was performed in this study to efficiently integrate the expression cassettes carrying the non-native promoters and controlling the GFP expression into the same predefined genomic region in S. cerevisiae, specifically in the X chromosome, location 2.

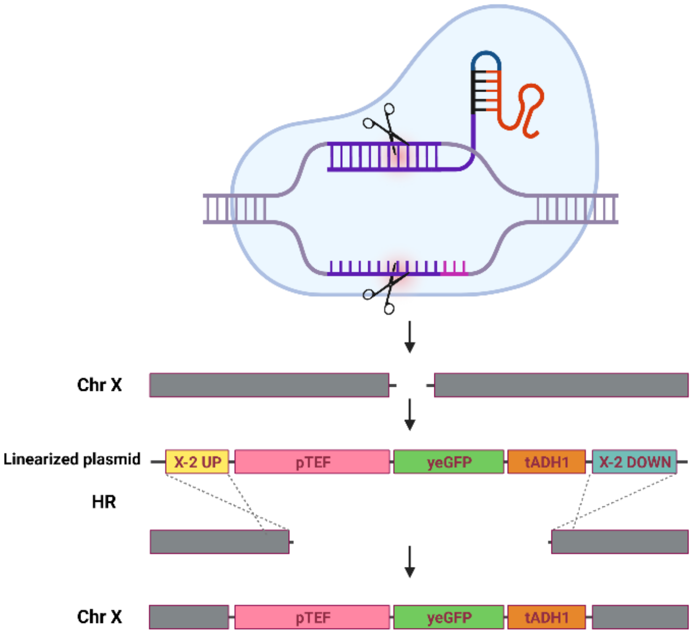


Figure 2. Integration into yeast genome. Linearized plasmid along with Cas9 and gRNA are transformed into *S. cerevisiae*. Cas9/gRNA induces double strand breaks (DSBs), which are repaired by homology-directed repair (HDR) pathway. The yeast DNA invades the homology arms of the linearized plasmid to serve as a primer for DNA repair synthesis. The linear fragments are stably integrated into the designated loci on the chromosomes.

1.7 Green fluorescent protein

Understanding the strength of mechanisms controlling transcription of a gene requires the characterization of regulatory elements. Green fluorescent protein (GFP) is a useful biological marker for measuring gene expression and quantifying it *in vivo* (Chudakov et al., 2010). The advantage of GFP is in its universality. Since GFP reaction does not require cofactors, the activity of the reporter does not depend on the metabolic status. When the reporter is cloned downstream of a gene or a regulatory region of interest on a plasmid that can enter the host cell, the reporter gene is co-expressed and co-regulated with the protein of interest (Zaslaver et al., 2006). In this study, a variant of GFP that expresses well in *S. cerevisiae*, yeast enhanced GFP (yeGFP), was used to assess the expression potential of the non-native promoters by measuring the fluorescence intensity upon growth of the yeast in a microtiter plate reader. The intensity of light emitted by a population of cells provides an indication of the quantity of the protein of interest, thus the strength of the promoter used.

1.8 Overview of workflow

In brief, this study focuses on the characterization of non-native promoters. Specifically, we examined eight TEF promoters obtained from various organisms: *Kluyveromyces marxianus*, *Cyberlindnera jardinii*, *Candida parapsilosis*, *Scheffersomyces stipites*, *Rhodotorula toruloides, Lipomyces starkeyl*, and *Yarrowla lipolytica*. Subsequently, we cloned these promoters with MoClo and integrated them into *S. cerevisiae* using EasyClone. The resulting strains were grown on glucose, sucrose, galactose, and ethanol to characterize their expression in *S. cerevisiae* across the different substrates. Our results show a diverse level of expression across the different non-native promoters, with high stability across the substrates. This indicates the promising future of non-native TEF promoters in industrial biotechnology.

2. AIMS OF THE PROJECT

The aims of this thesis are:

1) Isolate and clone TEF promoters from other yeast species: *Kluyveromyces marxianus*, *Cyberlindnera jardinii*, *Candida parapsilosis*, *Scheffersomyces stipites*, *Rhodotorula toruloides*, *Lipomyces starkeyi* and *Yarrowla lipolytica*.
2) Integrate GFP expression cassettes with non-native promoters into the *S. cerevisiae* genome.
3) Characterize the non-native promoters based on fluorescence in *S. cerevisiae*.
4) Test the strength of the promoters on different carbon sources.

3. EXPERIMENTAL PART

3.1 MATERIALS AND METHODS

3.1.1 Strains and molecular tools

The yeast strains used in this study were isogenic to CEN.PK113-7D, strains can be found in Supplementary Table 4. Escherichia coli DH5α was used for all bacterial transformations performed and storage of plasmids. Oligos, biobricks and plasmids are listed in Supplementary Table 1 to 3, respectively.

3.1.2 Media

Throughout this study, a variety of culture media were utilized to meet the specific requirements of each experimental stage.

Yeast Extract Peptone Dextrose (YPD) medium consisted of 10g/L yeast extract, 20g/L of peptone, and 100mL/L of sterile 20% (w/v) D-glucose solution.

Lysogeny Broth (LB) medium consisted of 20g/L of Difco LB broth powder (Difco) containing 10 g Tryptone, 5 g Yeast extract, and 5 g Sodium chloride.

Minimal medium consisted of a sugar source, 14.4 g/L of potassium dihydrogen orthophosphate, 7.5 g/L of armonium sulfate, 0.5 g/L of magnesium sulfate, 2 mL/L of trace elements solutions (containing 4.5 g/L $CaCl_2 \cdot 2H_2O$, 4.5 g/L $ZnSO_4 \cdot 7H_2O$, 3.0 g/L $FeSO_4 \cdot 7H_2O$ (Thermo Fisher), 1.0 g/L H_3Bo_3, 1.0 g/L $MnCl_2 \cdot 4H_2O$, 0.30 g/L $CuSO_4 \cdot 5H_2O$ and 0.10 g/L KI) and 1 mL of vitamin solution (containing 25 g/L myo-inositol, 1 g/L calcium pantothenate, 1 g/L nicotinic acid, 1 g/L pyridoxine HCl (Thermo Fisher), 1 g/L thiamine HCl, 0.2 g/L 4-aminobenzoic acid, 0.05 g/L biotin, and 40 ml/L of 0.1 M NaOH solution).

For the plate-reader experiment, 45 ml of the minimal media was supplemented with 5 ml of 2% of the respective carbon sources. The carbon sources used in the study were glucose, galactose, sucrose, and ethanol.

All materials were obtained from Sigma-Aldrich unless otherwise noted.

3.1.3 DNA amplification by polymerase chain reaction (PCR)

Candidate promoter sequences for integration were amplified to generate their respective Biobricks (Supplementary Table 2) using 1 µl of the template DNA, 25 µl of high-fidelity Phusion DNA Polymerase (Thermo Fisher Scientific), 2.5 µl of each correspondent primer from Supplementary Table 1, 19 µl of MilliQ water and using the following program:

Step	Temperature (°C)	Time (s)	
Initial denaturation	98	30	
Denaturation	98	10	30 cycles
Annealing	52	30	
Extension	72	60 per kb	
Final extension	72	300	
Storage	15	indefinite	

Following the PCR, the reactions were stained with 6X DNA Gel Loading Dye (Thermo Fisher Scientific). The samples over 500 bp were loaded on a 1% agarose gel and then DNA gel electrophoresis was run at a constant of 120 V. Fragment sizes were estimated using GeneRuler 1 kb Plus DNA Ladder (Thermo Fisher Scientific). Meanwhile, samples under 500 bp were loaded on 2% agarose gel, run at a constant of 120V, and estimated with a 50 bp GeneRuler Ladder. For Golden Gate assembly, fragments with the correct size were extracted using FavorPrep™ GEL/PCR Purification Kit (Favorgen) following the manufacturer's instructions. The concentrations of DNA were quantified using Marshall Scientific's NanoDrop Spectrophotometer ND-100.

3.1.4 Assembly with Golden Gate

Golden Gate cloning was used to assemble part plasmids and cassette plasmids (Supplementary Table 3). Assembly reactions were set up in the following manner: 20 fmol of each corresponding DNA parts/BioBricks, 1 µL of T4 DNA ligase buffer (Thermo Scientific), 0.5 µL T4 DNA ligase (Thermo Scientific), 0.5 µL FastDigest BpiI restriction enzyme or FastDigest Eco31I (Thermo Scientific) and up to 10 µL MilliQ water. The mixture was run on a ProFlex™ PCR System thermal cycler using the following program:

Steps	Temperature (°C)	Time (min)	
Initial Digestion	37	5	
Digestion	37	1	35 cycles
Ligation	16	2	
Final Digestion	37	4	
Heat inactivation	80	10	
Storage	15	indefinite	

All assemblies were used for bacterial transformation. To validate the efficiency of the cloning process, three colonies were chosen for colony PCR to check the sizes of the plasmid, using 1 µl of primers OL77 and OL78, 2 µl of FIREPol Master Mix Ready to Load, 6 µl of MilliQ water, and using the following program:

Steps	Temperature (°C)	Time (min)	
Initial denaturation	95	300	
Denaturation	95	20	30 cycles
Annealing	52	40	
Extension	72	60 per kb	
Final extension	72	300	
Storage	15	indefinite	

Colonies containing the correct constructs were grown overnight at 37°C using 3 mL of LB medium with appropriate selection. 500 µL of the cultures mixed with 500 µL 50% glycerol was stored at -80°C, and the remaining volume was used for plasmid purification using FavorPrep™ Plasmid DNA Extraction Mini Kit (Favorgen) following the manufacturer's instructions, only skipping the WP Buffer step. All constructs were validated by DNA sequencing.

3.1.5 Bacterial transformation

Escherichia coli DH5α cells were taken out of -80°C freezer and left on ice to thaw. After thawing, 3 µL of the Golden Gate reactions and 30 µL chemically competent E. coli DH5 were mixed and incubated on ice and then heat-shocked for 60 seconds at 42 °C in a heat block and put back on ice for 2-4 minutes. The mixture was plated on LB plates containing the corresponding antibiotic, either 100 µg/mL kanamycin or 25 µg/mL ampicillin, and incubated overnight at 37 °C.

3.1.6 Yeast transformation

To facilitate the integration of DNA into the yeast genome, the plasmids need to be linearized, creating ends that are similar to the yeast genome. This was done by mixing 5 µL of the cassette plasmids with 2 µL FastDigest buffer, 1 µL FastDigest NotI, and up to 20 µL MilliQ water, incubating at 37°C for one hour, and then heat-inactivating to 80°C for 5 minutes.

Competent yeast cells capable of taking up foreign DNA are prepared by growing the yeast strain overnight at 30 °C in 2 mL YPD medium with appropriate antibiotic selection. The titer of the yeast culture is then determined, and 2.5 × 10^8 cells are inoculated into 50 mL fresh YPD medium to give a concentration of 5 × 10^6 cells per mL culture (OD_{600}= 0.5 in the final 50mL YPD medium). The culture is then grown for 4 hours until it reaches a concentration of approximately 2 × 10^7 cells per mL culture (OD_{600}= 2). The cells are centrifuged, washed, and then centrifuged again. They are subsequently resuspended in 1 mL of sterile MilliQ water, pelleted by centrifugation, resuspended again in 1 mL of sterile MilliQ water, pipetted into 100 µL aliquots, and pelleted once more.

To transform the competent yeast cells, a transformation mixture is prepared containing 240 µL 50% (w/v) PEG mw, 36 µL 1 M lithium acetate, 20 µL ss-carrier DNA, 0.1 µg gRNA helper, 0.1-1 µg of the linearized DNA, and the mixture is topped up to 360 µL using sterile MilliQ water. The transformation mixture is added to the competent yeast cells, and the mixture is incubated at 42°C for 40 minutes. This step allows the foreign DNA to enter the yeast cells.

After the incubation, the cells are centrifuged, resuspended in YPD medium without selection, and incubated for 2 hours at 30°C while shaking for recovery and to allow for expression of the resistance gene. The cells are then centrifuged again, resuspended in sterile MilliQ water, and plated on YPD agar plates with nourseothricin and G418 selection.

Three colonies from each transformant had their genomic DNA (gDNA) extracted. This gDNA was subsequently used for colony PCR to check the homology arms.

A single correct yeast colony is inoculated in YPD medium without any selection, incubated at 30° C at 160 rpm for 16 hours. After overnight incubation, 500 µL of the cultures mixed with 500 µL 50% glycerol was stored at -80°C, and 3 µl of cells is pipette on a YPD plate and incubated at 30° C for two days. Individual colonies are restreaked on a YPD plate without selection and on YPD plates with nourseothricin antibiotic or G418 antibiotic. The cells are grown on YPD, YPDN, and YPDG to identify those that carry Cas9 and gRNA. Single colonies that only grew in YPD were chosen, as those grown in YPDG still have Cas9 protein, meaning incomplete loss of the plasmid, while those grown in YPDN still contain gRNA.

All strains used and constructed in this study can be found in Supplementary table 4.

3.1.7 Growth and promoter strength characterization

Biological replicates were inoculated in a 24-deep-well plate with 500 µl of minimal media with 2% glucose and 2% sucrose for precultures. The strains were grown at 30° C and 300 rpm for two days in a shaking incubator. While a 96-deep-well plate was used to inoculate replicas in 2% galactose and 2% (v/v) ethanol, which was shaken at 410 rpm.

The OD_{600} of the strain replicates in each carbon source was measured by using 10x dilution and 200 volume μL in a microtiter plate reader. Subsequently, samples were first diluted to 0.4 OD_{600}, as an intermediate dilution due to the imprecision of the machine with samples under 0.1 OD_{600}. After this first dilution, yeasts were diluted 20x to get OD_{600}nm of 0.02. In order to prevent evaporation and ensure reliable results, the edges of the microtiter plate were left free from samples and to help maintain a consistent temperature throughout the experiment, minimal media was added to them. The plates were shaken at the fast speed setting at 30° C while every 15 minutes OD_{600} and GFP fluorescence was measured, cells in galactose and ethanol were incubated for 48 hours, while cells in glucose and sucrose were incubated for 36 hours. The goal was to compare the OD_{600} and the GFP, during the exponential phase and middle of the stationary phase, to make sure they accumulate enough GFP.

Appropriate positive and negative control were included to validate the accuracy of flow cytometry measurements and to account for background fluorescence. The negative control used was a yeast strain, SBME1, without GFP expression, while the positive control was a strain, SBME8, expressing GFP under its own TEF1 promoter. Another reference strain, SBME40, expressing GFP under a supposed medium-strength promoter, ADH1 promoter, was also included for comparison.

3.2 RESULTS

3.2.1 Identification TEF promoters

An initial search for TEF promoters was performed using the TEF protein sequence of *S. cerevisiae* to find homologous genes through Basic Local Alignment Search Tool (BLAST) in other fungi/yeasts (*Kluyveromyces marxianus*, *Cyberlindnera jardinii*, *Candida parapsilosis, Scheffersomyces stipites*, *Rhodotorula toruloides*, *Lipomyces starkeyi* and *Yarrowla lipolytica*), The selection of organisms was based on their availability in-house, and also by their close relationship to *S. cerevisiae*, determined through phylogenetic analysis conducted by my supervisor. This approach allowed us to narrow down the potential candidates for TEF promoters. The TEF gene is regulated by a promoter sequence, which is located upstream of the coding region and contains specific regulatory elements that control the initiation of transcription. Through this process, we identified strains with promising TEF promoters that were readily accessible for experimental validation. Subsequently, 1000 bp upstream sequence from the translation start codon, AUG, was selected for further screening to identify the promoter sequence (Le et al., 2019). Organisms with the presence of another gene within those 1000 bp had their promoter sequence reduced.

3.2.2 Assembly part plasmids and cassette plasmids

Seven strains were selected to undergo genomic extraction and PCR using the corresponding primers containing BbsI and BsaI recognition site and the specific overhang correspondent for a Type 2 part plasmid to obtain their promoter sequence. Two PCR products were made for *Yarrowia lipolytica* and *Lipomyces starkeyi,* three for ADH1, and four for CYC1 promoter sequence to remove internal restriction enzymes. The TEF of the yeast *Candida parapsilosis* is coded by two genes, called TEF1 and TEF2, therefore we made a different experiment for each of them. CparTEF1 was also made using two biobricks to delete internal restriction sites. Results for the PCR amplification of the biobricks can be found in Supplementary figure 1 to 6. Biobricks for *Rhodotorula toruloides*, *Lipomyces starkeyi* and CYC1 did not work out.

Each promoter sequence is assembled into a universal entry vector, pSBME61, replacing the green fluorescent protein gene for expression in bacteria, and therefore allowing selection when grown in bacteria under kanamycin selection. Thus, the colonies that did not show fluorescence under UV light were picked (Supplementary Figure 7 to 9). The successful plasmids resulted in part plasmids type 2. There are four types of part plasmid used to construct our cassette plasmids by digesting each of them with BsaI; a backbone plasmid pSBME65 containing ampicillin resistance, as well as upstream and downstream homology arms for yeast, type 3 part plasmid pSBME95 with yeast enhanced green fluorescent protein gene to evaluate the strength of the promoter sequence when integrated in the genome, type 4 part plasmid pSBME98 with ADH1 terminator sequence, and of course the type 2 part plasmids containing the non-native promoters that need to be characterized in yeast. In total, six plasmids were assembled each with its respective non-native promoter and their full expression cassette driving yeGFP expression with the ADH1 terminator The assemblances were tested, and all colonies showed correct construction (Supplementary Figure 11 and 12).

3.2.3 Integration into the yeast genome

NotI restriction sites present in our cassette plasmids, were used to linearize the DNA for subsequent integration into the yeast genome, cutting the origin of replication and the ampicillin resistance gene out. The success of the linearization was visualized through DNA electrophoresis, which showed two distinct bands (Supplementary Figure 13). One band corresponded to the expression cassette to be integrated with the homology arms for repair, while the other represented the backbone. The linearized fragment, along with the gRNA helper expressing Cas9, were then transformed into *S. cerevisiae* cells. Cells that have only taken the gRNA vector up will cut the target site in the yeast genome, but since no repair fragment with homology arms is present, they will eventually die. On the other hand, cells that only took up the linearized DNA repair fragment, but not the gRNA vector containing the nourseothricin resistance cassette will be killed by the nourseothricin when the cells are plated. However, in cells that have taken both, the gRNA guides the Cas9 endonuclease to the chromosome and introduces a double-stranded break. The native homologous recombination machinery of S. cerevisiae recognizes the homology arms in our linearized DNA repair fragment and repairs the DSB. The linear fragments are stably integrated into chromosome X (Supplementary Figure 14 to, 16).

3.2.4 Green Fluorescence Protein expression

Three biological replicates were chosen to perform the experiment as they increase the robustness of the data analyzed and address if the effect of the different promoters is sustained under different biological variables. They were grown in 24-deep-well plates in glucose, sucrose, galactose, and ethanol mediums. Subsequently, each strain was diluted 20 times, except ethanol which underwent a 2-time dilution, and screened in a plate reader. The 2-time dilution for ethanol was necessary as yeast may face difficulties in metabolizing ethanol, leading to a slower growth rate.

The plate reader experiment was performed to measure the absorbance and GFP expression of the cell cultures every 15 minutes for two days. Optical density serves as an indirect measure of cell density in culture, as it quantifies the amount of light absorbed by a sample. Thus, after plotting the progression of the OD_{600} it is possible to assess the behavior of proliferating cells under different conditions.

One important consideration taken in the analysis of the results was that biological replicas grow at different speeds, due to slight differences in the lag phase of the cells. Especially on galactose, this made such a big difference that timepoints between strains could not be compared together. Therefore, the figures show one representative from each strain, chosen to represent the average of the biological replicates as good as possible.

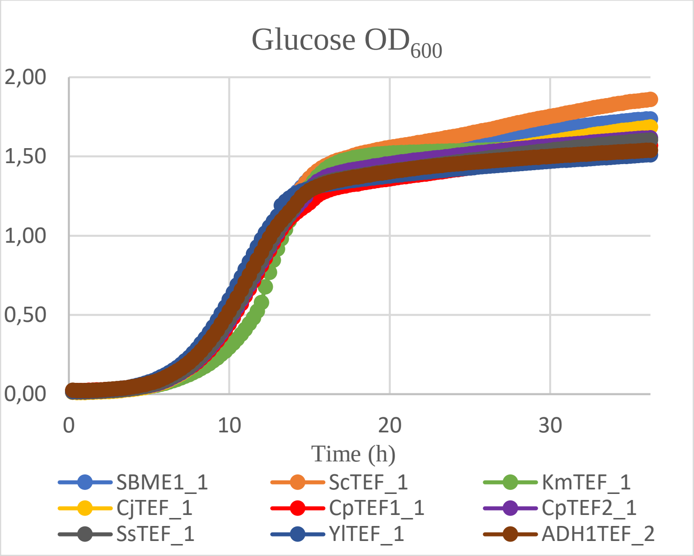


Figure 3. OD_{600} of representative strains under glucose conditions

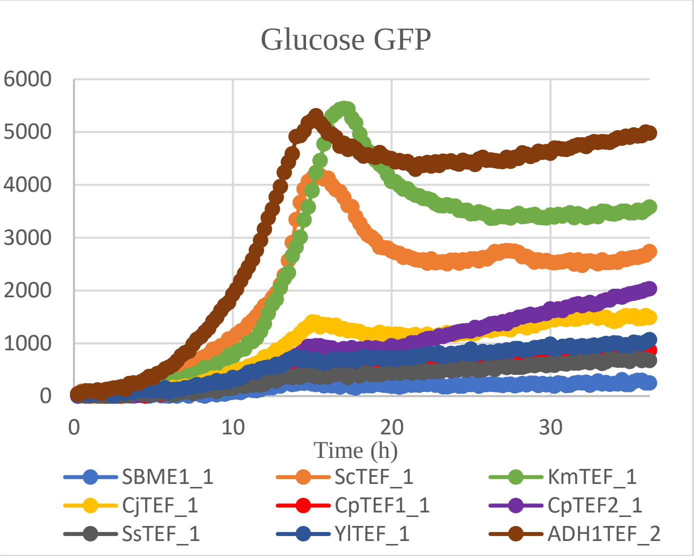


Figure 4. GFP expression of representative strains under glucose conditions

Most strains maintained a relatively stable GFP expression level, either with no significant change or a peak after 15 hours under glucose (Figure 3 and Figure 4), after reaching the end of the exponential phase, except for *Kluyveromyces marxianus*, which took slightly longer to reach its peak but ended up showing results 1.2 times higher than the strain under its own native TEF promoter, 6000 compared to 5000 from SBME8. *Candida parapsilosis* TEF2 displayed a continuous growth beyond 35 hours, surpassing the first peak achieved after 15 hours and reaching levels of 2000. Similarly, *Cyberlindnera jardinii* reached 1400 during its peak at 15 hours, then experience a small drop but after exhibite another increase, it maintained values around 1500 instead of displaying the exponential increase seen in *Candida parapsilosis TEF2*. *Candida parapsilosis* TEF1, *Scheffersomyces stipitis*, *Yarrowia lipolytica* all expressed growth pattern similar to *Cyberlindnera jardinii* but with a smaller drop and lower expression levels, 1000, 700, and 1000 respectively.

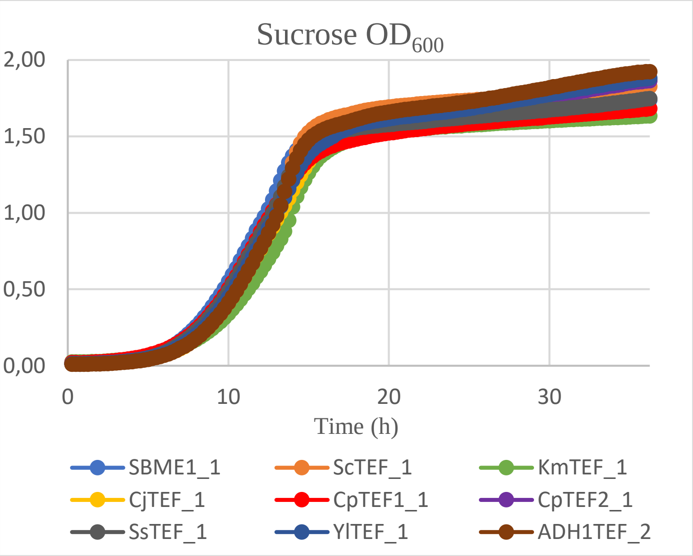


Figure 5. OD_{600} of representative strains under sucrose conditions

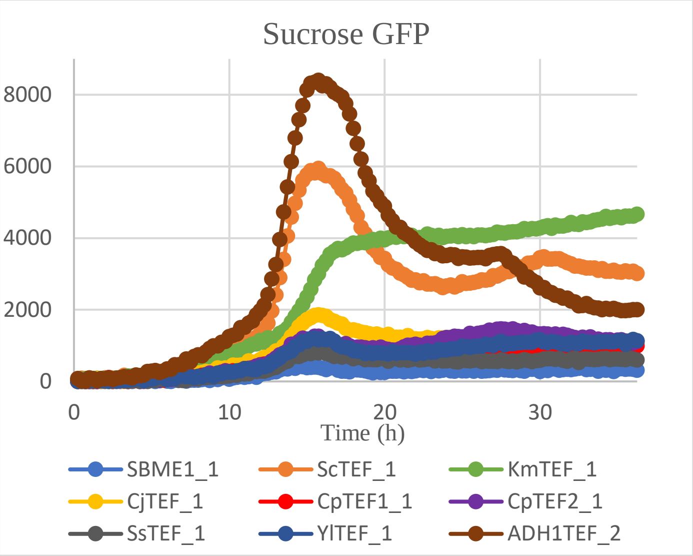


Figure 6. GFP expression of representative strains under sucrose conditions

Samples in sucrose consistently peaked at 15 hours (Figure 5 and Figure 6), after the end of the exponential phase, except for *Kluyveromyces marxianus*, which exhibited continuous growth in GFP expression instead of undergoing a drop in expression. The positive control showed its best results under sucrose, expressing values of 6000 and a much clearer drop before stabilizing the expression of the protein. Similarly, the supposed medium-strength promoter, ADH1 promoter, exhibited its highest GFP expression peak, approximately 8000, under sucrose conditions. *Candida parapsilosis* TEF2 seemed to reach its peak at 1500, experience a drop to 1000, then increased to 1500 before decreasing again. Similarly, *Cyberlindnera jardinii* promoter reached its peak also at 1800 but experienced a steady drop afterward. *Scheffersomyces* stipites showed maximal GFP levels of 800, and *Yarrowia lipolytica* achieved its peak at approximately 1300, followed by a decrease and subsequent constant growth, reaching 1400 after 35 hours. Among them, *Scheffersomyces stipites* and *Candida parapsilosis* TEF1 showed the lowest results with their maximal GFP values being 800 and 1000, respectively.

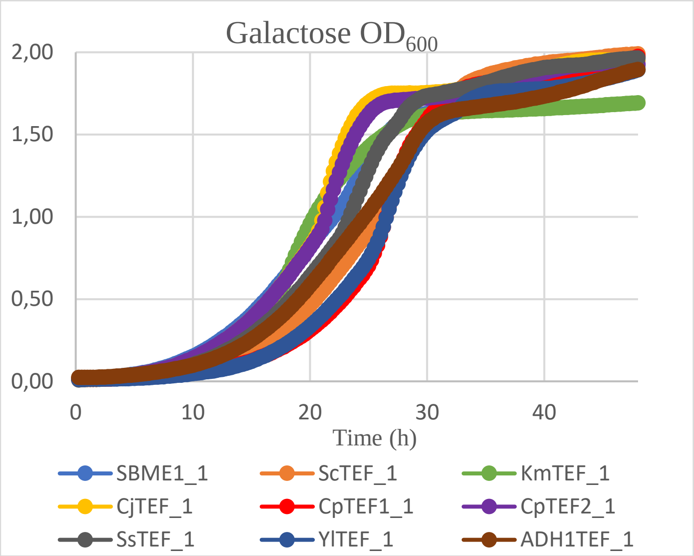


Figure 7. OD$_{600}$ of representative strains under galactose conditions

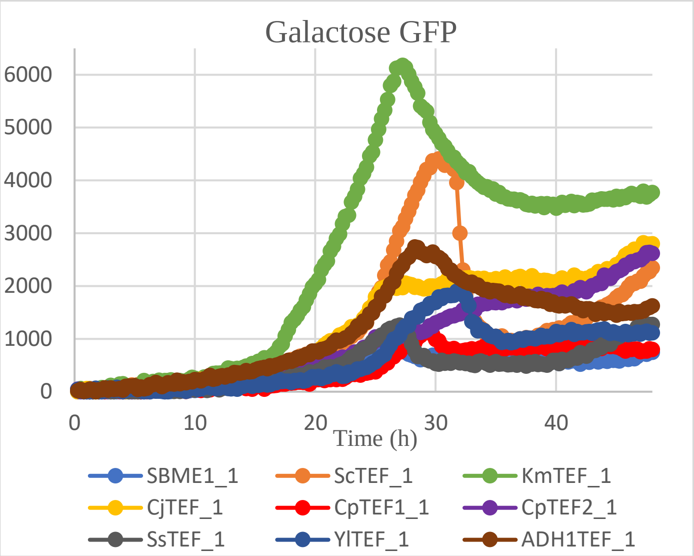


Figure 8. GFP expression of representative strains under galactose conditions

Strains under galactose reached their peak between 25 and 30 hours (Figure 7 and Figure 8), closely aligned with the end of the exponential phase. Native promoter exhibited medium-strength activity in comparison to the *Kluyveromyces marxianus* exhibited remarkably high GFP expression at 8000. The *Cyberlindnera jadinii* promoter reached its highest value of 2000 in galactose medium but experienced a modest decline compared to the constant GFP expression observed in other media. Nonetheless, it reached a similar expression level as the ADH1 promoter. *Candida parapsilosis* TEF1 promoter reached its peak GFP expression at around 1400. In contrast, galactose was not the optimal medium for the *Candida parapsilosis* TEF2 promoter. It displayed a peak after the exponential phase, followed by declining values. However, it subsequently demonstrated continuous growth, surpassing the previous peak, and reaching values similar to the peak of the supposed medium-strength promoter. *Scheffersomyces stipitis* showed maximal GFP levels of 1200, and *Yarrowia lipolytica* at approximately 1800, following a growth pattern similar to sucrose.

Ethanol showed almost no growth in any of the strains (Supplementary Figure 18), except for *Kluyveromyces marxianus* promoter, which seems to experience a slight growth after 35 hours. Therefore, the results are not presented.

Finally, we determined the fluorescence levels at an OD$_{600}$ of 1 and at peak fluorescent intensity for each strain, so that the variation in lag phase wouldn't be an issue. We normalized the values to the OD$_{600}$ and the average background fluorescence for each carbon source. Subsequently, for each strain on each carbon source the normalized average and standard deviation of the fold change over background were calculated (Figure 9).

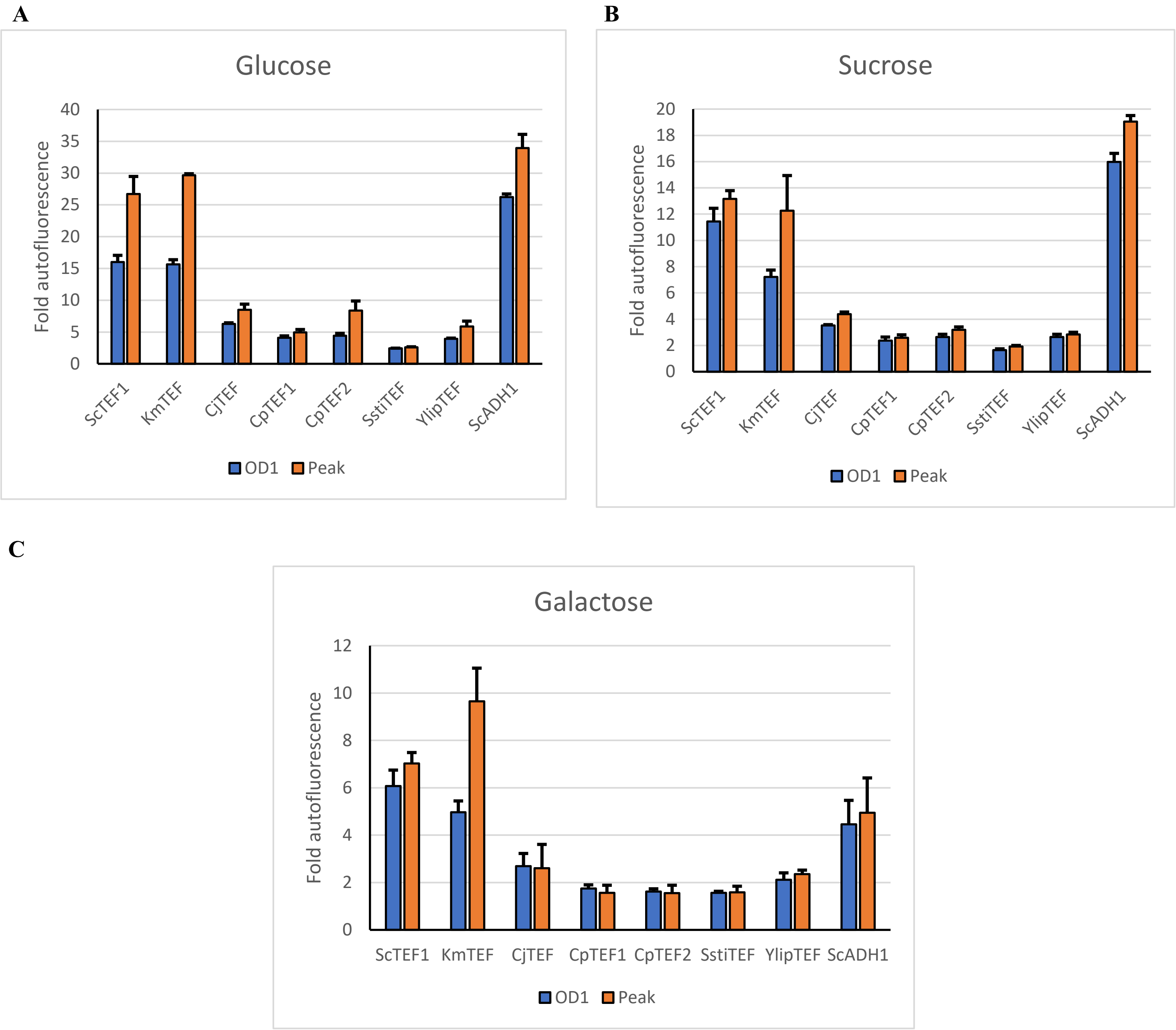


Figure 9. Fold increase in expression over the background for each strain and carbon substrate. The figures show the fluorescence levels at an OD_{600} of 1 and at peak fluorescent intensity under (A) glucose (B) sucrose (C) galactose.

In the figures, we can observe the diversity of strength of the different promoters and at the same time their trend in the different carbon sources. For instance, it is clear the strong performance of *Kluyveromyces marxianus* across the different media in comparison to the non-native promoters and the native promoter expressing GFP, and how *Cyberlindnera jardinii* can be described as a medium-stregth and *Candida parapsilosis* TEF2 can also be seen as a medium-stregth mostly in glucose and sucrose. It has to be noted that the background fluorescence levels were higher for sucrose and galactose, most likely due to morphological changes that happen due to the metabolism of the carbon sources. For example, in the case of galactose, flocculation, and that is the reason why the fold increases look smaller, even though the relative fluorescence units are similar between experiments.

3.3 DISCUSSION

We have developed a library of weak, medium-strength, and strong non-native promoters that could serve as a potential toolkit for the industry. During the obtaining of the promoter sequences, the BioBricks for the *Rhodotorula toruloides* and *Lipomyces starkeyi* TEF promoters may not have formed, as the available strains for these organisms were different from the strains for which we had the DNA sequence of the TEF promoter. Since there could be genetic variation between the strains, particularly in the primer binding sites or in the target sequence, this could have prevented the successful amplification of the product.

The alignment of the promoter sequences of the different organism has shown small overlaps of less than 11 base pairs that do not count as homology regions but can be a hint to possible sequences that have been evolutionary conserved and are necessary for the promoter region of the TEF gene. (Supplementary figure 18). However, even though the promoter sequence with more overlaps with the host genome is *Kluyveromyces marxianus*, but as strains like *Candida parapsilosis* TEF1, which happens to be a weak promoter, also contains one of the sequences with more small overlaps, we cannot say there is a possible explanation for the great performance of K related to the base pairs that have in common.

Aside from developing the promoters, we also examined the impact of sugars frequently encountered and starting to be desired in industrial processes on the promoters, as yeast pathways are subject to change under different cultivation conditions(Peng et al., 2015). During the experiment, the negative control still exhibits a low level of autofluorescence, which arises from the presence of endogenous fluorophores within the cells. Still, the strains demonstrated a preference for simple sugars, glucose and sucrose, as they can enter glycolysis directly. Meanwhile, cells have to enzymatically modify other carbon sources and adapt to them. For example, carbon sources like galactose strains exhibited a slower growth rate. Additionally, strains undergo flocculation in this type of carbon source(Nayyar et al., 2017), meaning yeast cells aggregate, and form large clusters that possess a higher capacity to absorb and reflect green fluorescence. Thereby, due to the high amount of autofluorescence in galactose conditions it is more difficult to see the difference between the negative control and the weaker promoters. Another important observation was that the strains do not seem to grow in ethanol conditions, but it may be possible that this outcome is because of not optimal growth conditions in the plate reader, or due to the use of 96% ethanol that may still contain components that may interfere with the growth. Moreover, the promoter ADH1 that we believed was medium-strength as it was described as one in previous studies, and therefore was chosen as an intermediate comparison showed to be a strong promoter in carbon sources like sucrose and glucose, but indeed medium-strength in carbon sources like galactose (Yan et al., 2022)(Peng et al., 2015).

In order to improve the reliability of the experiment, additional techniques, such as fluorescence-activated cell sorting (FACS), which analyzes individual cells based on their fluorescent properties, could be used to corroborate the results obtained from the plate reader experiment(Cormack et al., 1996). Furthermore, the use of 100% ethanol to check if the yeast would be more likely to grow in improved conditions should be checked. Performing the experiment on different days would be beneficial for further research as it would help to account for any variability that may arise due to external factors, such as changes in the laboratory conditions.

Future research could be continued by minimizing the promoters and doing computational analysis to compare promoter elements and identify elements within the promoters, such as regulatory sequences or binding sites, which play a role in controlling gene expression. These approaches could potentially provide valuable insights into the function of common and different promoters elements between organisms, helping to focus on the essential elements within the sequences to design more efficient and targeted promoters.

SUMMARY

We were able to obtain the TEF sequences and cloned them into expression cassettes of the organisms: *Kluyveromyces marxianus*, *Cyberlindnera jardinii*, *Candida parapsilosis*, *Scheffersomyces stipites*, and *Yarrowla lipolytica*. *Rhodotorula toruloides* and *Lipomyces starkeyl* were not possible to obtain after many tries.

All expression cassettes obtained were successfully integrated into the yeast genome.

Kluyveromyces marxianus appears to be a strong promoter, with higher values in all the substrates, even higher than our positive controls, except in sucrose where both the native promoter that has GFP and the supposed medium-strength promoter get higher values, ADH1 promoter reaching almost double the highest value of our promoter. Both *Cyberlindnera jardinii* and *Candida parapsilosis* TEF2 showed to be medium-strength promoters. Lastly, *Candida parapsilosis TEF1, Scheffersomyces stipites*, and *Yarrowla lipolytica* seemed to be weak promoters but nonetheless still gave values higher than our negative control demonstrating their efficiency in the yeast genome.

REFERENCES

Ahmed, M. Z., Badani, P., Reddy, R., & Mishra, G. (2021). Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas Advancement in Molecular Diagnostics and Signal Readout Approaches. In *Journal of Molecular Diagnostics* (Vol. 23, Issue 11, pp. 1433–1442). Elsevier B.V. https://doi.org/10.1016/j.jmoldx.2021.07.025

Akbarian, A., Andooz, A., Kowsari, E., Ramakrishna, S., Asgari, S., & Cheshmeh, Z. A. (2022). Challenges and opportunities of lignocellulosic biomass gasification in the path of circular bioeconomy. In *Bioresource Technology* (Vol. 362). Elsevier Ltd. https://doi.org/10.1016/j.biortech.2022.127774

Borodina, I., & Nielsen, J. (2014). Advances in metabolic engineering of yeast Saccharomyces cerevisiae for production of chemicals. In *Biotechnology Journal* (Vol. 9, Issue 5, pp. 609–620). Wiley-VCH Verlag. https://doi.org/10.1002/biot.201300445

Cherry, J. M., Hong, E. L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E. T., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S. R., Fisk, D. G., Hirschman, J. E., Hitz, B. C., Karra, K., Krieger, C. J., Miyasato, S. R., Nash, R. S., Park, J., Skrzypek, M. S., … Wong, E. D. (2012). Saccharomyces Genome Database: The genomics resource of budding yeast. *Nucleic Acids Research*, *40*(D1). https://doi.org/10.1093/nar/gkr1029

Cho, J. S., Kim, G. B., Eun, H., Moon, C. W., & Lee, S. Y. (2022). Designing Microbial Cell Factories for the Production of Chemicals. In *JACS Au* (Vol. 2, Issue 8, pp. 1781–1799). https://doi.org/10.1021/jacsau.2c00344

Choi, K. R., Jang, W. D., Yang, D., Cho, J. S., Park, D., & Lee, S. Y. (2019). Systems Metabolic Engineering Strategies: Integrating Systems and Synthetic Biology with Metabolic Engineering. *Trends in Biotechnology*, *37*(8), 817–837. https://doi.org/10.1016/j.tibtech.2019.01.003

Chudakov, D. M., Matz, M. V., Lukyanov, S., & Lukyanov, K. A. (2010). Fluorescent Proteins and Their Applications in Imaging Living Cells and Tissues. *Physiological Reviews*, *90*(3), 1103–1163. https://doi.org/10.1152/physrev.00038.2009

Cormack, B. P., Valdivia, R. H., & Falkow, S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*, *173*(1), 33–38. https://doi.org/10.1016/0378-1119(95)00685-0

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. https://doi.org/10.1111/1567-1364.12118

Jin, L., Nawab, S., Xia, M., Ma, X., & Huo, Y. (2019). Context-dependency of synthetic minimal promoters in driving gene expression: a case study. *Microbial Biotechnology*, *12*(6), 1476–1486. https://doi.org/10.1111/1751-7915.13489

Köbbing, S., Blank, L. M., & Wierckx, N. (2020). Characterization of Context-Dependent Effects on Synthetic Promoters. *Frontiers in Bioengineering and Biotechnology*, *8*. https://doi.org/10.3389/fbioe.2020.00551

Le, N. Q. K., Yapp, E. K. Y., Nagasundaram, N., & Yeh, H.-Y. (2019). Classifying Promoters by Interpreting the Hidden Information of DNA Sequences via Deep Learning and Combination of Continuous FastText N-Grams. *Frontiers in Bioengineering and Biotechnology*, *7*. https://doi.org/10.3389/fbioe.2019.00305

Lee, M. E., DeLoache, W. C., Cervantes, B., & Dueber, J. E. (2015). A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synthetic Biology*, *4*(9), 975–986. https://doi.org/10.1021/sb500366v

Li, M., & Borodina, I. (2015). Application of synthetic biology for production of chemicals in yeast Saccharomyces cerevisiae. In *FEMS Yeast Research* (Vol. 15, Issue 1). Oxford University Press. https://doi.org/10.1111/1567-1364.12213

Ma, Y., Zhang, L., & Huang, X. (2014). Genome modification by CRISPR/Cas9. In *FEBS Journal* (Vol. 281, Issue 23, pp. 5186–5193). Blackwell Publishing Ltd. https://doi.org/10.1111/febs.13110

Matanović, A., Arambašić, K., Žunar, B., Štafa, A., Miklenić, M. S., Šantek, B., & Svetec, I. K. (2022). Toolbox for Genetic Transformation of Non-Conventional Saccharomycotina Yeasts: High Efficiency Transformation of Yeasts Belonging to the Schwanniomyces Genus. *Journal of Fungi*, *8*(5). https://doi.org/10.3390/jof8050531

Nayyar, A., Walker, G., Wardrop, F., & Adya, A. K. (2017). Flocculation in industrial strains of *Saccharomyces cerevisiae* : role of cell wall polysaccharides and lectin-like receptors. *Journal of the Institute of Brewing*, *123*(2), 211–218. https://doi.org/10.1002/jib.421

Nevoigt, E. (2008). Progress in Metabolic Engineering of Saccharomyces cerevisiae . *Microbiology and Molecular Biology Reviews*, *72*(3), 379–412. https://doi.org/10.1128/mmbr.00025-07

Nielsen, J. (2013). Production of biopharmaceutical proteins by yeast. *Bioengineered*, *4*(4), 207–211. https://doi.org/10.4161/bioe.22856

Nielsen, J. (2019). Yeast Systems Biology: Model Organism and Cell Factory. In *Biotechnology Journal* (Vol. 14, Issue 9). Wiley-VCH Verlag. https://doi.org/10.1002/biot.201800421

Nielsen, J., & Keasling, J. D. (2016). Engineering Cellular Metabolism. *Cell*, *164*(6), 1185–1197. https://doi.org/10.1016/j.cell.2016.02.004

Otto, M., Skrekas, C., Gossing, M., Gustafsson, J., Siewers, V., & David, F. (2021). Expansion of the Yeast Modular Cloning Toolkit for CRISPR-Based Applications, Genomic Integrations and Combinatorial Libraries. *ACS Synthetic Biology*, *10*(12), 3461–3474. https://doi.org/10.1021/acssynbio.1c00408

Parapouli, M., Vasileiadis, A., Afendra, A. S., & Hatziloukas, E. (2020). Saccharomyces cerevisiae and its industrial applications. In *AIMS Microbiology* (Vol. 6, Issue 1, pp. 1–31). AIMS Press. https://doi.org/10.3934/microbiol.2020001

Patel, J. R., Oh, J., Wang, S., Crawford, J. M., & Isaacs, F. J. (2022). Cross-kingdom expression of synthetic genetic elements promotes discovery of metabolites in the human microbiome. *Cell*, *185*(9), 1487-1505.e14. https://doi.org/10.1016/j.cell.2022.03.008

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*(1). https://doi.org/10.1186/s12934-015-0278-5

Redden, H., & Alper, H. S. (2015). The development and characterization of synthetic minimal yeast promoters. *Nature Communications*, *6*. https://doi.org/10.1038/ncomms8810

Redden, H., Morse, N., & Alper, H. S. (2015). The synthetic biology toolbox for tuning gene expression in yeast. In *FEMS Yeast Research* (Vol. 15, Issue 1). Oxford University Press. https://doi.org/10.1111/1567-1364.12188

Rugbjerg, P., Myling-Petersen, N., Porse, A., Sarup-Lytzen, K., & Sommer, M. O. A. (2018). Diverse genetic error modes constrain large-scale bio-based production. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-03232-w

Schirmaier, F., & Philippsen, P. (1984). Identification of two genes coding for the translation elongation factor EF-la of S. cerevisiae. In *The EMBO Journal* (Vol. 3, Issue 13).

Sewalt, V., Shanahan, D., Gregg, L., La Marta, J., & Carillo, R. (2016). The Generally Recognized as Safe (GRAS) process for industrial microbial enzymes. In *Industrial Biotechnology* (Vol. 12, Issue 5, pp. 295–302). Mary Ann Liebert Inc. https://doi.org/10.1089/ind.2016.0011

Sheldon, R. A. (2016). Engineering a more sustainable world through catalysis and green chemistry. *Journal of the Royal Society Interface*, *13*(116). https://doi.org/10.1098/rsif.2016.0087

Steiner, S., & Philippsen, P. (1994). *Sequence and promoter analysis of the highly expressed TEF gene of the filamentous fungus Ashby gossypii* (Vol. 242).

Stewart, G. G. (2014). Saccharomyces: Saccharomyces cerevisiae. In *Encyclopedia of Food Microbiology: Second Edition* (pp. 309–315). Elsevier Inc. https://doi.org/10.1016/B978-0-12-384730-0.00292-5

Weber, E., Engler, C., Gruetzner, R., Werner, S., & Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE*, *6*(2). https://doi.org/10.1371/journal.pone.0016765

Weeks, A. M., & Chang, M. C. Y. (2011). Constructing de Novo Biosynthetic Pathways for Chemical Synthesis inside Living Cells. *Biochemistry*, *50*(24), 5404–5418. https://doi.org/10.1021/bi200416g

Yan, C., Yu, W., Zhai, X., Yao, L., Guo, X., Gao, J., & Zhou, Y. J. (2022). Characterizing and engineering promoters for metabolic engineering of Ogataea polymorpha. *Synthetic and Systems Biotechnology*, *7*(1), 498–505. https://doi.org/10.1016/j.synbio.2021.12.005

Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M. G., & Alon, U. (2006). A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. *Nature Methods*, *3*(8), 623–628. https://doi.org/10.1038/nmeth895

APPENDIX

Supplementary Table 1. List of Primer used in this study.

ID	Name	Sequence 5' to 3'
OL351	2_Kmar_TEF1_fwd	GCATGAAGACATTCGGTCTCAAACGACACCGATGAAGCAAAGAAG
OL352	2_Kmar_TEF1_rev	ATCGAAGACATGGTCTCACATATCTTCTTTAATGTTACTTCTCTTGGAGTTAG
OL353	2_Cjar_pTEF1_fwd	GCATGAAGACATTCGGTCTCAAACGCCGTTTGTATATGTCTCCAAGT
OL354	2_Cjar_pTEF1_rev	ATCGAAGACATGGTCTCACATATCTTTGTGATTTATTACTATCAAATTTATCTAACTTCT
OL355	2_Cpar_pTEF1_part1_fwd	GCATGAAGACATTCGGTCTCAAACGTGTTTTTTTAACATCTATACATCAATGAG
OL356	2_Cpar_pTEF1_part1_rev	ATCGAAGACATGTCAATTCCTCTCAATGACACAAAA
OL357	2_Cpar_pTEF1_part2_fwd	GCATGAAGACATTGACAGACCAAATCAGAAACATC
OL358	2_Cpar_pTEF1_part2_rev	ATCGAAGACATGGTCTCACATATCTTGATTAGATTGATTTAAACTTTTAGTTTGTAAAGA
OL359	2_Cpar_pTEF2_fwd	GCATGAAGACATTCGGTCTCAAACGCCCTATTACTACAGACCACGT
OL360	2_Cpar_pTEF2_rev	ATCGAAGACATGGTCTCACATATCTTTTTTGTTGTTGCTATATAGATTGAATACT
OL361	2_Lsta_pTEF1_part1_fwd	GCATGAAGACATTCGGTCTCAAACGACCATTAAGATTCACTGTCCTTG
OL362	2_Lsta_pTEF1_part1_rev	ATCAGAAGACATACTTCCCGGTGCCTCAG
OL363	2_Lsta_pTEF1_part2_fwd	GCATGAAGACATAAGTCAAAGACAAAAAAATGCGG
OL364	2_Lsta_pTEF1_part2_rev	ATCGAAGACATGGTCTCACATATCTTTTTGAAAGATTAAAGAGGATGTATAAAAGA
OL367	2_Rtor_pTEF1_fwd	GCATGAAGACATTCGGTCTCAAACGGACGGCGTGAGGAGC
OL368	2_Rtor_pTEF1_rev	ATCGAAGACATGGTCTCACATATCTTCTATCCCCCCCAGTACACA
OL369	2_Ssti_pTEF1_fwd	GCATGAAGACATTCGGTCTCAAACGCATATACTTATGGACAAGGTTCACT
OL370	2_Ssti_pTEF1_rev	ATCGAAGACATGGTCTCACATATCTTTGTAGATAGACTTAGATTGTATGAAACTAA
OL371	2_Ylip_pTEF1_part1_fwd	GCATGAAGACATTCGGTCTCAAACGAAGAGGCAGAGATGGAGC
OL372	2_Ylip_pTEF1_part1_rev	AGTCGAAGACATTCTGTGGTCGACTTCCAGA
OL373	2_Ylip_pTEF1_part2_fwd	GCATGAAGACATCAGACCGGGTTGGCG
OL374	2_Ylip_pTEF1_part2_rev	ATCGAAGACATGGTCTCACATATCTTTTTGAATGATTCTTATACTCAGAAGGA
OL488	2_pADH1_part1_fwd	ATGAAGACATTCGGTCTCAAACGGGGTGTACAATATGGACTTCCTCTTTTCTGGCAACCA
OL489	2_pADH1_part1_rev	CAGAAGACATGTATTATAGGAATCCCGATGTATGGGTTTGGTTGCCAGAAAAGAGGAAGT

OL490	2_pADH1_part2_fwd	ATGAAGACATATACCTTCGTTGGACTCCCTAACA
OL491	2_pADH1_part2_rev	ACGAAGACATTCTACCATCATTTTTTTGTCATTGCGG
OL492	2_pADH1_part3_fwd	ATGAAGACATTAGACACTAAAGGAAAAAATTAACGACAAAGA
OL493	2_pADH1_part3_rev	ACGAAGACATGGTCTCACATATCTTTTGTATATGAGATAGTTGATTGTATGCTTG
OL494	2_pCYC1_part1_fwd	ATGAAGACATTCGGTCTCAAACGGGAAAAAATTTTTAGGCTCGAGAACAATAG
OL495	2_pCYC1_part1_rev	ACGAAGACATCACGGAAATATAATTCATTTCTTGGTTTTCC
OL496	2_pCYC1_part2_fwd	ATGAAGACATCGTGTGTGACGACATCGTCGAATATGATTCAGGGTAACAGTATTGATGTA
OL497	2_pCYC1_part2_rev	ACGAAGACATCATCTTGATCTTGCTCCCGGGAATTTTAGATTCAGGTAGGAAATTGATTACATCAATACTGTTACCCTGAATCATATTCG
OL498	2_pCYC1_part3_fwd	GATGTTTTCACCGATCTTTCCGGACTCTTTGGCCGGGGTTTACGGACGATGGCAGATGAC
OL499	2_pCYC1_part3_rev	TTTGGTCATCTGCCATCGTCCGTAAACCCCGGCCAAAGAGTCCGGAAAGATCGGTGAAAA
OL500	2_pCYC1_part4_fwd	ATGAAGACATCAAAGCGCCAGTTCATTTGGC
OL501	2_pCYC1_part4_rev	ACGAAGACATGGTCTCACATATATTAATTTAGTGTGTGTATTTGTGTTTGTGTGT
OL77	Part_check_fwd	ATGTTCTTTCCTGCGTTATC
OL78	Part_check_rev	TCGATGAGTTTTTCTAATCAGAAT
OL86	X2_UP_cPCR_fwd	ATCCTTACATTCGGACGAAG
OL84	X2_UP_cPCR_rev	GATCCTTATACTCGATGCCT
OL85	X2_DOWN_cPCR_fwd	GGCGATTCTATAAGAGTTTGG
OL87	X2_DOWN_cPCR_rev	GATTCTGTGAGCCTCTTACC
OL108	Int_cPCR_fwd	ATCCGTCACATATGAGAAGG
OL109	Int_cPCR_rev	CACCAGTCCAAACTCTTATAGA

Supplentary table 2. List of Biobricks used in this study

ID	Name	FWD primers	REV primers	Template DNA
BRICK98	2_Kmar_pTEF1	OL351	OL352	Kluyveromyces marxianus gDNA
BRICK99	2_Cjar_pTEF1	OL353	OL354	Cyberlindnera jardinii gDNA
BRICK100	2_Cpar_pTEF1_part1	OL355	OL356	Candida parapsilosis gDNA
BRICK101	2_Cpar_pTEF1_part2	OL357	OL358	Candida parapsilosis gDNA
BRICK102	2_Cpar_pTEF2	OL359	OL360	Candida parapsilosis gDNA
BRICK103	2_Ssti_pTEF1	OL369	OL370	Scheffersomyces stiptis gDNA
BRICK104	2_Rtor_pTEF1	OL367	OL368	Rhodotorula toruloides gDNA
BRICK105	2_Lsta_pTEF1_part1	OL361	OL362	Lipomyces starkeyl gDNA

BRICK106	2_Lsta_pTEF1_part2	OL363	OL364	Lipomyces starkeyl gDNA
BRICK107	2_Ylip_pTEF1_part1	OL371	OL372	Yarrowla lipolytica gDNA
BRICK108	2_Ylip_pTEF1_part2	OL373	OL374	Yarrowla lipolytica gDNA
BRICK109	2_pADH1_part1	OL488	OL489	-
BRICK110	2_pADH1_part2	OL490	OL491	CEN.PK genomic DNA
BRICK111	2_pADH1_part3	OL492	OL493	CEN.PK genomic DNA
BRICK112	2_pCYC1_part1	OL494	OL495	CEN.PK genomic DNA
BRICK113	2_pCYC1_part2	OL496	OL497	-
BRICK114	2_pCYC1_part3	OL498	OL499	-
BRICK115	2_pCYC1_part4	OL500	OL501	CEN.PK genomic DNA

Supplementary Table 3. List of plasmids used in this study.

ID	**Name**	**Description**	**Template**	**Biobricks**	**Plasmids**	**Source**
Integrative vectors						
pSBME61	BpiI_KanR	Backbone plasmid for entry of parts.				Steven A. van der Hoek inventory
pSBME65	AmpR-X2	Integrative vector for X2 site with Amp resistance				Steven A. van der Hoek inventory
gRNA vectors						
pcfB3020	X-2-gRNA					(Jessop-Fabre et al., 2016)
Part plamids						
pSBME277	2_Kmar_pTEF	Type 2 vector - Kluyveromyces marxianus TEF promoter	pSBME61	BRICK98		This study
pSBME278	2_Cjar_pTEF	Type 2 vector - Cyberlindnera jadinii TEF promoter	pSBME61	BRICK99		This study
pSBME279	2_Cpar_pTEF1	Type 2 vector - Candida parapsilosis TEF1 promoter	pSBME61	BRICK100, BRICK101		This study
pSBME280	2_Cpar_pTEF2	Type 2 vector - Candida parapsilosis TEF2 promoter	pSBME61	BRCK102		This study

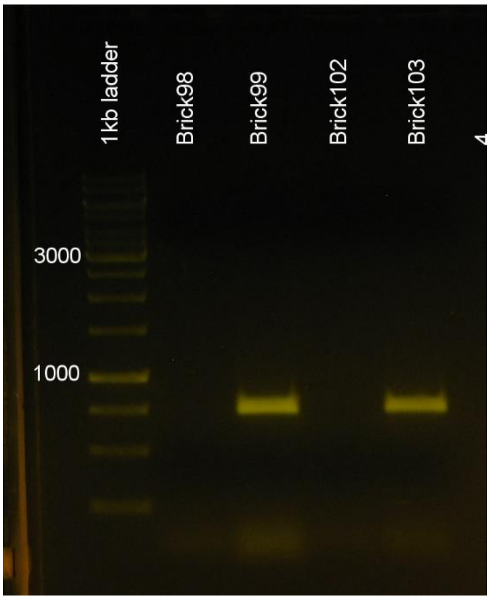
pSBME281	2_Ssti_pTEF	Type 2 vector - Scheffersomyces stiptis TEF promoter	pSBME61	BRCK103		This study
pSBME291	2_Rtor_pTEF	Type 2 vector - Rhodotorula toruloides TEF promoter	pSBME61	BRCK104		This study
pSBME292	2_Lsta_pTEF	Type 2 vector - Lipomyces starkeyi TEF promoter	pSBME61	BRCK105,106		This study
pSBME293	2_Ylip_pTEF	Type 2 vector - Yarrowia Lipolytica TEF promoter	pSBME61	BRICK107,BRICK108		This study
pSBME294	2_pADH1	Type 2 vector - ADH1 promoter	pSBME61	BRICK109,BRICK110,BRICK111		This study
pSBME295	2_pCYC1	Type 2 vector - CYC1 promoter	pSBME61	BRICK112,BRICK113,BRICK114,BRICK115		This study
pSBME95	3_yeGFP	Type 3 vector-yeGFP				Steven A. van der Hoek inventory
pSBME98	4_tADH1	Type 4 vector - ADH1 terminator				Steven A. van der Hoek inventory
Assembled plasmids						
pSBME286	X2-Kmar_pTEF-yeGFP-tADH1	Integrative vector that integrates yeGFP under control of Kmar-pTEF	pSBME65		pSBME277, pSBME95, pSBME98	This study
pSBME287	X2-Cjad_pTEF-yeGFP-tADH1	Integrative vector that integrates yeGFP under control of Cjad-pTEF	pSBME65		pSBME278, pSBME95, pSBME98	This study
pSBME288	X2-Cpar_pTEF1-yeGFP-tADH1	Integrative vector that integrates yeGFP under control of Cpar-pTEF1	pSBME65		pSBME279, pSBME95, pSBME98	This study
pSBME289	X2-Cpar_pTEF2-yeGFP-tADH1	Integrative vector that integrates yeGFP under	pSBME65		pSBME280, pSBME95,	This study

		control of Cpar-pTEF2			pSBME98	
pSBME290	X2-Ssti_pTEF-yeGFP-tADH1	Integrative vector that integrates yeGFP under control of Ssti-pTEF	pSBME65		pSBME281, pSBME95, pSBME98	This study
pSBME296	X2-Rtor_pTEF-yeGFP-tADH1	Integrative vector that integrates yeGFP under control of Rtor-pTEF	pSBME65		pSBME291, pSBME95, pSBME98	This study
pSBME297	X2-Lsta_pTEF-yeGFP-tADH1	Integrative vector that integrates yeGFP under control of Lsta-pTEF	pSBME65		pSBME292, pSBME95, pSBME98	This study
pSBME298	X2-Ylip_pTEF-yeGFP-tADH1	Integrative vector that integrates yeGFP under control of Ylip-pTEF	pSBME65		pSBME293, pSBME95, pSBME98	This study
pSBME299	X2-pADH1-yeGFP-tADH1	Integrative vector that integrates yeGFP under control of pADH1	pSBME65		pSBME294, pSBME95, pSBME98	This study
pSBME300	X2-pCYC1-yeGFP-tADH1	Integrative vector that integrates yeGFP under control of pCYC1	pSBME65		pSBME295, pSBME95, pSBME98	This study

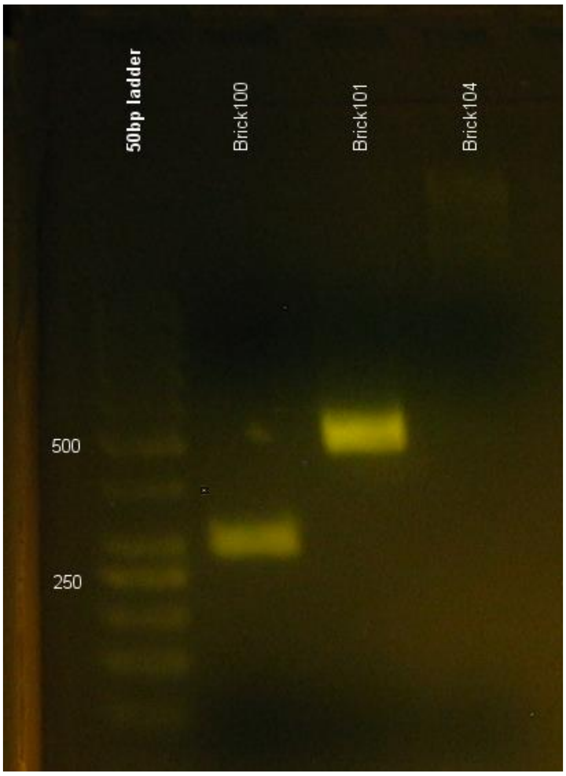
Supplementary Table 4. List of all the strains used in this study.

Strain	**Characteristics**	**Strain specifics**	**Parent strain**	**Genetic edit**	**Source**
SBME1	CEN.PK113-7D	Original strain			Steven A. van der Hoek inventory
SBME6	X4-SpCas9	Background strain			Steven A. van der Hoek inventory
SBME8 (ScTEF)	X4-SpCas9-X2-yeGFP	Strain expressing GFP under own promoter			Steven A. van der Hoek inventory
SBME34 (KmTEF)	X4-SpCas9_X2_Kmar_pTEF_yeGFP_tADH1	Strain expressing GFP under Kmar promoter	SBME6	pSBME286	This study

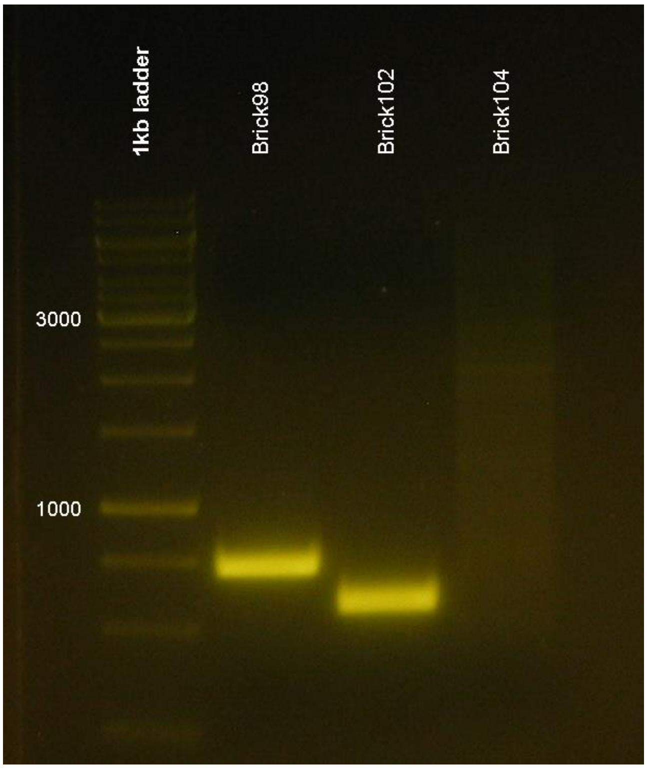
SBME35 (CjTEF)	X4-SpCas9_X2 Cjad_pTEF_yeGTP_tADH1	Strain expressing GFP under Cjad promoter	SBME6	pSBME287	This study
SBME36 (CpTEF1)	X4-SpCas9_X2 Cpar_pTEF1_yeGFP_tADH1	Strain expressing GFP under Cpar TEF1 promoter	SBME6	pSBME288	This study
SBME37 (CpTEF2)	X4-SpCas9_X2 Cpar_pTEF2_yeGFP_tADH1	Strain expressing GFP under Cpar TEF2 promoter	SBME6	pSBME289	This study
SBME38 (SsTEF)	X4-SpCas9_X2 Sstl_pTEF_yeGFP_tADH1	Strain expressing GFP under Sstl promoter	SBME6	pSBME290	This study
SBME39 (YlTEF)	X4-SpCas9_X2_Ylip_pTEF_yeGFP_tADH1	Strain expressing GFP under Ylip promoter	SBME6	pSBME298	This study
SBME40 (ADH1)	X4-SpCas9_X2-pADH1-yeGFP-tADH1	Strain expressing GFP under ADH1 promoter	SBME6	pSBME299	This study



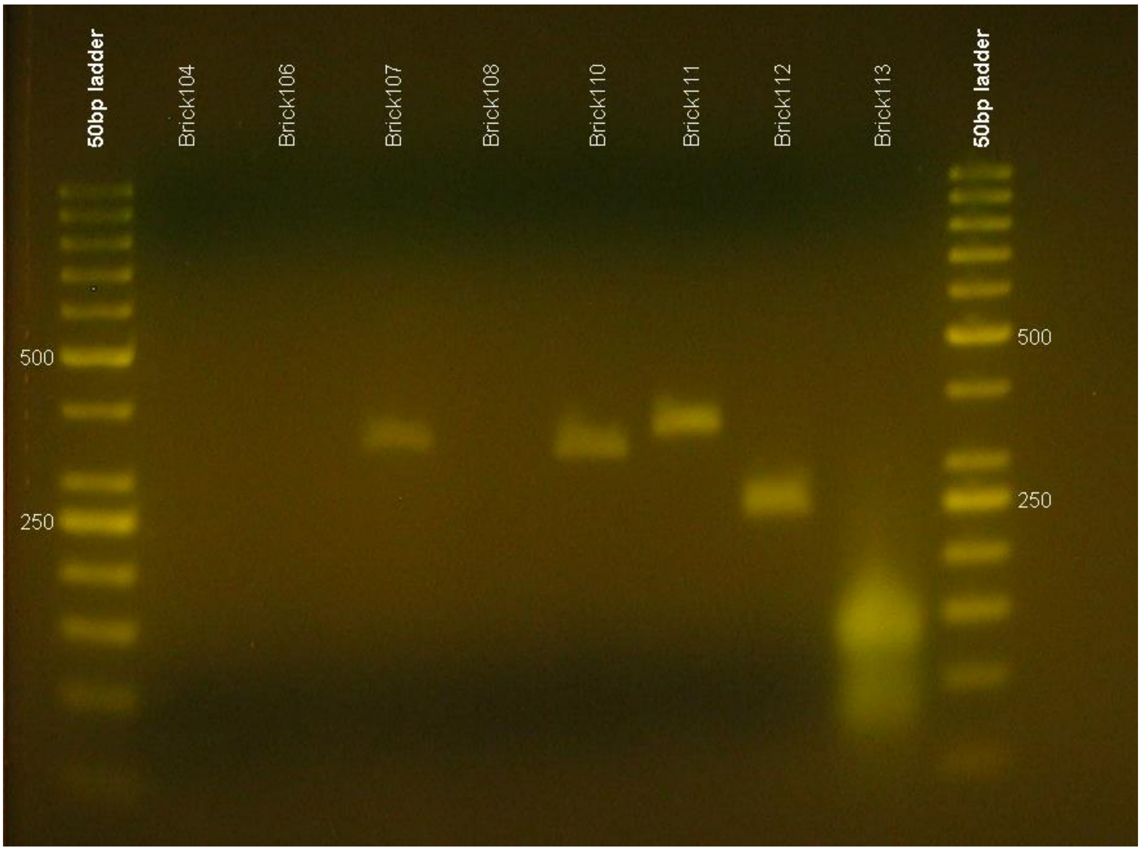
Supplementary Figure 1. Agarose gel electrophoresis of the first set of promoter Biobricks. Biobrick99 and Biobrick103 were as expected, around 750 base pairs. Biobrick98 and 102 were unsuccessful.



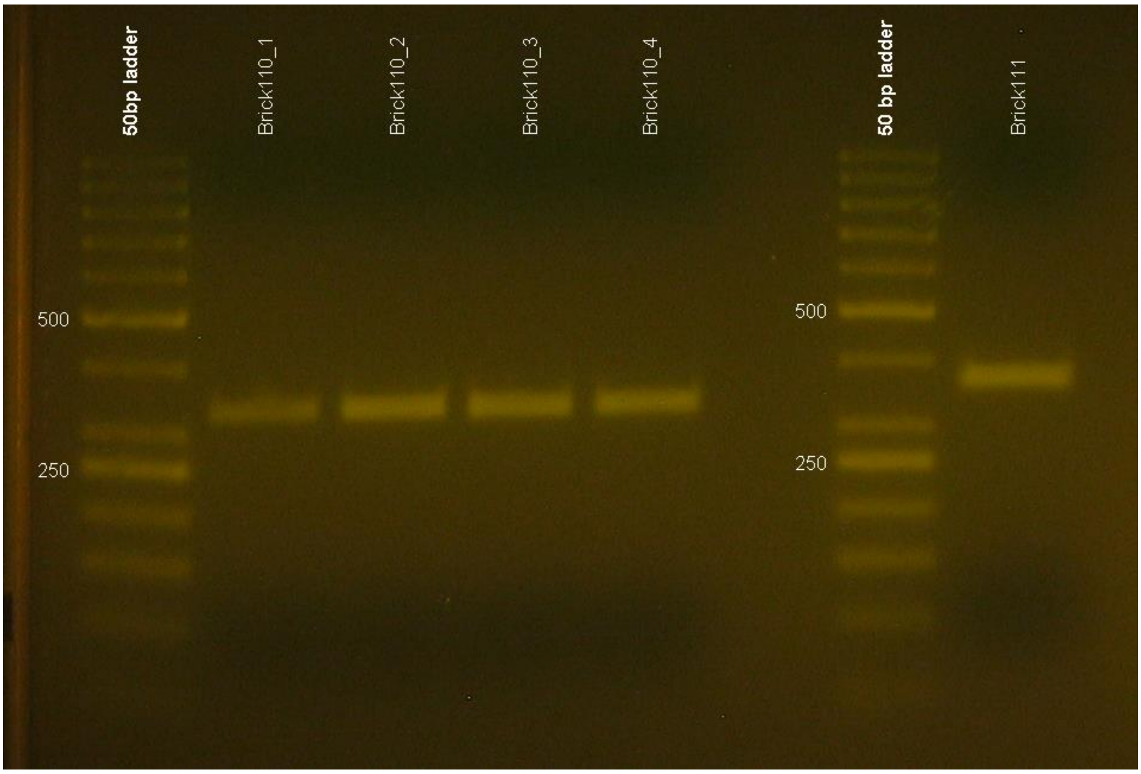
Supplementary Figure 2. Agarose gel electrophoresis of the second set of promoter Biobricks. Biobrick100 and Biobrick101 were as expected, around 290 and 488 base pairs, respectively. Biobrick104 was unsuccessful.



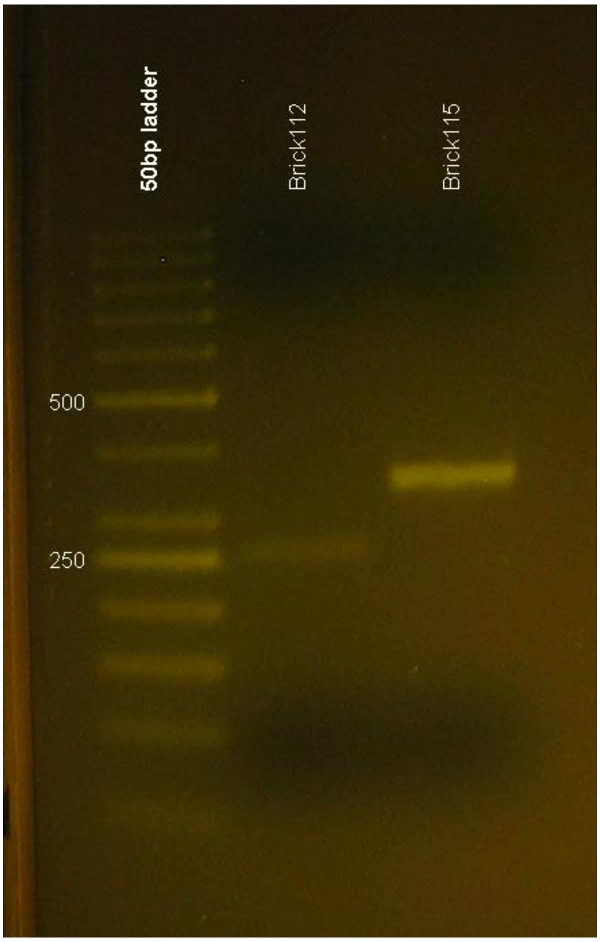
Supplementary Figure 3. Agarose gel electrophoresis, successful obtention of Brick98 and 102, as expected, around 750 and 603 base pairs, respectively. Biobrick104 unsuccessful.



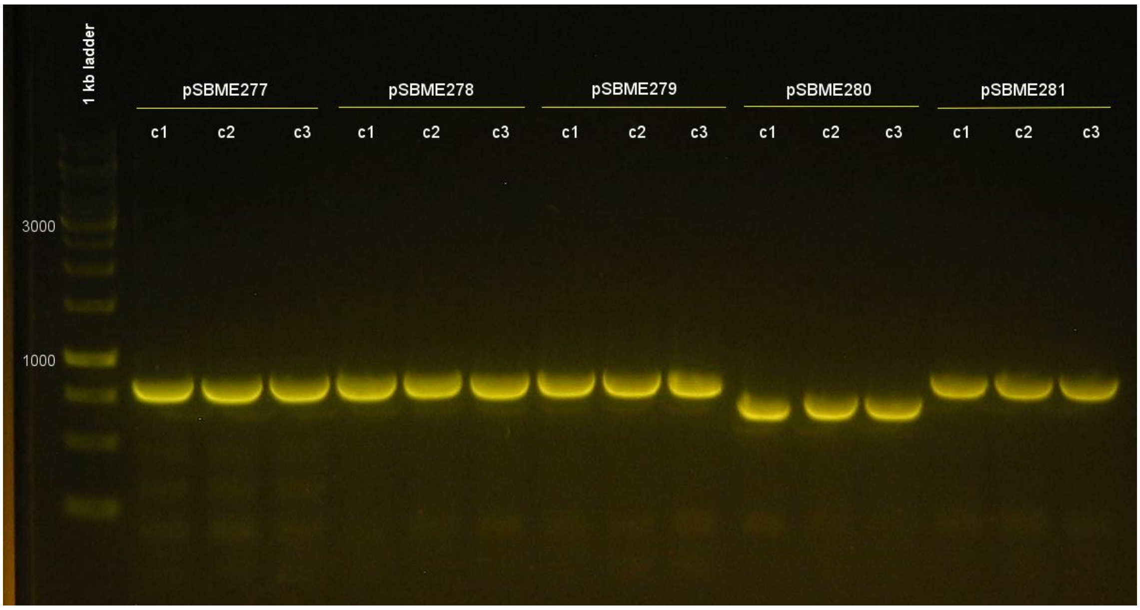
Supplementary Figure 4. Agarose gel for second set of promoters Biobricks. Biobrick 107,110,111 and 112 were successful. Biobrick104, 106 and 113 did not work out.



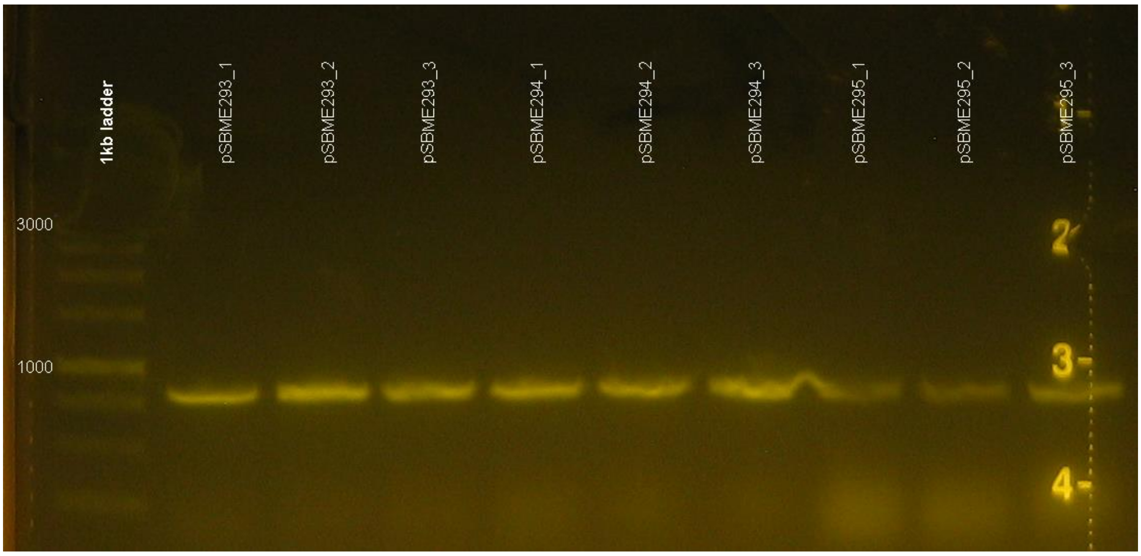
Supplementary Figure 5. Agarose gel for Biobrick110 and Biobrick 111. All samples have correct sizes.



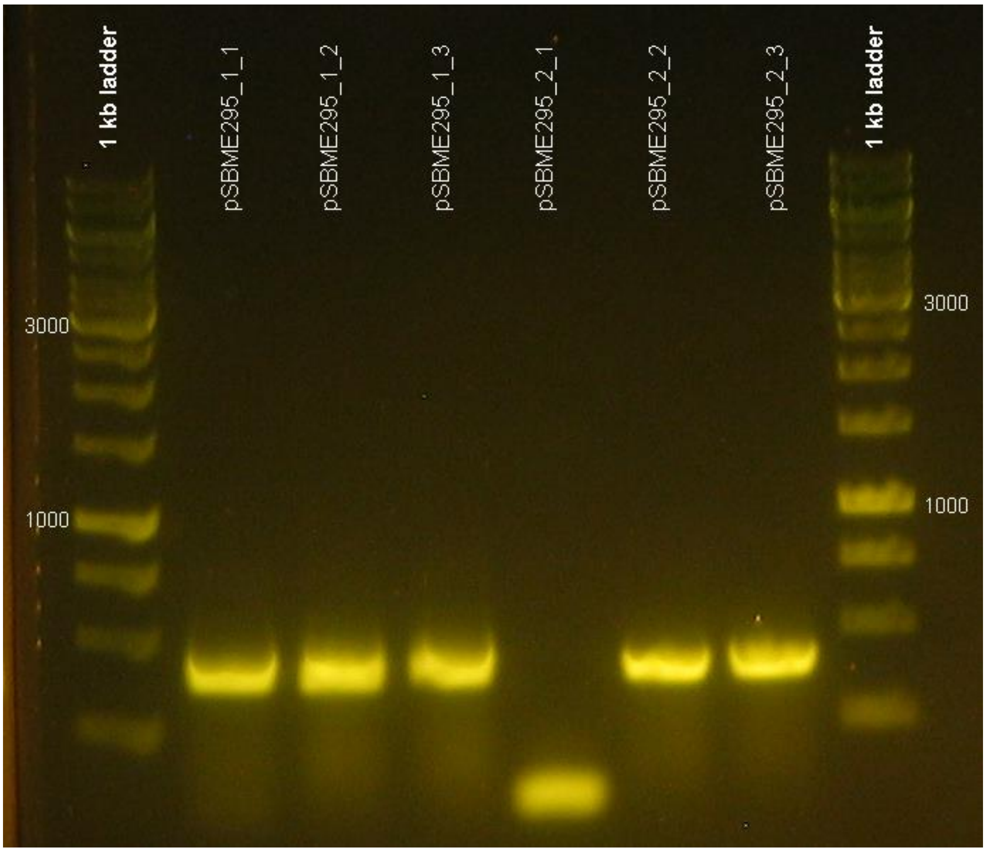
Supplementary Figure 6. Agarose gel for Biobricks 112 and 115. All sizes are correct.



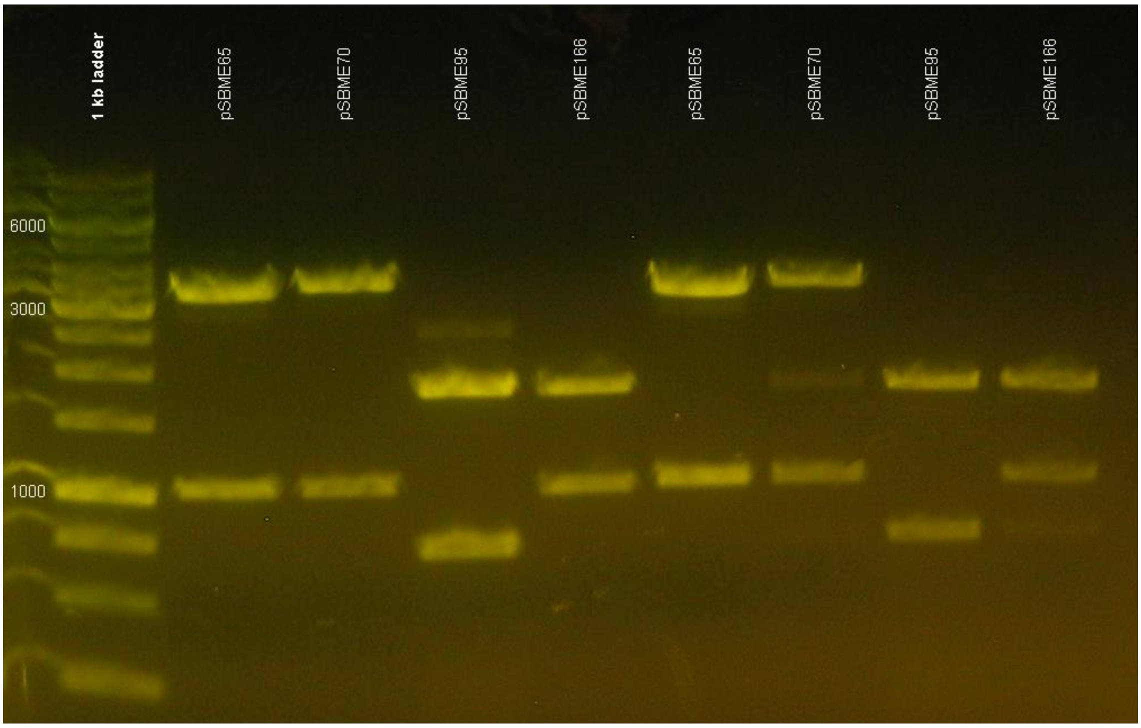
Supplementary Figure 7. Agarose gel for colony PCR for first set of part plasmids. The results were as expected.



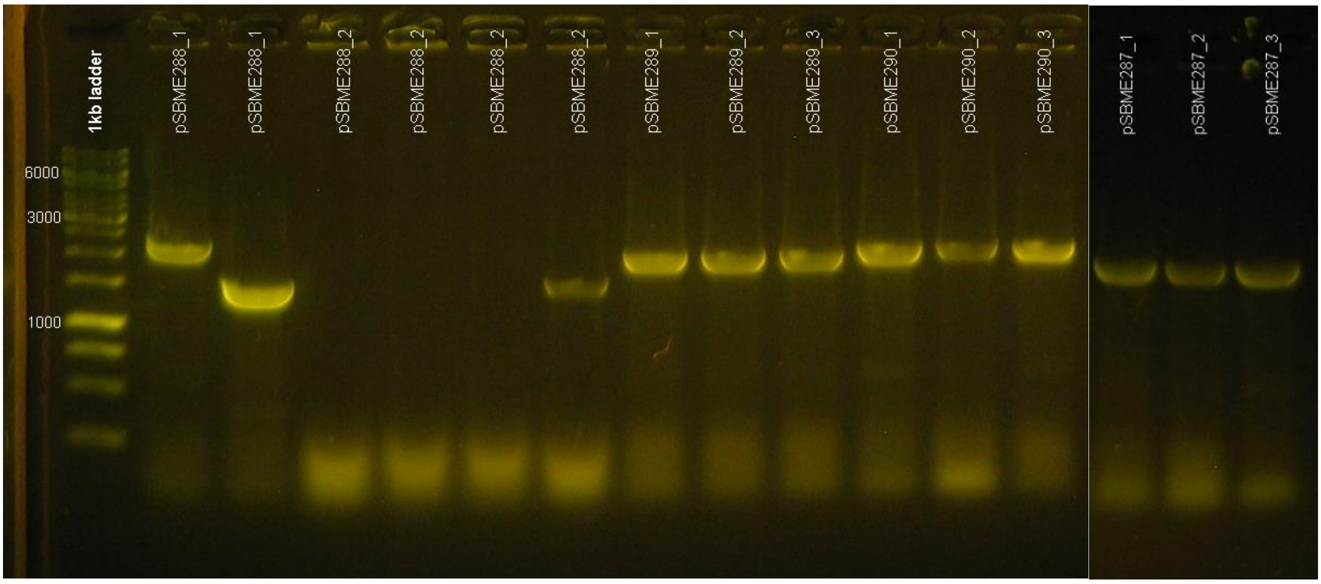
Supplementary Figure 8. Agarose gel for cPCR for second set of part plasmids. All samples have correct sizes, and colony number 1 was selected from each plate for the next experiment



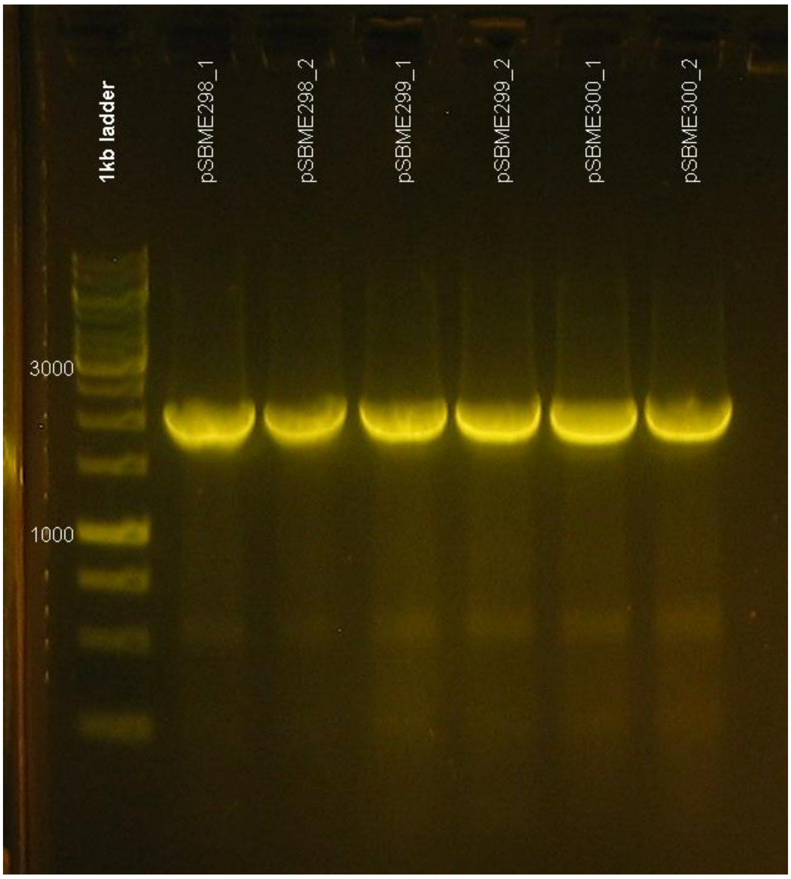
Supplementary Figure 9. Agarose gel for part plasmids expressing the promoter of CYC1. Samples should be 2479 base pairs, the plasmids are smaller than they should be, thus we don't continue the experiment.



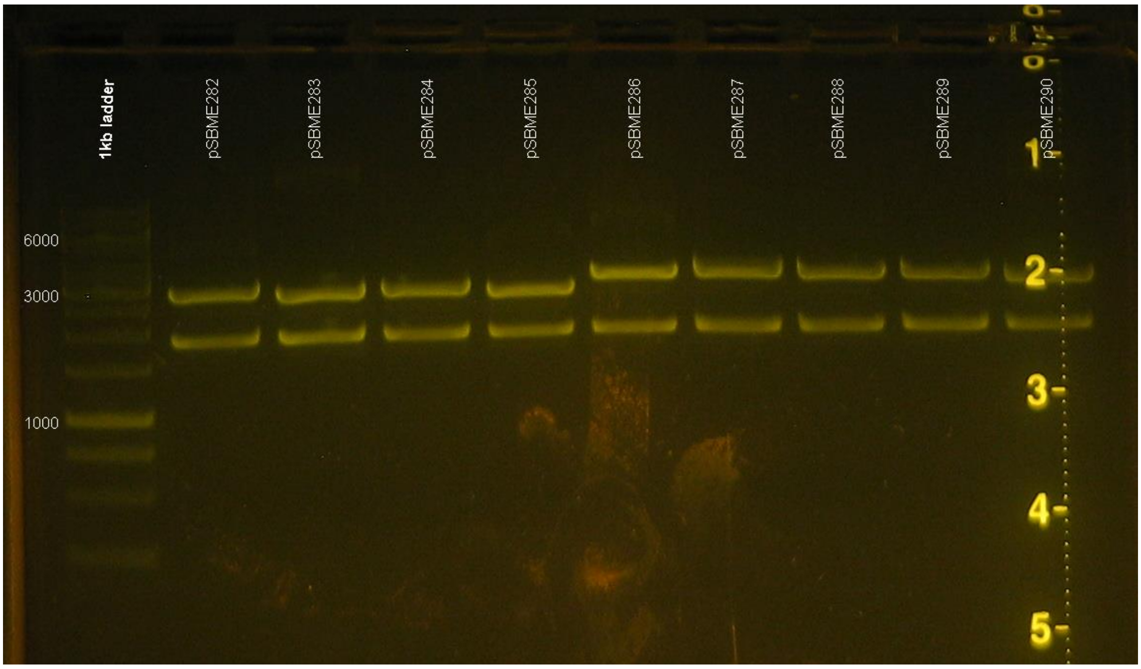
Supplementary Figure 10. Agarose gel for checking if restriction enzymes are working properly. Our cassette plasmid backbone, another similar backbone, our part plasmid expressing GFP, and another plasmid cut by BsaI are treated with the restriction enzyme. If restriction enzymes are working, two bands should be displayed.



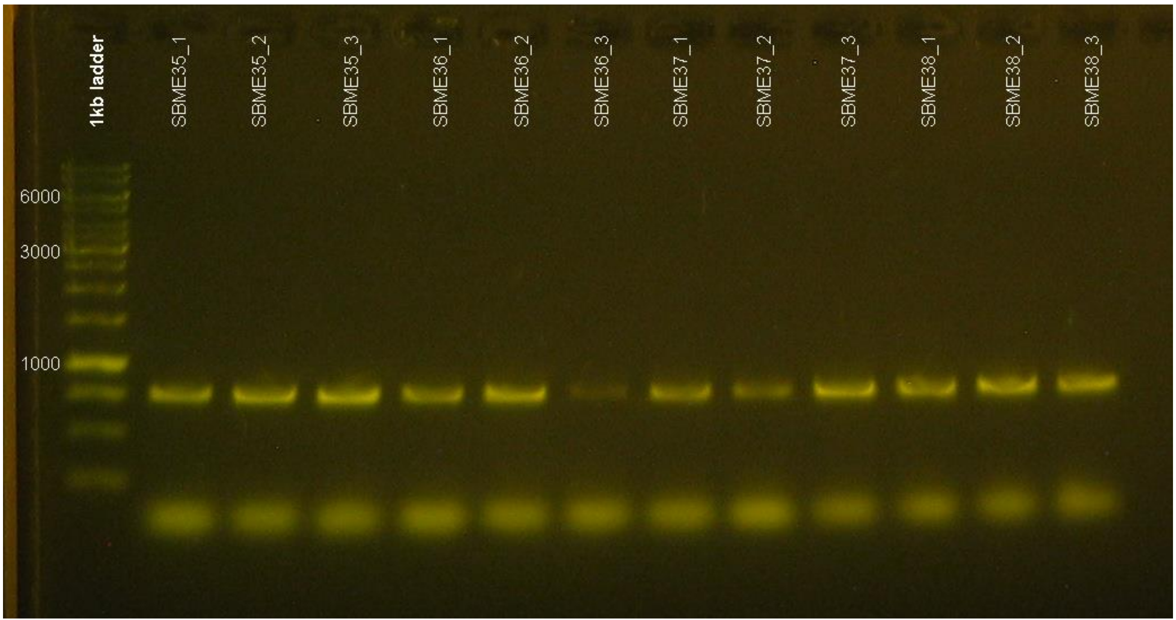
Supplementary Figure 11. Agarose gel for the first set of cassette plasmid. Three colonies for each strain were used. Colony number one was chosen for all strains to continue the experiment.

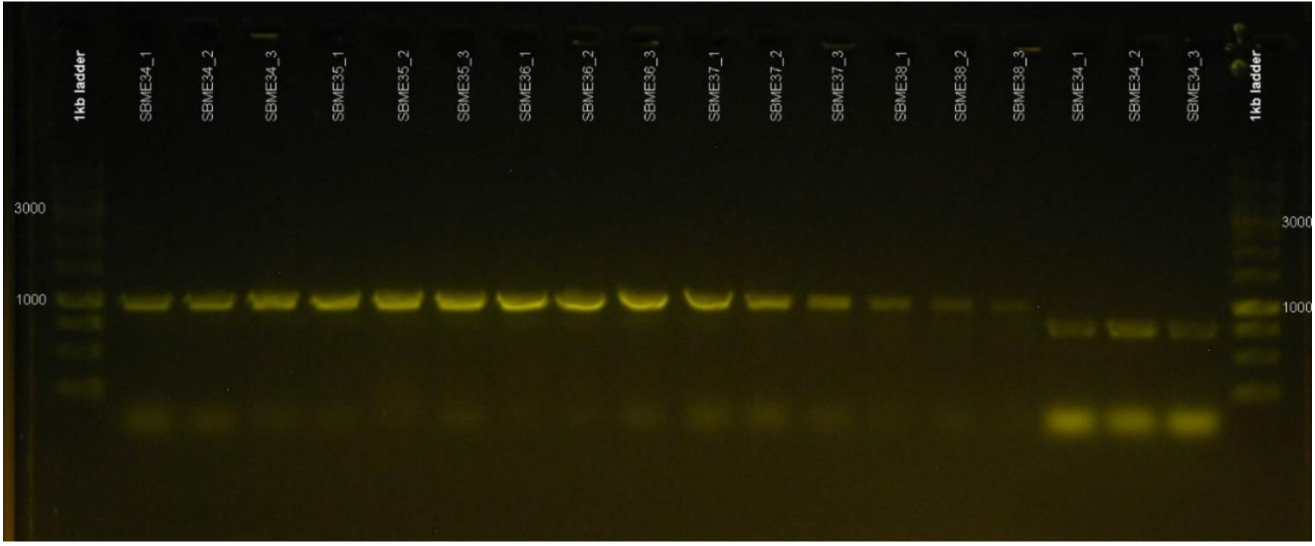


Supplementary Figure 12. Agarose gel for the second set of cassette plasmids. All samples have the correct size.

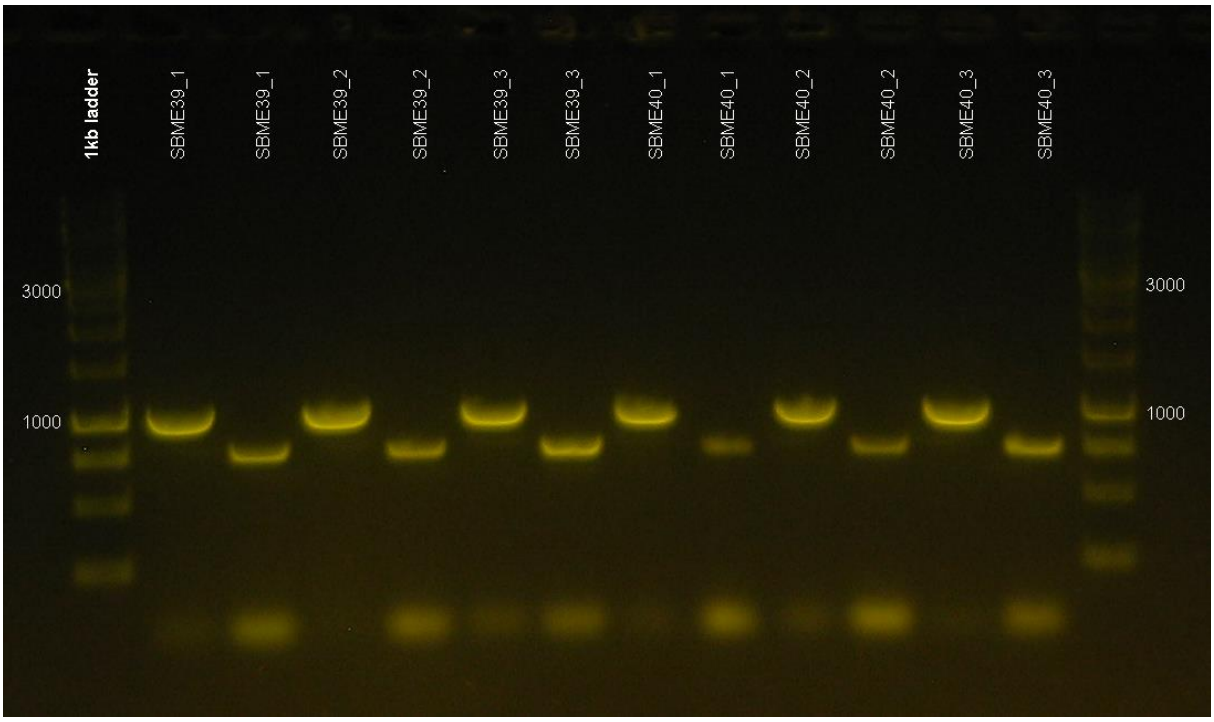


Supplementary Figure 13. Agarose gel for the confirmation of the linearization of the cassette plasmids for yeast transformation. The first band will be the backbone and the second our vector.

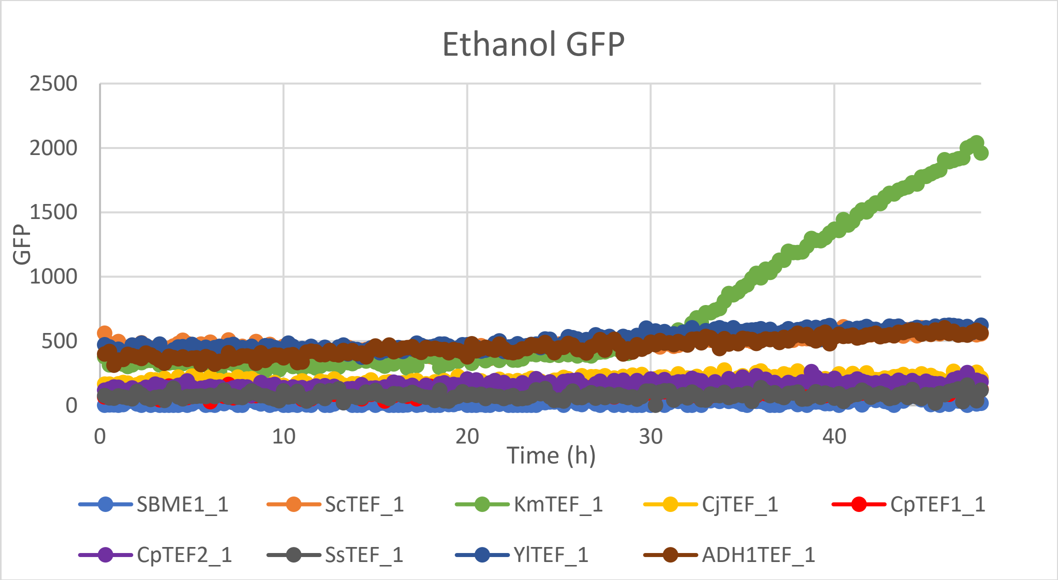




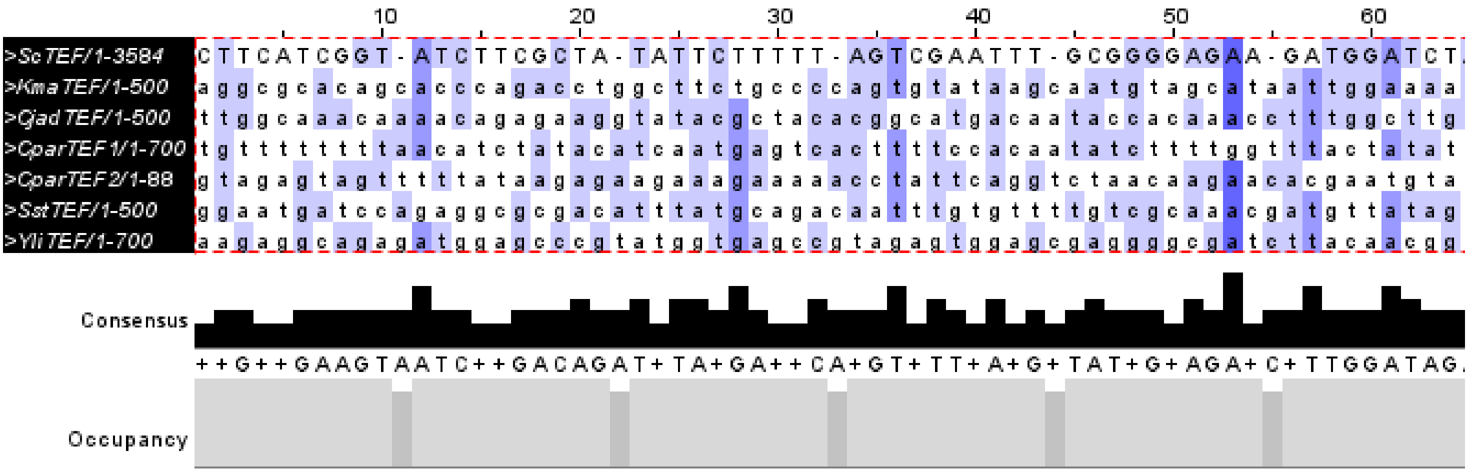
Supplementary figure 14 and 15. Agarose gel for CPCR to verify each colony for integration of X2 UP and DOWN homology arm.



Supplementary Figure 16. Agarose gel to check the integration of the X2 UP and DOWN homology arms. First lane for each strain is with the primers for X2 UP, second lane for X2 DOWN.



Supplementary Figure 17. GFP production of all the strains under Ethanol conditions. KmTEF does show expression of the protein after 30 hours.



.

Supplementary figure 18. Alignment of the promoter sequences with Saccharomyces cerevisiae to check for similarities in pattern and behavior.

NON-EXCLUSIVE LICENCE TO REPRODUCE THESIS AND MAKE THESIS PUBLIC

I, ___Paula Bayón Buján___,

(*author's name*)

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to

reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

___Characterization of non-native promoters in *Saccharomyces cerevisiae*___,

(*title of thesis*)

supervised by ___Steven A. van der Hoek___.

(*supervisor's name*)

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 3.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Paula Bayón Buján

author's name

24/05/2023