



Identification and characterization of sequence signatures in the *Bacillus subtilis* promoter P_{ylb} for tuning promoter strength

Jiangtao Xu · Xiaoqing Liu · Xiaoxia Yu · Xiaoyu Chu · Jian Tian · Ningfeng Wu

Received: 11 July 2019 / Accepted: 18 October 2019 / Published online: 5 November 2019
© The Author(s) 2019

Abstract

Objective To thoroughly characterize the P_{ylb} promoter and identify the elements that affect the promoter activity.

Result The sequences flanking the – 35 and – 10 box of the P_{ylb} promoter were divided into six segments, and six random-scanning mutant promoter libraries fused to an enhanced green fluorescent protein EGFP were made and analyzed by flow cytometry. Our results showed that the four nucleotides flanking the – 35 box could mostly influence the promoter activity, and this influence was related to the GC content. The promoters mutated in these regions were successfully used for expressing the gene *ophc2* encoding organophosphorus hydrolase (OPHC2) and the gene *kata* encoding catalase (KatA).

Conclusion Our work identified and characterized the sequence signatures of the P_{ylb} promoter that could tune the promoter strength, providing further information for the potential application of this promoter. Meanwhile, the sequence signatures have the potential to be used for tuning gene expression in enzyme production, metabolic engineering, and synthetic biology.

Keywords *Bacillus subtilis* · Promoter · Random-scanning mutant · P_{ylb}

Introduction

Bacillus subtilis has been developed as a convenient host for the production of heterologous proteins and industrial enzymes (van Dijn and Hecker 2013). As the promoter is one of the key factors for efficiently expressing heterologous proteins (Blazek and Alper 2013), a series of native promoters from *B. subtilis* have been identified and successfully used in *B. subtilis* for gene expression, such as the promoter P43 (Zhang et al. 2005), P_{xyIA} (Bhavsar et al. 2001), P_{sacB} (Steinmetz et al. 1985) and P_{glv} (Ming et al. 2010). However, more endogenous promoters need to be explored for producing heterologous proteins.

Meanwhile, efforts have been made to enhance promoter strength. The most conventional strategy is modifying moderately conserved sequences of the

Jiangtao Xu and Xiaoqing Liu have contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10529-019-02749-4>) contains supplementary material, which is available to authorized users.

J. Xu · X. Liu · X. Yu · X. Chu · J. Tian (✉) · N. Wu (✉)

Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China
e-mail: tianjian@caas.cn

N. Wu
e-mail: wuningfeng@caas.cn

promoter, such as the UP element, the core region (− 35 and − 10 motifs), and the 16 region (TRTG motif) within existing promoters (Cheng et al. 2016; Guan et al. 2016; Lee et al. 2010; Phan et al. 2012, 2015; Zhou et al. 2019). Furthermore, mutagenesis of the spacer region between the − 35 and − 10 regions in *E. coli* and *Lactobacillus (Lb.) plantarum* has successfully created high-coverage synthetic promoter libraries (De Mey et al. 2007). Saturation mutagenesis of a *Lactococcus lactis* promoter drastically modulates expression (Jensen and Hammer 1998). In addition, randomization of the spacer region of P_{ymdA} and P_{serA} in *B. subtilis* has generated mostly strong to medium promoters (Guiziou et al. 2016). Thus, mutagenesis of the spacer region is considered a rational methodology to modify prokaryotic promoter strength.

Previously, we had identified a highly efficient stationary phase promoter P_{y1b} in *B. subtilis*. The P_{y1b} promoter could induce higher levels of expression of active β -galactosidase, EGFP, RFP, pullulanase, and organophosphorus hydrolase from the late log phase to the stationary phase than that of the widely used P43 promoter, demonstrating strong potential to be used for the overexpression of useful proteins in *B. subtilis*. By scanning site-directed mutagenesis on the P_{y1b} promoter, we found that not only the − 35 and − 10 regions, but also the regions flanking the − 35 and − 10 regions were directly involved in the transcriptional activity of P_{y1b} (Yu et al. 2015). In this study, we focused on the flanking sequences to thoroughly characterize the P_{y1b} promoter and identify the elements that affect the promoter activity. Our work clarified the effect of the identified promoter sequence signatures on the activity of the P_{y1b} promoter. Meanwhile, the promoters libraries with the mutation in the sequence signatures have a wide range of activities, and have the potential to be used for tuning gene expression in enzyme production, metabolic engineering, and synthetic biology.

Materials and methods

Bacterial strains, plasmids and growth conditions

The bacterial strains and plasmids used in this study are listed in Supplementary Table 1. The *Escherichia coli* strain DH5 α was used as the host for gene cloning

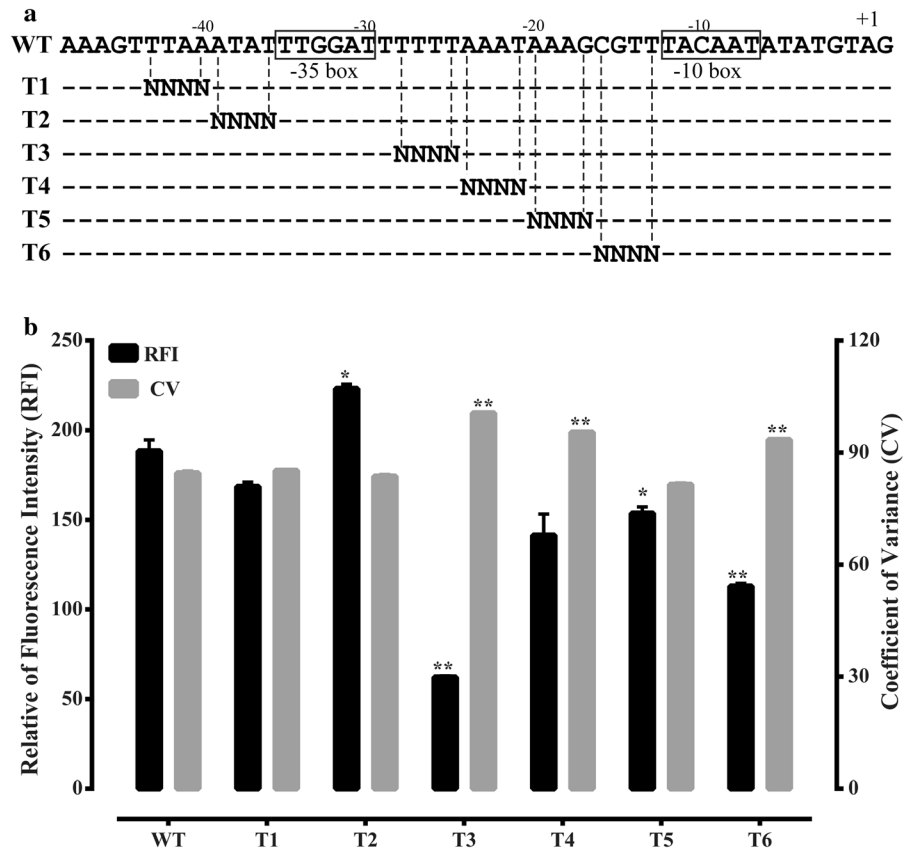
and the construction of the mutated promoter libraries. The *B. subtilis* strain WB600 was used for promoter screening and gene expression. All the strains were cultured in Luria–Bertani medium (LB) at 37 °C under constant shaking (200 rpm). The concentrations of antibiotics used for selection were as follow: 100 μ g/mL ampicillin (Amp), 10 μ g/mL kanamycin (Kan), 50 μ g/mL tetracycline (Tet).

Construction of P_{y1b} random-scanning mutagenesis libraries

Six P_{y1b} random-scanning mutagenesis libraries were constructed by randomly mutating the four nucleotides in six segments, which located in the P_{y1b} region from position − 43 to − 10 (relative to TSS) (Fig. 1a). First, using the plasmid P_{y1b} -G-P43-R-pUBC19 in which the promoter P_{y1b} controls the transcription of the reporter gene *egfp* encoding enhanced green fluorescent protein EGFP as the template, six mutant promoter libraries were generated by PCR with six degenerate primers P16-F-1, -2, -3, -4, -5 and -6 and the reverse primer V-R (Supplementary Table 2), the linear vector fragments were amplified with the forward primers P16-Fr-1, -2, -3, -4, -5 and -6 and the same reverse primer V-F (Supplementary Table 2). Following, the gel-purified promoter and vector fragments were assembled through POE-PCR (You and Percival Zhang 2012). The reaction solution had a total volume of 50 μ L and contained 0.2 mM dNTP, 400–500 ng promoter fragment, the vector fragment (equimolar with the promoter fragment), 1.5 μ L DMSO and 0.04 U/ μ L NEB high-fidelity Phusion DNA polymerase. The POE-PCR conditions were set as follows: 98 °C for 30 s, followed by 30 cycles of 10 s at 98 °C, 30 s at 55 °C, and 4.5 min at 72 °C, with a final extension step at 72 °C for 10 min. The POE-PCR products were isopropanol precipitated, resuspended in 10 μ L ddH₂O and then transformed into competent *E. coli* DH5 α cells. All the *E. coli* transformants were collected together with sterile ddH₂O and the plasmids from these cells were extracted by the TIANprep Mini Plamid Kit II (Tiangen Biotech, Beijing, China). Lastly, the plasmids of these P_{y1b} mutant libraries were transformed into the *B. subtilis* WB600 strain (Spizizen 1958) to construct the corresponding random-scanning mutagenesis libraries T1–T6.

Fig. 1 Construction and characterization of the P_{y1b} random-scanning mutagenesis libraries.

a Schematic representation of the random-scanning mutagenesis target sites on the P_{y1b} promoter to construct the libraries. The promoter sequences were randomly mutagenized by the degenerate sequences within the corresponding primers. The -35 and -10 boxes are denoted by the open squares. The $+1$ mark represents the transcription start site (TSS). **b** Fluorescence intensity output of mutated library promoters driving expression of *egfp* as determined by flow cytometry in triplicate. * $p < 0.01$; ** $p < 0.001$, relative to the WT control



Characterization of promoter libraries by flow cytometry

To measure the activity of the promoters in the six random-scanning mutagenesis libraries of P_{y1b} , flow cytometry was used to quantify EGFP fluorescence. The overnight cultures of all the recombinant *B. subtilis* strains were centrifuged at $5000 \times g$ for 10 min, and the cells were resuspended in ice-cold $1 \times$ phosphate-buffered saline (PBS) buffer (pH 7.4). Flow cytometry analysis was carried out on a FACSCalibur apparatus (Becton Dickinson Biosciences) using excitation at 488 nm. Three independent replicas were analyzed from each promoter library, and each cytometric measurement was performed on 100,000 cells. The data were acquired and analyzed using the CELLQuest software with defined gates that were measured by tight forward scatter/side scatter light to ensure a homogenous population size. The average relative fluorescence intensity (RFI) given represent the mean values of the geometric mean expression values of each replicate; the standard deviation was

also calculated from the replicate geometric mean values. The value of coefficient of variation (CV) for each promoter library was obtained by dividing the standard deviation of all replicate values by the mean of all replicate values.

Screening of the T2 and T3 mutant libraries

The T2 and T3 mutant libraries were screened for P_{y1b} mutants with respectively higher and lower activities by using a BD Influx flow cytometer. The separation gates in the dot plots of the two-parameter histogram were reset beyond the negative control (*B. subtilis* WB600) values and the positive control (wild type P_{y1b}) values. P4 gate was used to collect up-regulated P_{y1b} mutants, while the P2 gate collected the extremely weak mutants (Supplementary Fig. 1). The fluorescent intensity per OD_{600} (FI/ OD_{600}) of these selected mutants were measured by a SpectraMax M2 (Molecular Devices, USA) microplate reader. The excitation/emission was read at 484/507 nm. The promoters of these selected mutants were sequenced on an ABI

3730 DNA analyzer (Qingke Biotechnology Co., Ltd.) with the sequencing primers F4-cx-up and F4-cx-down (Supplementary Table 2). All the sequences were analyzed using the WebLogo online software (<https://weblogo.berkeley.edu>) to explore the relationship between signature sequences and the strength of P_{ylb} .

Construction of site-directed mutagenesis promoters

The promoter P_{ylb} in the plasmids P_{ylb} -G-P43-R-pUBC19, P_{ylb} -R-pUBC19, P_{ylb} -ophc2-pUBC19 and P_{ylb} -katA-pUBC19 were mutated into P_{ylb} -T2G, P_{ylb} -T2G and P_{ylb} -T3G, respectively (Supplementary Table 1). To do this, these four plasmids were used as templates to amplify the fragments containing the promoters P_{ylb} -T2G, P_{ylb} -T2G and P_{ylb} -T3G by the primer pairs P16-T2GF/V-R, P16-T2CF/V-R and P16-T3GF/V-R, and the vector fragments by the primer pairs P16-Fr-2 (P16-Fr-3)/V-F (Supplementary Table 2). The corresponding promoter and vector fragments were assembled through POE-PCR, resulting in the 12 plasmids pGT2G, pGT2C, pGT3C, pRT2G, pRT2C, pRT3C, pOT2G, pOT2C, pOT3C, pKT2G, pKT2C and pKT3C (Supplementary Table 1). The sequences of these mutated plasmids were confirmed by DNA sequencing, and then transformed into *B. subtilis* WB600.

Determination of the transcription levels of *egfp* by quantitative reverse transcription-PCR (qRT-PCR)

To determine the promoter's activity, total RNA was isolated from 18 h cultures of the transformed *B. subtilis* strains EGFP-WT, EGFP-T2G, EGFP-T2C and EGFP-T3C, harboring the recombinant plasmids P_{ylb} -G-P43-R-pUBC19, pGT2G, pGT2C, and pGT3C, using a RNAPrep PureCell/Bacteria Kit (TIANGEN Biotech, Beijing, China). The first strand of cDNA was synthesized using TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix (TransGen Biotech, Beijing, China) with the gene-specific primers *egfp*-rtR and BS16S-rtR (Supplementary Table 2). The qRT-PCR was performed using the SYBR Green Real-time PCR Master Mix Plus (Toyobo, Osaka, Japan) with the primers *egfp*-rtF and *egfp*-rtR (Supplementary Table 2) for the *egfp* gene and

primers BS16S-rtF and BS16S-rtR (Supplementary Table 2) for the *16S*rDNA gene.

Determination of the expression levels of green fluorescent protein EGFP and red fluorescent protein mApple

To determine the yield of fluorescent protein, the overnight cultures of the recombinant *B. subtilis* strains were inoculated into a 96-well microtiter plate containing 200 μ L LB liquid medium with 10 μ g/mL kanamycin per well, and then incubated at 37 °C with shaking at 750 rpm in an incubator 1000 (Heidolph, Germany) for 20 h. Fluorescence intensity and cell density (OD₆₀₀) were measured using a SpectraMax M2 (Molecular Devices, USA) microplate reader. The excitation/emission was read at 484/507 nm for EGFP and 562/598 nm for mApple. Samples were collected from 20 h cultures for SDS-PAGE analysis.

Heterologous protein expression and enzyme activity analysis

A fresh overnight culture of the recombinant *B. subtilis* strains OPHC2-WT, OPHC2-T2G, OPHC2-T2C, and OPHC2-T3C, KatA-WT, KatA-T2G, KatA-T2C, and KatA-T3C were inoculated into 50 mL LB liquid medium containing 10 μ g/mL kanamycin and cultivated at 37 °C, at 200 rpm, for 28 h. Organophosphorus hydrolase (OPHC2) activities were determined as described previously (Yu et al. 2015). One unit of OPHC2 activity was defined as the amount of the enzyme required to liberate 1 μ mol of *p*-nitrophenol per minute at 37 °C. For catalase activity, the reaction was performed as described previously (Goldblith and Proctor 1950) and one unit of catalase activity was defined as the amount of enzyme required to catalyze the decomposition of 1 μ mol H₂O₂ per minute at 30 °C.

Results

Construction and characterization of the P_{ylb} random-scanning mutagenesis libraries

To identify the key elements that influence the promoter strength in the regions flanking the – 35 and – 10 regions of P_{ylb} , six mutant promoter

libraries, T1–T6, were constructed (Fig. 1a) and EGFP fluorescence was quantified using flow cytometry to determine the *egfp* expression levels. The T2 library had the highest RFI (18.7% increase), while the T3 library had the lowest RFI (67.0% decrease), relative to the wild type P_{y1b} (Fig. 1b). Meanwhile, the highest CV was observed in the T3 library (Fig. 1b). These data suggested that the four nucleotides flanking the – 35 box of P_{y1b} have the most critical influence on promoter strength and randomized mutations in the T3 locus of P_{y1b} were likely to cause more obvious changes in the promoter strength.

Characterization of P_{y1b} mutants collected from the T2 and T3 libraries

To determine the effect of the four nucleotides flanking the – 35 box of P_{y1b} on the promoter activity, the T2 and T3 libraries were selected to sort out the up-regulated and down-regulated variants by flow cytometry and the sequences of these collected promoters were determined by DNA sequencing. 58 clones, of which FI/OD₆₀₀ increased by 30% (Supplementary Fig. 2a), were collected from the T2 library, and these represented 39 different mutated promoters (Supplementary Fig. 3a). 39 clones, with FI/OD₆₀₀ 80% lower than the wild type (Supplementary Fig. 2b), were collected from the T3 library, representing 20 different mutated promoters (Supplementary Fig. 3b). To detect any features of residues at the T2 and T3 locus, all these sequences were subjected to WebLogo analysis. Our findings revealed that the sequences in the T2 and T3 locus of these collected promoter mutants were GC-rich and C-rich, respectively (Fig. 2). These results suggested that the GC bases in the T2 and the C content in the T3 locus were respectively positively

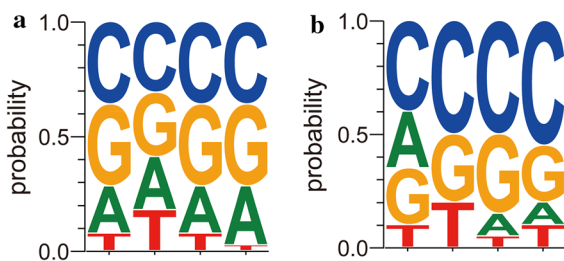


Fig. 2 Sequence features of the P_{y1b} mutants collected from the **a** T2 and the **b** T3 library; the data was analyzed by the WebLogo online software (<https://weblogo.berkeley.edu>)

and negatively correlated with the transcriptional activity of P_{y1b} .

Characterization of the effect of the identified T2 and T3 locus signatures on the promoter activity

To further test the effect of the GC content in the T2 locus and the C content in T3 locus on the promoter activity, the P_{y1b} was mutated into P_{y1b} -T2G and P_{y1b} -T2C by replacing “ATAT” in the T2 locus by “GGGG” and “CCCC”, respectively, and into P_{y1b} -T3C by replacing “TTTT” in the T3 locus by “CCCC”. First, the activities of these three mutant promoters and the native P_{y1b} were measured by qRT-PCR. The data showed that the activities of the P_{y1b} -T2G and P_{y1b} -T2C were higher than that of P_{y1b} , while the activities of the P_{y1b} -T3C were much lower than that of P_{y1b} (Fig. 3a). Then, the activities of these promoters were further analyzed by assessing the expression of the reporter gene *egfp* in the plasmids P_{y1b} -G-P43-R-pUBC19, pGT2G, pGT2C, and pGT3C in the *B. subtilis* strain WB600. The expression of EGFP in these strains displayed a continuous increase, and the overall expression levels of EGFP at each time point in the *B. subtilis* strains EGFP-T2G and EGFP-T2C were the highest, followed by EGFP-WT, and EGFP-T3C (Fig. 3b). Compared to EGFP-WT, the expression levels of EGFP in the EGFP-T2G and EGFP-T2C mutants after 20 h of culturing increased by 51.5% and 55.1%, respectively, while in the EGFP-T3C it decreased by about 74.6% (Fig. 3b). Meanwhile, the relative fluorescent intensity (FI/OD₆₀₀) was measured and the data showed that the expression levels of EGFP in EGFP-T2G and EGFP-T2C was 1.35-fold and 1.40-fold higher than that of EGFP-WT; while the expression levels of EGFP-T3C dropped by 73.2% (Fig. 3c). The final protein expression levels measured by SDS-PAGE were consistent with those measured by fluorescence intensity (Fig. 3d).

To further test the reliability of the signature sequences, the three mutant promoters P_{y1b} -T2G, P_{y1b} -T2C, P_{y1b} -T3C and the native P_{y1b} were used to express the reporter gene *mApple* using the plasmids pRT2G, pRT2C, pRT3C and P_{y1b} -R-pUBC19 in the *B. subtilis* strain WB600. These three mutated promoters have similar regulatory trends between the expression of *mApple* and EGFP (Fig. 4). The expression levels of *mApple* in the *mApple*-T2G and *mApple*-T2C was respectively 1.57-fold and 1.83-fold

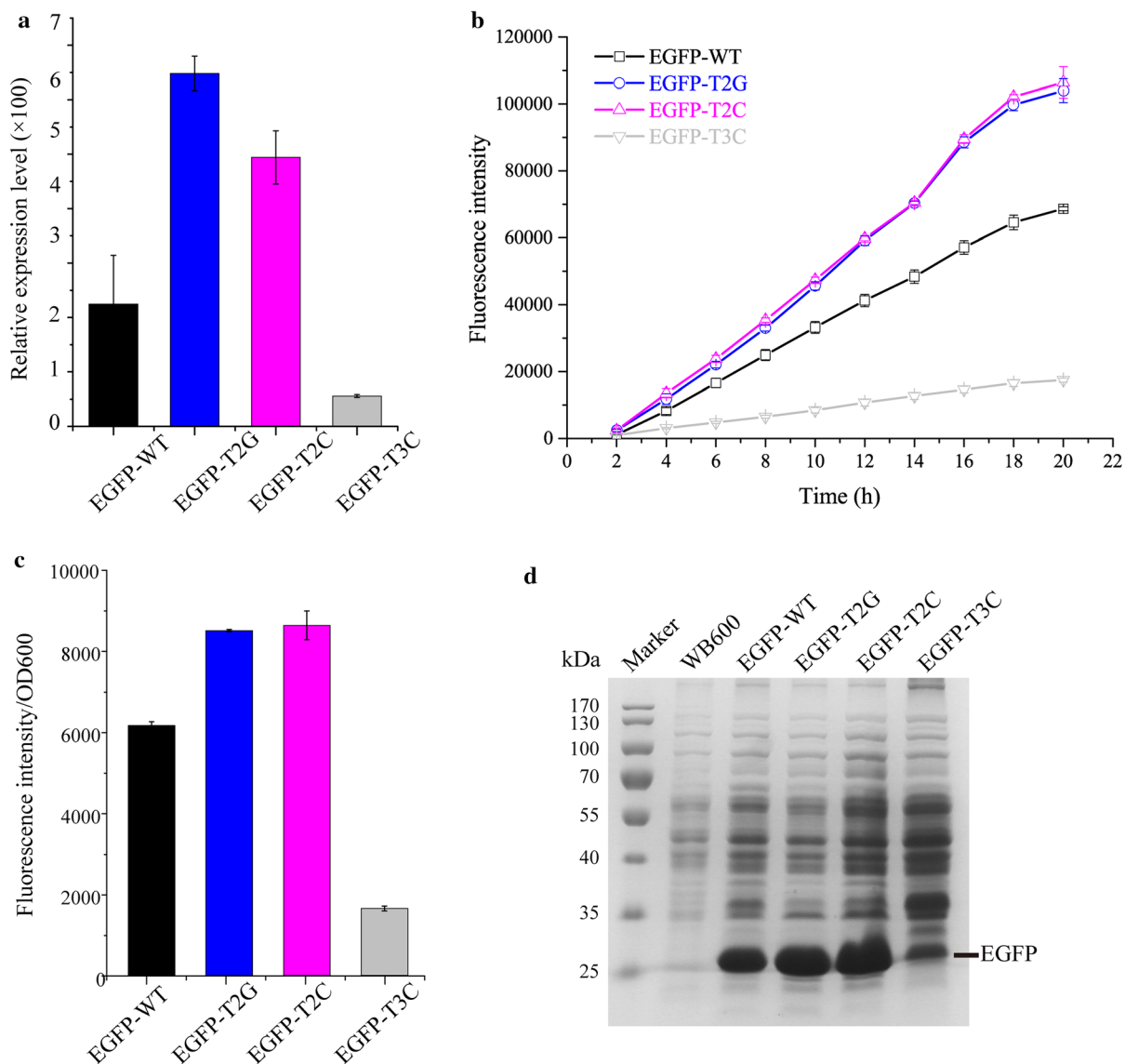


Fig. 3 Characterization of P_{ylb} -T2G, P_{ylb} -T2C, P_{ylb} -T3C variants and native P_{ylb} using the reporter gene *egfp*. **a** qRT-PCR to analyze the relative expression levels of EGFP under the control of wild type and mutated promoters at 18 h. The *16 S rRNA* gene was used as an internal control for normalization. The data were analyzed by the $2^{-\Delta\Delta Ct}$ method. All the data were independently repeated in triplicates. **b** The yields of total fluorescence intensity (FI) in the *B. subtilis* strains harboring

P_{ylb} -G-P43-R-pUBC19 (EGFP-WT), pGT2G (EGFP-T2G), pGT2C (EGFP-T2C) and pGT3C (EGFP-T3C) plasmids during the whole culturing period. **c** The relative fluorescent intensity measured in the EGFP-WT, EGFP-T2G, EGFP-T2C and EGFP-T3C strains, after 20 h of culturing, is represented by the FI per OD₆₀₀. **d** SDS-PAGE analysis of the expression level of EGFP in the EGFP-WT, EGFP-T2G, EGFP-T2C and EGFP-T3C strains after 20 h of culturing

higher than that in the mApple-WT, while the expression levels of mApple in the mApple-T3C decreased by 82.6% (Fig. 4b).

The above results supported that the extracted signature sequences in the T2 and T3 locus could differentially regulate the activity of the P_{ylb} promoter.

The application prospect of the signature sequences

To evaluate the applicability of the signature sequences, the three mutant promoters P_{ylb} -T2G, P_{ylb} -T2C, P_{ylb} -T3C and the native P_{ylb} were used to

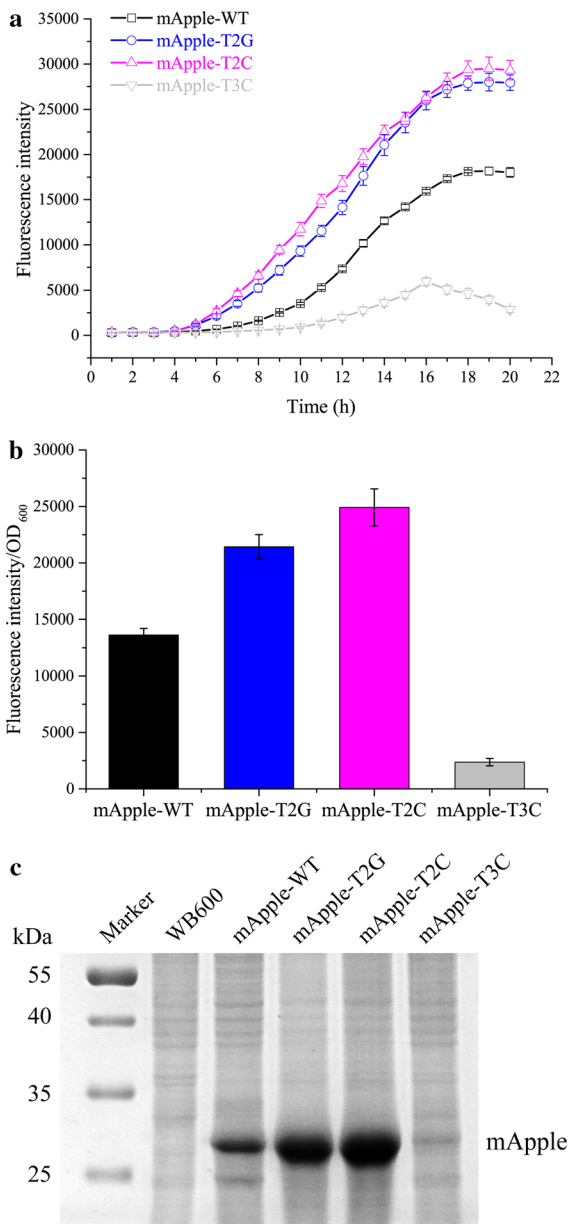


Fig. 4 Characterization of the P_{yib} -T2G, P_{yib} -T2C, P_{yib} -T3C variants and native P_{yib} using the reporter gene *mApple*. **a** The total fluorescence intensity yields in the *B. subtilis* strains harboring the P_{yib} -R-pUBC19 (mApple-WT), pRT2G (mApple-T2G), pRT2C (mApple-T2C) and pRT3C (mApple-T3C) plasmids during the whole culturing period. **b** The relative fluorescent intensity in the mApple-WT, mApple-T2G, mApple-T2C and mApple-T3C strains after 20 h culturing, which is represented by the FI per OD₆₀₀. **c** SDS-PAGE analysis of the expression levels in the mApple in mApple-WT, mApple-T2G, mApple-T2C and mApple-T3C strains after 20 h culturing

express organophosphorus hydrolase OPHC2 in the *B. subtilis* strains OPHC2-T2G, OPHC2-T2C, OPHC2-T3C and OPHC2-WT, and catalase KatA in the *B. subtilis* strains KatA-T2G, KatA-T2C, KatA-T3C and KatA-WT. The OPHC2 activities detected in strain OPHC2-T2G and OPHC2-T2C were 22% and 38% higher than that of strain OPHC2-WT, while there was no obvious enzyme activity in the OPHC2-T3C (Fig. 5a). The catalase activities detected in strain KatA-T2G and KatA-T2C were about 63% lower than that of strain KatA-WT, while the catalase activity detected in strain KatA-T3C was 125% higher than that of strain KatA-WT (Fig. 5c). The results of SDS-PAGE analysis of OPHC2 and KatA in the culture supernatant were shown in Fig. 5b and d. These data indicated that the strong promoters P_{yib} -T2G, P_{yib} -T2C were more suitable for overexpression of OPHC2, whereas, the weak promoter P_{yib} -T3C was more suitable for overexpression of KatA. This means that the signature sequences could be used to tune the promoter strength to meet the requirement of expressing different proteins.

Discussion

Promoters are important regulatory elements for expressing protein. To date, some high-quality native promoters have been widely used in *B. subtilis* for gene expression. However, most of the endogenous promoters in *B. subtilis* were not well explored and comprehensively analyzed. The promoter P_{yib} was found and proved to be a highly active promoter of *B. subtilis* in our lab (Yu et al. 2015). In this study, for better application of the promoter P_{yib} , the elements that influence its activity were researched.

One commonly used strategy for exploring promoter element is to randomly mutate within the whole promoter sequence. It is time-consuming and easily missing information. Our previous work shown that the promoter regions I (position - 38 to - 23, relative to TSS) and II (position - 18 to - 3, relative to TSS) that contain the - 35 and the - 10 regions were particularly important for the transcriptional activity of P_{yib} (Yu et al. 2015). Therefore, we focused on the sequence flanking the - 35 and the - 10 regions to identify and characterize elements, apart from the core

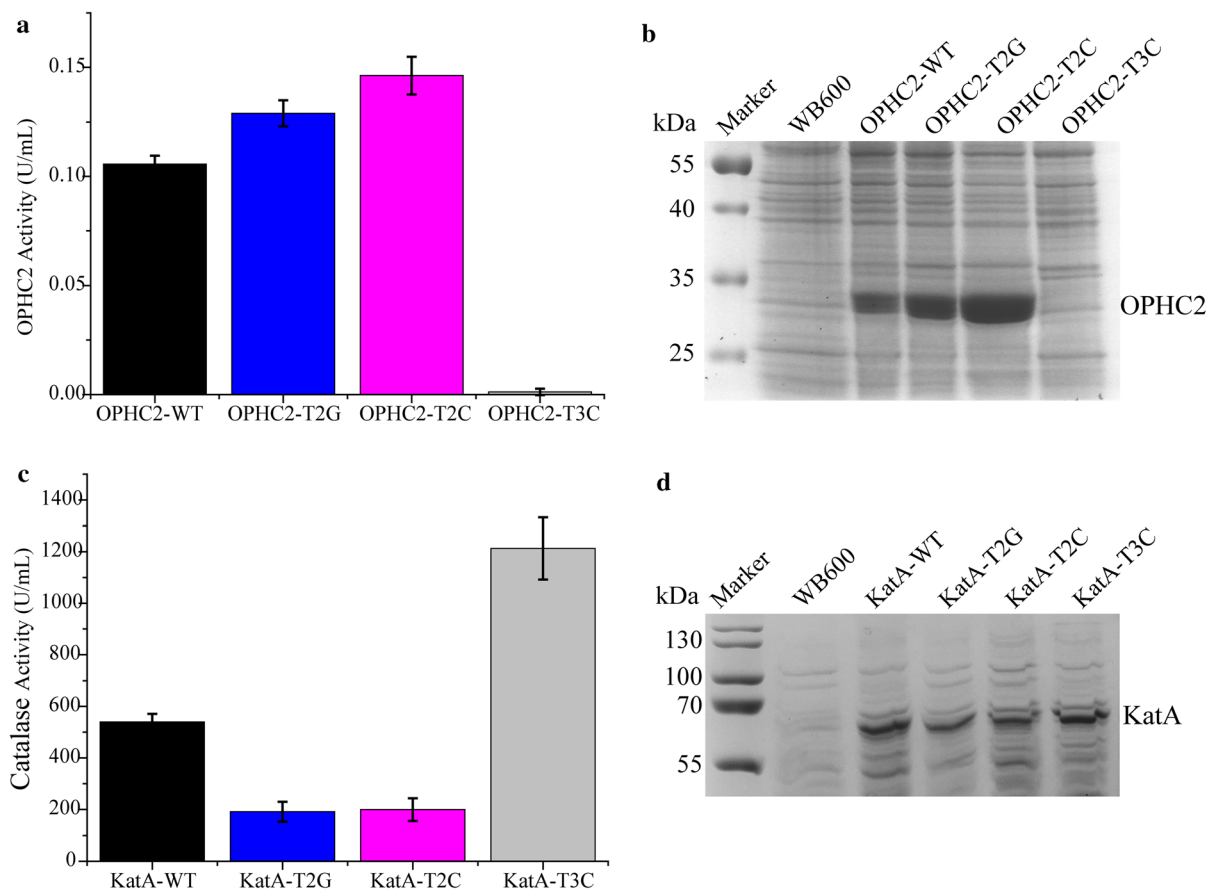


Fig. 5 Expression of organophosphorus hydrolase OPHC2 and catalase KatA in the recombinant *B. subtilis* strain under the control of the variant promoters P_{y1b} -T2G, P_{y1b} -T2C, P_{y1b} -T3C and the native promoter P_{y1b} . **a** The enzymatic activity yields of OPHC2 in the *B. subtilis* strains harboring the P_{y1b} -ophc2-pUBC19 (OPHC2-WT), pOT2G (OPHC2-T2G), pOT2C (OPHC2-T2C) and pOT3C (OPHC2-T3C) plasmids. **b** SDS-PAGE analysis of the expression levels of the OPHC2 in the

OPHC2-WT, OPHC2-T2G, OPHC2-T2C and OPHC2-T3C strains. **c** The enzymatic activity yields of KatA in the *B. subtilis* strains harboring the P_{y1b} -katA-pUBC19 (KatA-WT), pKT2G (KatA-T2G), pKT2C (KatA-T2C) and pKT3C (KatA-T3C) plasmids. **d** SDS-PAGE analysis of the expression levels of the catalase in the KatA-WT, KatA-T2G, KatA-T2C and KatA-T3C strain

elements, which influence the promoter strength. In our study, the sequence to be interested in were divided into six groups and each group contained four nucleotides (Fig. 1a), so that each random mutagenesis library potentially contained 256 mutant promoters. Meanwhile, the transformation efficiency of the WB600 strain was 768 cfu/ μ g. Thus, the transformants were predicted to cover more than 99% of the possible mutants. Together with the high-throughput screening by flow cytometry, our methodological approach could thoroughly explore signature elements with higher efficiency, as compared to the traditional methods of studying promoter activity by random mutagenesis within the whole promoter sequence.

Thus it can be used to systematically determine the promoter sequence features which were influenced transcriptional activity in *B. subtilis*.

Previous studies have demonstrated that the spacer region between the – 35 and the – 10 box within the bacterial promoter is critical to regulate the expression levels of heterologous genes (Guiziou et al. 2016; Han et al. 2017; Jensen and Hammer 1998). Our work demonstrated that the GC content in the sequence flanking the – 35 box in the T2 locus enhanced the promoter activity, while the C content in the T3 locus reduced the promoter activity. However, the rationale for this influence in transcriptional activities is obscure. As the -35 element is responsible for

recruiting RNA polymerase (Hawley and McClure 1983), the GC content in the four nucleotides flanking the –35 box may impact RNA polymerase recruiting. This could involve higher-level regulatory effects like DNA looping (Cournac and Plumbridge 2013).

A promoter with a single characteristic is unlikely to satisfy each exogenous gene since, for example, strong overexpression is not always optimal for every given gene. We can see that, the strong promoters $P_{y1b-T2G}$ and $P_{y1b-T2C}$ weren't suitable for the expression of catalase KatA. As the catalase KatA in *B. subtilis* is an extracellular enzyme without known signal peptide (Naclerio et al. 1995), it may be secreted via non-classical secretion pathways. However, the mechanisms of non-classical secretion are unidentified. We speculated that the catalase KatA expressed by the strong promoter could not be folded correctly in time, which affected the secretion, resulting in the decreased extracellular expression.

The T2 and T3 libraries included mutant promoters that induced variable transcriptional levels, ranging from high to low (Supplementary Fig. 4). The randomization of the four nucleotides flanking the –35 box gave rise to dynamic expression that covered a wide range of levels, so every promoter with differential activity in the T2 and T3 libraries could be potentially used to express genes that require the corresponding transcriptional activity. Thus, the T2 and T3 libraries have great application potentials for enzyme production, metabolic engineering, and synthetic biology.

In conclusion, our thorough characterization of the P_{y1b} sequence provides further information for the potential application of this promoter, and contributes to the enrichment of the *B. subtilis* promoter library. Particularly, the T2 and T3 libraries cover a wide range of promoters with dynamic activities which have great potential to be applied for highly effective tuning of gene expression and regulation of complex metabolic pathways.

Acknowledgements This work was supported by the National Natural Science Foundation of China (NSFC, Grant no. 31500064).

Supplementary information Supplementary Table 1—Plasmids and strains used in this study.

Supplementary Table 2—Primers used in this study.

Supplementary Fig. 1—Cell sorting of *B. subtilis* mutant strains from the T2 and T3 randomly mutated promoter libraries.

Supplementary Fig. 2—The relative fluorescent intensity of EGFP driven by selected promoter variants of P_{y1b} from the T2 and T3 libraries.

Supplementary Fig. 3—Sequences of the P_{y1b} mutants collected from the T2 and T3 library.

Supplementary Fig. 4—Distribution of the expression levels of EGFP driven by promoter variants from the T2 and T3 libraries.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Bhavsar AP, Zhao X, Brown ED (2001) Development and characterization of a xylose-dependent system for expression of cloned genes in *Bacillus subtilis*: conditional complementation of a teichoic acid mutant. *Appl Environ Microbiol* 67:403–410. <https://doi.org/10.1128/AEM.67.1.403-410.2001>
- Blazcek J, Alper HS (2013) Promoter engineering: recent advances in controlling transcription at the most fundamental level. *Biotechnol J* 8:46–58. <https://doi.org/10.1002/biot.201200120>
- Cheng J, Guan C, Cui W, Zhou L, Liu Z, Li W, Zhou Z (2016) Enhancement of a high efficient autoinducible expression system in *Bacillus subtilis* by promoter engineering. *Protein Expr Purif* 127:81–87. <https://doi.org/10.1016/j.pep.2016.07.008>
- Cournac A, Plumbridge J (2013) DNA looping in prokaryotes: experimental and theoretical approaches. *J Bacteriol* 195:1109–1119. <https://doi.org/10.1128/JB.02038-12>
- De Mey M, Maertens J, Lequeux GJ, Soetaert WK, Vandamme EJ (2007) Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering. *BMC Biotechnol* 7:34. <https://doi.org/10.1186/1472-6750-7-34>
- Goldblith SA, Proctor BE (1950) Photometric determination of catalase activity. *J Biol Chem* 187:705–709
- Guan C, Cui W, Cheng J, Zhou L, Liu Z, Zhou Z (2016) Development of an efficient autoinducible expression system by promoter engineering in *Bacillus subtilis*.

- Microb Cell Fact 15:66. <https://doi.org/10.1186/s12934-016-0464-0>
- Guiziou S, Sauveplane V, Chang HJ, Clerte C, Declerck N, Jules M, Bonnet J (2016) A part toolbox to tune genetic expression in *Bacillus subtilis*. Nucleic Acids Res 44:7495–7508. <https://doi.org/10.1093/nar/gkw624>
- Han LC et al (2017) Fabrication and characterization of a robust and strong bacterial promoter from a semi-rationally engineered promoter library in *Bacillus subtilis*. Process Biochem 61:56–62. <https://doi.org/10.1016/j.procbio.2017.06.024>
- Hawley DK, McClure WR (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res 11:2237–2255
- Jensen PR, Hammer K (1998) The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. Appl Environ Microbiol 64:82–87
- Lee SJ, Pan JG, Park SH, Choi SK (2010) Development of a stationary phase-specific autoinducible expression system in *Bacillus subtilis*. J Biotechnol 149:16–20. <https://doi.org/10.1016/j.jbiotec.2010.06.021>
- Ming YM, Wei ZW, Lin CY, Sheng GY (2010) Development of a *Bacillus subtilis* expression system using the improved P_{glv} promoter. Microb Cell Fact 9:55. <https://doi.org/10.1186/1475-2859-9-55>
- Naclerio G, Baccigalupi L, Caruso C, De Felice M, Ricca E (1995) *Bacillus subtilis* vegetative catalase is an extracellular enzyme. Appl Environ Microbiol 61:4471–4473
- Phan TT, Nguyen HD, Schumann W (2012) Development of a strong intracellular expression system for *Bacillus subtilis* by optimizing promoter elements. J Biotechnol 157:167–172. <https://doi.org/10.1016/j.jbiotec.2011.10.006>
- Phan TT, Tran LT, Schumann W, Nguyen HD (2015) Development of Pgrac100-based expression vectors allowing high protein production levels in *Bacillus subtilis* and relatively low basal expression in *Escherichia coli*. Microb Cell Fact 14:72. <https://doi.org/10.1186/s12934-015-0255-z>
- Spizizen J (1958) Transformation of biochemically deficient strains of *Bacillus subtilis* by deoxyribonucleate. Proc Natl Acad Sc USA 44:1072–1078. <https://doi.org/10.1073/pnas.44.10.1072>
- Steinmetz M, Le Coq D, Aymerich S, Gonzy-Treboul G, Gay P (1985) The DNA sequence of the gene for the secreted *Bacillus subtilis* enzyme levansucrase and its genetic control sites. Mol Gen Genet 200:220–228
- van Dijk JM, Hecker M (2013) *Bacillus subtilis*: from soil bacterium to super-secreting cell factory. Microb Cell Fact 12:3. <https://doi.org/10.1186/1475-2859-12-3>
- You C, Percival Zhang YH (2012) Easy preparation of a large-size random gene mutagenesis library in *Escherichia coli*. Anal Biochem 428:7–12. <https://doi.org/10.1016/j.ab.2012.05.022>
- Yu X et al (2015) Identification of a highly efficient stationary phase promoter in *Bacillus subtilis*. Sci Rep 5:18405. <https://doi.org/10.1038/srep18405>
- Zhang XZ, Cui ZL, Hong Q, Li SP (2005) High-level expression and secretion of methyl parathion hydrolase in *Bacillus subtilis* WB800. Appl Environ Microbiol 71:4101–4103. <https://doi.org/10.1128/AEM.71.7.4101-4103.2005>
- Zhou C, Ye B, Cheng S, Zhao L, Liu Y, Jiang J, Yan X (2019) Promoter engineering enables overproduction of foreign proteins from a single copy expression cassette in *Bacillus subtilis*. Microb Cell Fact 18:111. <https://doi.org/10.1186/s12934-019-1159-0>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.