**RESEARCH**                                                                **Open Access**

# Prediction and characterization of promoters and ribosomal binding sites of *Zymomonas mobilis* in system biology era

Yongfu Yang[1†], Wei Shen[1†], Ju Huang[1], Runxia Li[1], Yubei Xiao[1], Hui Wei[2], Yat-Chen Chou[3], Min Zhang[2], Michael E. Himmel[2], Shouwen Chen[1], Li Yi[1], Lixin Ma[1] and Shihui Yang[1*]

## Abstract

**Background:** *Zymomonas mobilis* is a model bacterial ethanologen with many systems biology studies reported. Besides lignocellulosic ethanol production, *Z. mobilis* has been developed as a platform for biochemical production through metabolic engineering. However, identification and rigorous understanding of the genetic origins of cellular function, especially those based in non-coding region of DNA, such as promoters and ribosomal binding sites (RBSs), are still in its infancy. This knowledge is crucial for the effective application of *Z. mobilis* to new industrial applications of biotechnology for fuels and chemicals production.

**Results:** In this study, we explored the possibility to systematically predict the strength of promoters based on systems biology datasets. The promoter strength was clustered based on the expression values of downstream genes (or proteins) from systems biology studies including microarray, RNA-Seq and proteomics. Candidate promoters with different strengths were selected for further characterization, which include 19 strong, nine medium, and ten weak ones. A dual reporter-gene system was developed which included appropriate reporter genes. These are the *opm-Cherry* reporter gene driven by the constitutive P*lacUV5* promoter for calibration, and *EGFP* reporter gene driven by candidate promoters for quantification. This dual reporter-gene system was confirmed using the inducible promoter, P*tet*, which was used to determine the strength of these predicted promoters with different strengths. In addition, the dual reporter-gene system was applied to determine four synthetic RBSs with different translation initiation rates based on the prediction from bioinformatics server RBS calculator. Our results showed that the correlations between the prediction and experimental results for the promoter and RBS strength are relatively high, with $R^2$ values more than 0.7 and 0.9, respectively.

**Conclusions:** This study not only identified and characterized 38 promoters and four RBSs with different strengths for future metabolic engineering in *Z. mobilis*, but also established a flow cytometry-based dual reporter-gene system to characterize genetic elements including, but not limited to the promoters and RBSs studied in this work. This study also suggested the feasibility of predicting and selecting candidate genetic elements based on omics datasets and bioinformatics tools. Moreover, the dual reporter-gene system developed in this study can be utilized to characterize other genetic elements of *Z. mobilis*, which can also be applied to other microorganisms.

*Correspondence: Shihui.Yang@hubu.edu.cn
†Yongfu Yang and Wei Shen contributed equally to this work
[1] State Key Laboratory of Biocatalysis and Enzyme Engineering, Hubei Collaborative Innovation Center for Green Transformation of Bio-resources, Environmental Microbial Technology Center of Hubei Province, and School of Life Sciences, Hubei University, Wuhan 430062, China
Full list of author information is available at the end of the article

Yang *et al. Biotechnol Biofuels*      (2019) 12:52

Page 2 of 13

## Background

With the increasing consumption of fossil energy, significant efforts have focused on the development of sustainable alternative renewable energy needed to meet the energy demands for economic development and environmental protection. Correspondingly, biofuels produced through biochemical conversion from biomass-based feedstock by microorganisms have attracted significant attention. *Z. mobilis* is a natural ethanologenic bacterium with many desirable characteristics necessary to produce lignocellulosic biofuels and their intermediates through metabolic engineering, including ethanol and 2,3-butanediol [1–4].

To meet the needs of metabolic engineering and synthetic biology approaches, *Z. mobilis* genetic elements from coding regions (genes) and non-coding regions [e.g., promoters, ribosomal binding site (RBS), untranslated region (UTR), and terminators] have been broadly investigated [5]. Different from those in the coding region, genetic elements from the non-coding region can affect gene expression at the transcriptional or translational levels; as well as modulate their activity in response to environmental conditions [6–9]. Although some noncoding sequence elements of *Z. mobilis* were discovered [7, 10], there are no systematic and efficient approaches to identify and quantify elements that were already discovered. Many genetic elements remain undiscovered today.

With the development and deployment of technology, such as next-generation sequencing (NGS) and mass spectrometry, many systems biology studies were carried out with enormous omics data accumulated [11–13], including studies of *Z. mobilis* [13–27]. These systems biology datasets are vast, and thus contain information useful for deep mining and modeling [28, 29]. For example, genetic elements such as promoters can be sorted out by multiple omics data analysis [30]. However, promoters of *Z. mobilis* have not been systematically characterized, although the accurate information of genome and plasmid sequence; as well as genome annotation of *Z. mobilis* are available [14, 31, 32].

Traditionally, native promoters and RBSs are often discovered as a result of random genomic digestion, which is further facilitated by genome sequencing and annotation [6]. These genetic elements of different strengths are often required in metabolic engineering practices. For example, strong promoters were usually used to overexpress target genes to increase the titer of heterologous products [33]. Up to now, only a small number of strong promoters, such as P*gap*, P*pdc*, and P*eno*, have been investigated and used to construct metabolic pathways into wild-type strains suitable for efficient xylose utilization or enhanced inhibitor tolerance in *Z. mobilis* [34–43].

The application of strong promoters on metabolic engineering could cause metabolic burden on cellular growth, and thus reduce the titer, yield, or productivity. To address this issue, inducible promoters, such as P*tet*, were recently used to direct metabolic flux for balanced cellular growth and production [44]. However, the strengths of these promoters have not been investigated quantitatively, making it difficult to meet the needs of metabolic engineering and synthetic biology for genetic elements with different strengths [45]. Moreover, no RBS sequences have been specifically studied for *Z. mobilis*.

Some in vitro methods, such as electrophoresis mobility shift assay (EMSA) [46] and atomic force microscope (AFM) [47], have been developed to characterize promoter activity, but are not widely used due to their low accuracy in characterization of genetic elements quantitatively in vivo. The classical strategy to quantify the promoter activity in vivo is to measure the expression of downstream reporter genes driven by the candidate promoter [48]. Reporter genes, such as β-glucuronidase, β-galactosidase and fluorescent proteins, are commonly used for this purpose.

Fluorescent reporter proteins, which were established in 1996 [49] and developed later [50, 51], are preferred to be used with their unique advantages, which include being relatively small and nontoxic [52]. However, the expression of fluorescent reporter genes is affected by many internal and external factors, including mRNA degradation, translation, protein folding; as well as plasmid copy numbers [53], which then influence the promoter characterization in vivo [54]. These combined factors make it difficult to accurately measure the promoter activity in vivo, especially when using a single report gene in individual experiment without an intrinsic control [55, 56]. To reduce these influences, a dual reporter-gene system was developed and used for promoter characterization in different species, such as *E. coli*, *Clostridium* sp., and virus [55, 57, 58].

In this study, a dual reporter-gene system was developed which contains: (1) the *opmCherry* reporter gene driven by the constitutive *PlacUV5* promoter for calibration, and (2) the *EGFP* reporter gene for quantifying the

Yang *et al. Biotechnol Biofuels*    (2019) 12:52

Page 3 of 13

candidate genetic elements located upstream of *EGFP*. A series of promoters with different strengths were predicted based on previous systems biology data and four synthetic RBSs with different translation initiation rates were predicted and synthesized. These promoters and RBSs were then experimentally characterized using the dual reporter-gene system to explore the relationship between the prediction and experimental data.

## Results

### Establishment of the dual reporter-gene system

To develop a reliable dual reporter-gene system, fluorescent protein pairs should meet some requirements. Firstly, the pair should have distinguishable excitation and emission spectrum to avoid interplay between channels. Secondly, similar mature time rates are preferred, since different mature time rates will affect the simultaneous measurement. Thirdly, bright fluorescence intensity is preferred to help detect the signal intensity of relatively weak promoters. Based on these criteria, seven single fluorescent proteins, including EGFP, mCherry, RFP and CFP; as well as codon-optimized EGFP (opEGFP), mCherry (opmCherry), and CFP (opCFP), were synthesized and characterized under the control of a constitutive promoter P*lacUV5* using the shuttle vector pEZ15Asp [44] (Fig. 1a). The emission and excitation wavelength of CFP were 433 nm and 475 nm, which was failed to detect by our detector and not included in this work. Flow cytometry results indicated that EGFP and opmCherry produced the brightest fluorescence signal among the seven fluorescent proteins tested (Fig. 1b).

Since opmCherry (excitation maximum 587 nm, emission maximum 610 nm) and EGFP (excitation maximum 488 nm, emission maximum 507 nm) are spectrally distinguishable with good photostability [59, 60] and relatively comparable faster mature rates (40 min and 25 min, respectively) [59, 61] than that of RFP (100 min) [62], they were then selected to construct the dual reporter-gene system with *opmCherry* driven by the constitutive promoter P*lavUV5* as the intrinsic control and *EGFP* under the control of the targeted genetic element, such as the candidate promoter or RBS (Fig. 2a).

To confirm whether the dual reporter-gene system works, we first tested the system using an inducible promoter P*tet* to simulate the strengths of different promoters and determined the appropriate tetracycline inducer concentration. Our results indicated that the cellular growth decreased when tetracycline concentration was greater than 2 μg/mL, whereas the tetracycline concentration within 0–1 μg/mL did not affect cellular growth significantly (Additional file 2: Figure S1).

The correlation between the concentration of tetracycline inducer and the promoter strength was then
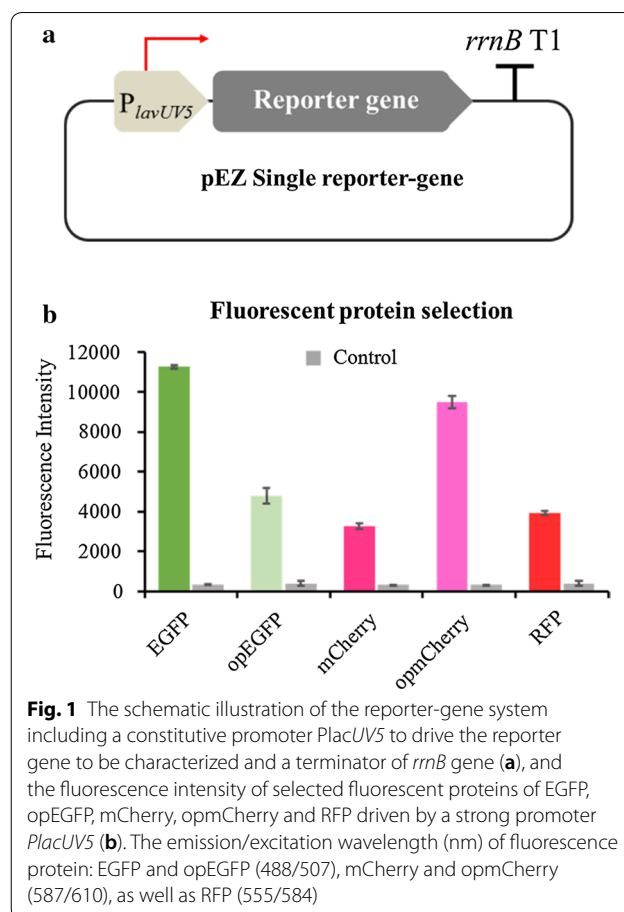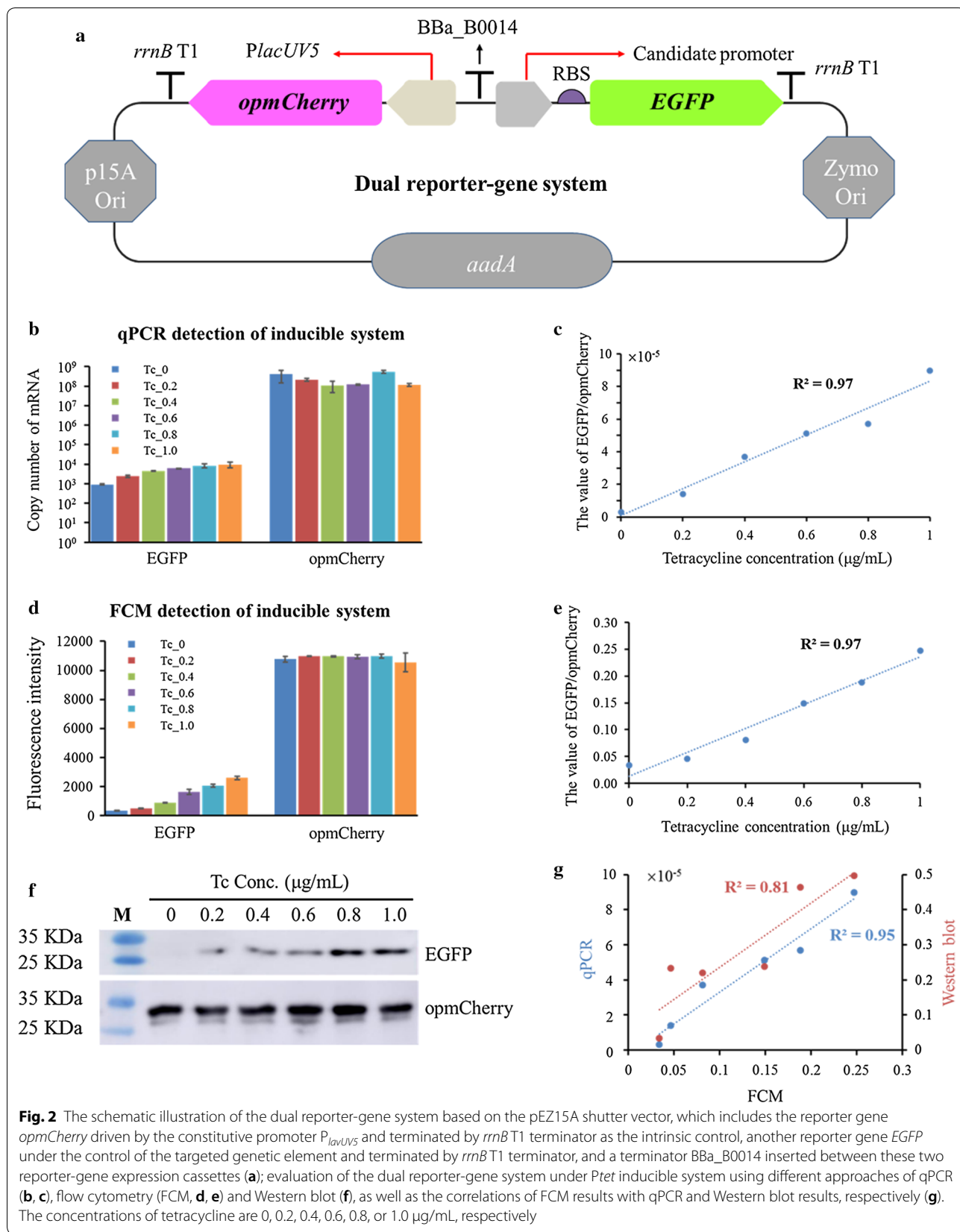


**Fig. 1** The schematic illustration of the reporter-gene system including a constitutive promoter PlacUV5 to drive the reporter gene to be characterized and a terminator of *rrnB* gene (**a**), and the fluorescence intensity of selected fluorescent proteins of EGFP, opEGFP, mCherry, opmCherry and RFP driven by a strong promoter *PlacUV5* (**b**). The emission/excitation wavelength (nm) of fluorescence protein: EGFP and opEGFP (488/507), mCherry and opmCherry (587/610), as well as RFP (555/584)

investigated. The ratio of EGFP fluorescence intensity for tested promoter versus the opmCherry calibration fluorescence intensity was used to represent the promoter strength. qPCR, Western blot and flow cytometry analyses were carried out to evaluate the target gene expression at the transcriptional and translational levels. The results indicated that the expression of *opmCherry* is at a relatively constant level, whereas the expression of *EGFP* was enhanced correspondingly to the increase of the tetracycline inducer concentration within 0–1 μg/mL (Fig. 2b, d, f). In addition, the value of EGFP/opmCherry, a normalized ratio to eliminate the internal and external noises [55], has a linear correlation with the gradient tetracycline concentrations (Fig. 2c, e).

The correlations between the results of flow cytometry and other technologies, such as qPCR or Western blot, were also relatively high (Fig. 2g), which supports the possibility of utilizing the high-throughput quantitative approach of flow cytometry to quantify different candidate genetic elements.

**Fig. 2** The schematic illustration of the dual reporter-gene system based on the pEZ15A shutter vector, which includes the reporter gene *opmCherry* driven by the constitutive promoter P*lavUV5* and terminated by *rrnB* T1 terminator as the intrinsic control, another reporter gene *EGFP* under the control of the targeted genetic element and terminated by *rrnB* T1 terminator, and a terminator BBa_B0014 inserted between these two reporter-gene expression cassettes (**a**); evaluation of the dual reporter-gene system under P*tet* inducible system using different approaches of qPCR (**b**, **c**), flow cytometry (FCM, **d**, **e**) and Western blot (**f**), as well as the correlations of FCM results with qPCR and Western blot results, respectively (**g**). The concentrations of tetracycline are 0, 0.2, 0.4, 0.6, 0.8, or 1.0 µg/mL, respectively

Yang *et al. Biotechnol Biofuels*      (2019) 12:52

Page 5 of 13

## Identification of candidate promoters with different strengths using omics datasets

Systems biology datasets from 109 microarrays, eight RNA-seq, and four proteomic experiments of previous studies were used to identify potential promoters with different strengths. The strength of upstream promoters was presented as the normalized average value of downstream gene expression in normal conditions of each omics dataset. To select promoters with different strengths, we sorted the normalized value and defined strength of promoters arbitrarily based on their quantile values: in each omics dataset, those sorted in the top 90% were defined as strong promoters, those sorted in the last 10% were defined as weak promoters, and those sorted within 40–60% were defined as medium ones.

Then, hierarchical cluster analysis cross different omics data was used to identify genes with consistent expression patterns (strong, medium, or weak), as shown in Venn diagram (Fig. 3). Six other weak promoters were also selected for a broad representation of weak promoters. Considering the existence of operon may affect the determination of the location for its upstream promoter, we used DOOR2 to predict the operons [63, 64]. A total of 38 candidate promoters with diverse strengths from strong to weak were then selected, which include 19 strong, nine medium, and ten weak ones (Table 1). In this study, the entire intergenic region of those genes was extracted as candidate promoters. These candidate promoters are listed in Table 1, and the information of promoter sequence and length are listed in supplementary materials (Additional file 1: Table S1).

## Characterization of candidate promoters quantitatively using flow cytometry

Selected candidate promoters were assembled into the upstream of EGFP reporter gene in the dual reporter-gene system using a modified Gibson assembly approach [65]. The relative strength of these promoters was quantified in exponential and stationary phase by flow cytometry (Table 1). Each sample was operated in triplicates with at least two independent experiments conducted by two researchers in the lab at different times.

The results showed that P*gap*, P*pdc*, and P*eno* have high EGFP/opmCherry ratios as determined by our dual reporter-gene system, which actually is consistent with our omics data-driven predictions and with previous studies that showed P*gap*, P*pdc*, P*eno* and P*adhB* are strong promoters in *Z. mobilis* [42, 43, 66]. The P*gap*, encoding the glyceraldehyde-3-phosphate dehydrogenase (ZMO0177) in glycolysis pathway, is the strongest promotor in *Z. mobilis* characterized in this study (Table 1). Other promoters driving genes in glycolysis pathway also have high EGFP/opmCherry ratios (Table 1).

Our results demonstrated that different flow cytometry profiles were identified and can be classified into strong, medium, and weak categories (Fig. 4a). This result suggested that the flow cytometry-based dual reporter-gene system can clearly distinguish different strengths of promoters based on their profiles. In addition, the correlations between omics data prediction and EGFP/opm-Cherry ratios are relatively high, with an $R^2$ more than 0.7 for RNA-Seq and proteomics datasets (Fig. 4b).
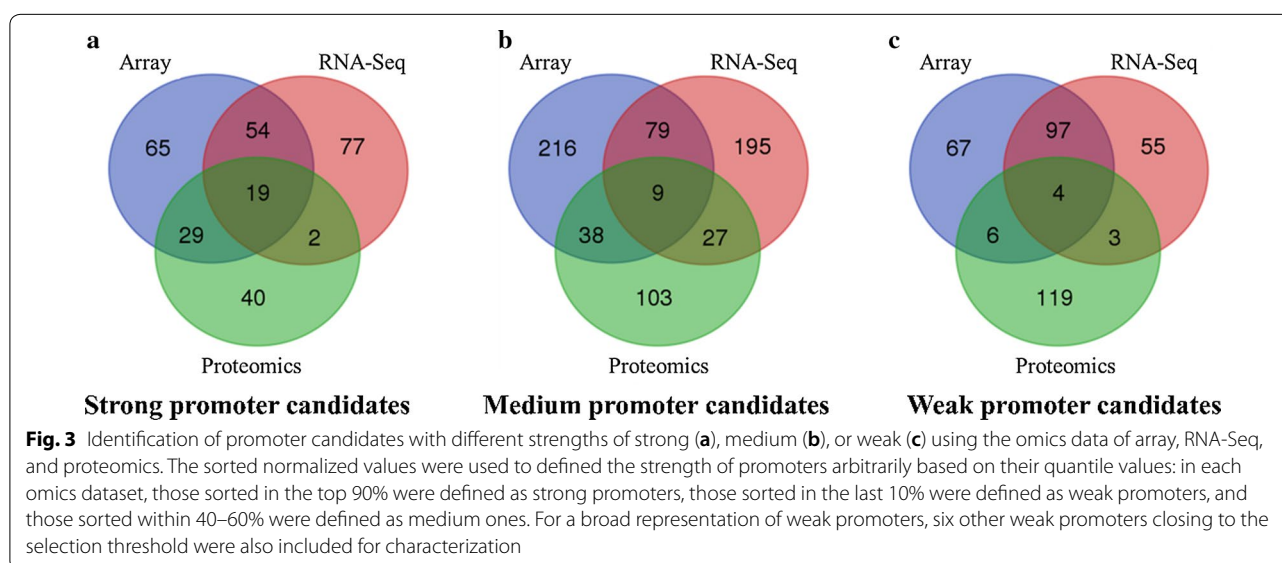


**Fig. 3** Identification of promoter candidates with different strengths of strong (**a**), medium (**b**), or weak (**c**) using the omics data of array, RNA-Seq, and proteomics. The sorted normalized values were used to defined the strength of promoters arbitrarily based on their quantile values: in each omics dataset, those sorted in the top 90% were defined as strong promoters, those sorted in the last 10% were defined as weak promoters, and those sorted within 40–60% were defined as medium ones. For a broad representation of weak promoters, six other weak promoters closing to the selection threshold were also included for characterization

Yang *et al. Biotechnol Biofuels*     (2019) 12:52

Page 6 of 13

**Table 1 Candidate promoters with different strengths selected based on omics datasets**

| Gene ID | Gene name | Operon | Gene function | Array | RNA-Seq | Proteomics | EGFP/opmCherry (log phase) | EGFP/opmCherry (sta phase) |
|---|---|---|---|---|---|---|---|---|
| Candidate promoters with strong strength | | | | | | | | |
| ZMO0177 | *gap* | | Glyceraldehyde-3-phosphate dehydrogenase, type I | 15.06 | 11.2 | 9.23 | 0.38 ± 0.0202 | 0.47 ± 0.0191 |
| ZMO1360 | *pdc* | | Thiamine pyrophosphate protein TPP binding domain-containing protein | 14.52 | 11.78 | 8.59 | 0.24 ± 0.0220 | 0.22 ± 0.0086 |
| ZMO0516 | *Tuf* | ND | Elongation factor Tu | 15.33 | 11.58 | 8.08 | 0.24 ± 0.0162 | 0.30 ± 0.0148 |
| ZMO1608 | *eno* | | Phosphopyruvate hydratase | 15.22 | 13.2 | 8.97 | 0.23 ± 0.0145 | 0.34 ± 0.0093 |
| ZMO0997 | *eda* | Yes | 2-Dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate aldolase | 14.97 | 14.6 | 8.46 | 0.18 ± 0.0178 | 0.15 ± 0.0167 |
| ZMO0367 | *zwf* | | Glucose-6-phosphate 1-dehydrogenase | 14.92 | 11.19 | 7.38 | 0.16 ± 0.0057 | 0.14 ± 0.0124 |
| ZMO1719 | *frk* | | ROK family protein | 15.06 | 12.2 | 6.85 | 0.12 ± 0.0038 | 0.17 ± 0.0098 |
| ZMO1609 | | | Hypothetical protein | 15.26 | 12.77 | 5.78 | 0.12 ± 0.0061 | 0.14 ± 0.0038 |
| ZMO0689 | *gfo* | | Oxidoreductase domain-containing protein | 14.49 | 11.68 | 6.38 | 0.10 ± 0.0075 | 0.13 ± 0.0130 |
| ZMO1721 | *gloA3* | ND | Glyoxalase/bleomycin resistance protein/dioxygenase | 14.35 | 12.66 | 5.97 | 0.09 ± 0.0044 | 0.11 ± 0.0022 |
| ZMO0514 | *rpsG* | Yes | 30S ribosomal protein S7 | 15.31 | 10.79 | 5.54 | 0.07 ± 0.0062 | 0.07 ± 0.0015 |
| ZMO0515 | | Yes | Elongation factor G | 15.07 | 11.14 | 5.83 | 0.07 ± 0.0062 | 0.07 ± 0.0015 |
| ZMO1596 | *adhB* | | Iron-containing alcohol dehydrogenase | 15.32 | 10.98 | 7.07 | 0.07 ± 0.0084 | 0.07 ± 0.0030 |
| ZMO1141 | *ilvC* | Yes | Ketol-acid reductoisomerase | 15.41 | 12.39 | 6.76 | 0.05 ± 0.0012 | 0.04 ± 0.0006 |
| ZMO0241 | *atpD* | Yes | F0F1 ATP synthase subunit beta | 15.09 | 11.73 | 7.77 | 0.05 ± 0.0030 | 0.04 ± 0.0029 |
| ZMO0244 | | | Histone family protein DNA-binding protein | 14.79 | 12.98 | 6.1 | 0.04 ± 0.0013 | 0.04 ± 0.0007 |
| Po1721 | | | | | | | 0.04 ± 0.0020 | 0.03 ± 0.0007 |
| ZMO0493 | *glnA* | Yes | Glutamine synthetase, type I | 14.53 | 10.02 | 6.17 | 0.03 ± 0.0023 | 0.03 ± 0.0005 |
| ZMO1779 | | Yes | Antibiotic biosynthesis monooxygenase | 15.08 | 11.24 | 7.52 | 0.02 ± 0.0005 | 0.02 ± 0.0008 |
| Candidate promoters with medium strength | | | | | | | | |
| ZMO1351 | *clcD1* | Yes | Carboxymethylenebutenolidase | 12.93 | 6.81 | 3.14 | 0.14 ± 0.0064 | 0.16 ± 0.0036 |
| ZMO0056 | *glmS* | | Glucosamine–fructose-6-phosphate aminotransferase | 12.93 | 6.91 | 2.45 | 0.12 ± 0.0021 | 0.11 ± 0.0024 |
| ZMO0559 | | | Hypothetical protein | 12.68 | 6.75 | 3.1 | 0.11 ± 0.0064 | 0.08 ± 0.0039 |
| ZMO1385 | | | Toxic anion resistance family protein | 12.83 | 6.93 | 2.55 | 0.06 ± 0.0012 | 0.05 ± 0.0018 |
| ZMO0127 | | Yes | S1/P1 nuclease | 12.84 | 7.11 | 3.13 | 0.05 ± 0.0015 | 0.05 ± 0.0014 |
| ZMO1100 | | Yes | Nucleotidyl transferase | 12.58 | 7.18 | 2.82 | 0.05 ± 0.0012 | 0.05 ± 0.0023 |
| ZMO1392 | | | Hypothetical protein | 12.46 | 7.43 | 2.45 | 0.04 ± 0.0012 | 0.04 ± 0.0012 |
| ZMO0326 | | | 7-cyano-7-deazaguanine reductase | 12.65 | 7.41 | 2.7 | 0.03 ± 0.0018 | 0.04 ± 0.0012 |
| ZMO0570 | *prmA* | | Ribosomal L11 methyltransferase | 12.38 | 7.32 | 2.45 | 0.03 ± 0.0030 | 0.04 ± 0.0013 |
| Candidate promoters with weak strength | | | | | | | | |
| ZMO1231 | *recJ* | | Single-stranded-DNA-specific exonuclease RecJ | 11.03 | 5.7 | 0.07 | 0.08 ± 0.0023 | 0.07 ± 0.0029 |
| ZMO1980 | *gidB* | Yes | Methyltransferase GidB | 10.59 | 5.27 | 0.07 | 0.05 ± 0.0025 | 0.05 ± 0.0029 |
| ZMO1484 | | | UvrD/REP helicase | 10.93 | 5.4 | 0.07 | 0.05 ± 0.0013 | 0.05 ± 0.0026 |
| ZMO0145 | | Yes | Peptidase M28 | 11.37 | 4.98 | 0.07 | 0.04 ± 0.0018 | 0.04 ± 0.0006 |
| ZMO0101 | | | NAD-dependent epimerase/dehydratase | 10.61 | 5.05 | 0.07 | 0.04 ± 0.0013 | 0.04 ± 0.0014 |
| ZMO1194 | *dprA* | Yes | DNA protecting protein DprA | 10.7 | 4.77 | 0.07 | 0.04 ± 0.0011 | 0.03 ± 0.0006 |
| ZMO1644 | | | DEAD/DEAH box helicase domain-containing protein | 10.61 | 4.98 | 0.07 | 0.03 ± 0.0013 | 0.03 ± 0.0007 |
| ZMO1582 | | | Uracil-DNA glycosylase superfamily protein | 10.33 | 4.63 | 0.07 | 0.03 ± 0.0004 | 0.03 ± 0.0003 |
| ZMO0005 | *cysD* | Yes | Sulfate adenylyltransferase subunit 2 | 11.5 | 5.28 | 0.07 | 0.03 ± 0.0012 | 0.03 ± 0.0009 |
| ZMO0300 | *xseA* | | Exodeoxyribonuclease VII large subunit | 11.6 | 4.71 | 0.07 | 0.03 ± 0.0010 | 0.03 ± 0.0013 |

The operon information is predicted by DOOR2 prediction server. The values of array, RNA-Seq and proteomics of each gene are the average $\log_2$-transformed values under different growth conditions

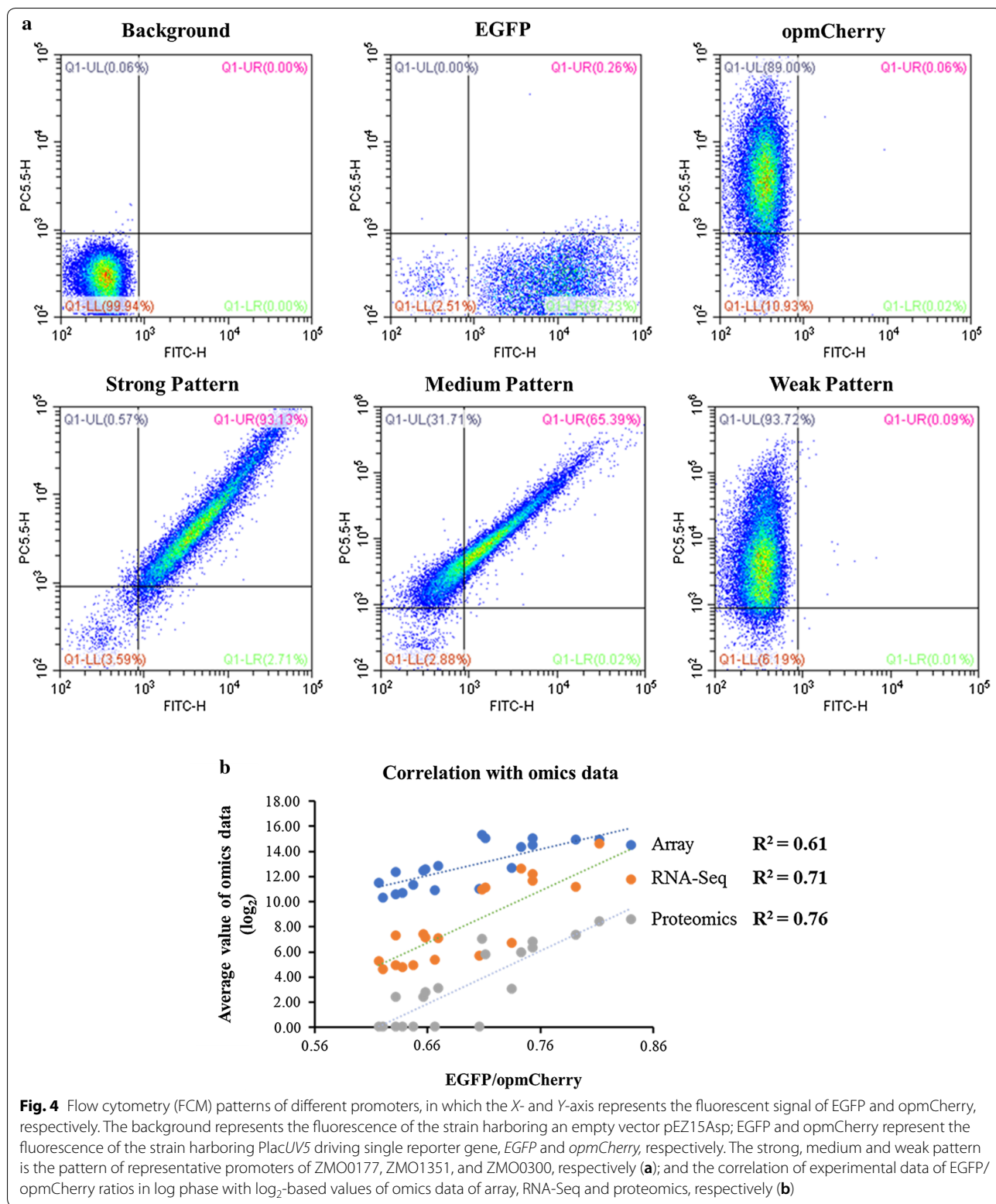*ND* non-determined, *Log phase* exponential phase, *Sta phase* stationary phase

Yang *et al. Biotechnol Biofuels*     (2019) 12:52

Page 7 of 13



**Fig. 4** Flow cytometry (FCM) patterns of different promoters, in which the *X*- and *Y*-axis represents the fluorescent signal of EGFP and opmCherry, respectively. The background represents the fluorescence of the strain harboring an empty vector pEZ15Asp; EGFP and opmCherry represent the fluorescence of the strain harboring Plac*UV5* driving single reporter gene, *EGFP* and *opmCherry,* respectively. The strong, medium and weak pattern is the pattern of representative promoters of ZMO0177, ZMO1351, and ZMO0300, respectively (**a**); and the correlation of experimental data of EGFP/opmCherry ratios in log phase with $\log_2$-based values of omics data of array, RNA-Seq and proteomics, respectively (**b**)

Yang *et al. Biotechnol Biofuels*　　(2019) 12:52

Page 8 of 13

**Table 2 The nucleotide sequences of synthetic RBS with different translation initiation rates based on the prediction using RBS calculator (https://salislab.net/software/forward)**

| Name | Strength | Predicted RBS sequence (5'–3') |
|---|---|---|
| ZM4-Ptet-GFP-10 | 10 | CCATAATCTAGAGAAAGTAAGCAC |
| ZM4-Ptet-GFP-1000 | 1000 | AGGCTAAGAACTAACGGAGAGGTAAAT |
| ZM4-Ptet-GFP-10000 | 10,000 | ATCACAGGGTCTAGAAGGAGGTCGAA |
| ZM4-Ptet-GFP-Max | 15,000 | GAGCGAGAAGGAGGTAAAGT |

### Quantification of predicted RBSs with different translation initiation rates

The dual reporter-gene system developed in this study was further explored to measure the RBSs with different translation initiation rates. The promoter used in this experiment was the inducible P*tet*; however, the RBS sequence was replaced by the synthesized RBS sequences with different translation initiation rates predicted by RBS calculator V2.0 (https://salislab.net/software/), which is based on the 16S rRNA sequence of *Z. mobilis*. From weak to strong translation initiation rates, four synthetic RBSs (Table 2) were selected for quantification in *Z. mobilis* using the dual reporter-gene system (Fig. 2).

The results showed that within a fixed strength of translation initiation rate, the relative strength of RBS was enhanced with the increase of tetracycline inducer concentration gradient (Fig. 5a, Additional file 2: Figure S2). In addition, the predicted RBS strengths and tetracycline concentrations also had relatively high correlations with an $R^2$ value of more than 0.93 (Fig. 5b). These results demonstrated that the dual reporter-gene system performed well in characterizing the strength of RBS, which suggests that the dual reporter-gene system can be used to characterize the non-coding sequence elements of RBSs.

## Discussions

In this study, a flow cytometry-based dual reporter-gene system was developed (Fig. 1), which was confirmed with an inducible promoter Ptet using different approaches of qPCR, flow cytometry, and Western blot (Fig. 2), and then applied to quantify biological parts such as promoters and RBS predicted by bioinformatics approaches with good correlations (Figs. 3, 4, 5). The high correlation between different concentrations of the tetracycline inducer and EGFP/opmCherry ratios suggested the capability of this dual reporter-gene system to characterize the strengths of different genetic elements, and the relatively high correlations between predicted candidates and experimental results for promoters and RBSs also indicate that it is practicable to predict the strengths of genetic elements based on omics datasets
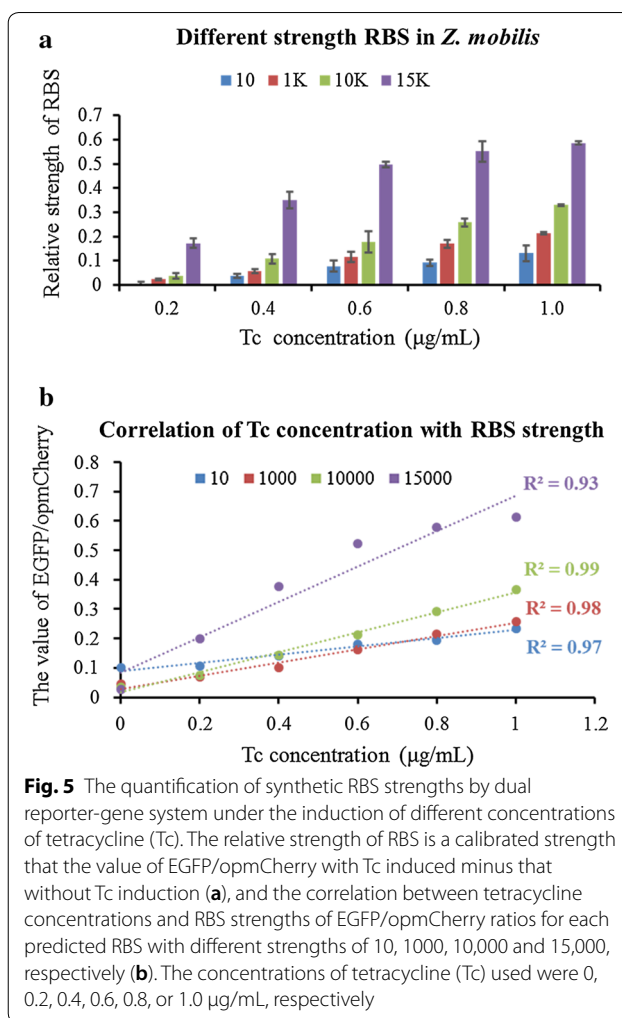


**Fig. 5** The quantification of synthetic RBS strengths by dual reporter-gene system under the induction of different concentrations of tetracycline (Tc). The relative strength of RBS is a calibrated strength that the value of EGFP/opmCherry with Tc induced minus that without Tc induction (**a**), and the correlation between tetracycline concentrations and RBS strengths of EGFP/opmCherry ratios for each predicted RBS with different strengths of 10, 1000, 10,000 and 15,000, respectively (**b**). The concentrations of tetracycline (Tc) used were 0, 0.2, 0.4, 0.6, 0.8, or 1.0 μg/mL, respectively

and bioinformatics tools, and therefore can be used to guide the selection of promoters with different strengths at the transcriptional level in other microbial systems.

Additionally, our result indicated that the codon-optimized EGFP actually had a decreased fluorescence, while the codon-optimized mCherry had enhanced fluorescence intensity (Fig. 1b). This system can, therefore, be applied to select and test reporter genes with desired fluorescence through protein engineering approaches such as codon optimization, rational design, and directed evolution.

Furthermore, this approach is straightforward with high reproducibility, and the fluorescence intensity of both EGFP and opmCherry as well as the corresponding correlation between results from different researchers in our group at different operation times is high with an $R^2$ value of 0.92 (Additional file 2: Figure S3). However, future work is still needed to further develop this dual reporter-gene system for broad applications such as the inclusion of short-life reporter genes and advanced

Yang *et al. Biotechnol Biofuels* (2019) 12:52

Page 9 of 13

imaging techniques to monitor the dynamic expression changes of diverse genetic elements including promoters and RBSs investigated in this work as well as other genetic elements to be explored such as UTRs, terminators, and sRNA.

## Conclusion

In this study, we conducted the systems biology data mining and bioinformatics analyses to predict the strengths of genetic elements, such as promoters and RBSs. We also developed a flow cytometry-based dual reporter-gene platform for *Z. mobilis* to measure the strengths of selected promoters and synthesized RBSs. We found a relatively high correlation between prediction and experimental results for the strengths of promoters and RBSs. The feasibility of identifying and characterizing predicted genetic elements, such as promoters and RBSs, demonstrated in this study suggests that other genetic elements from *Z. mobilis*, such as 5'UTRs and terminators, could also be quantitatively characterized by similar approaches. Furthermore, the high correlation between prediction and experimental results indicates that a similar strategy could be applied to other microorganisms leading to the identification of genetic elements using systems biology data and bioinformatics tools. Moreover, the seven native and synthesized fluorescent proteins and genetic elements including 38 promoters and four RBSs characterized in this study can be used for future metabolic engineering and synthetic biology practices in *Z. mobilis*.

## Methods

### Strains, vectors, and media

*Escherichia coli* XL10-Gold from Invitrogen (USA), was cultured in Lysogeny broth (LB, 10 g/L NaCl, 10 g/L tryptone, 5 g/L yeast extract). Medium was prepared according to the description in the Manual of Molecular Cloning [67]. Wild type *Z. mobilis* ZM4 was revived from frozen glycerol stocks in rich media with 5% glucose (RMG5: 50 g/L glucose, 10 g/L yeast extract, and 2 g/L $KH_2PO_4$) at 30 °C for 6–8 h without shaking. Shuttle vector pEZ15Asp includes origins of replication from both *E. coli* and *Z. mobilis* [44]. Strains and plasmids used in this study are listed in Table 3.

### Selection of candidate promoters

Public and in-house transcriptomic datasets from RNA-seq and microarray studies, as well as proteomic datasets were retrieved, collected, and hierarchically clustered separately using the JMP Genomics software 6.0 (SAS Inc., NC) with the conventional processing flows using default parameters [15, 16, 21, 23]. The common promoters showing strong, medium, or weak expression levels or abundance based on gene expression intensity or protein peptide hits abundance of downstream genes or proteins were selected using the Venn diagram generator at Venny 2.0.2 program (http://bioinfogp.cnb.csic.es/tools/venny/).

In this study, promoters were loosely defined as the region between the gene of interest and the preceding open reading frame (ORF). If genes were found to be transcribed as part of an operon, the operon promoter was selected. Genome-wide operon prediction in *Z. mobilis* is available at DOOR2 website server (http://csbl.bmb.uga.edu/DOOR/).

### Plasmid construction of candidate genetic elements

The vector pEZ15Asp containing an origin of replication with promoters for *E. coli* as well as *Z. mobilis* [44] was modified to construct the candidate genetic elements into the dual reporter-gene system. The reporter gene *opmCherry* driven by a constitutive promoter *LacUV5* was assembled to the suffix sequences of *Z. mobilis* replication promoter [68]. The reporter gene *EGFP* driven by tetracycline inducible promoter P*tet* was assembled to the prefix sequences of *E. coli* replication promoter.

For construction of candidate promoters into the dual reporter-gene system, the P*tet* was replaced by candidate promoters individually. All candidate promoters

**Table 3 Strains and plasmids used in this study**

| Strains and plasmids | Description | Source |
|---|---|---|
| Strains | | |
| *E. coli* XL10-Gold | Ultracompetent cells (tetracycline and chloramphenicol resistant) | Invitrogen |
| *Z. mobilis* ZM4 | *Z. mobilis* wild-type strain | Lab stock |
| Plasmids | | |
| pEZ15Asp | *P15A_ori, Zymo_Ori, P~tet~, Spe^R^* | [44] |
| pEZ15A-PlacUV5-EGFP | *P15A_ori, Zymo_Ori, P~lacUV5~::EGFP, Spe^R^* | This work |
| pEZ15A-PlacUV5-opmCherry | *P15A_ori, Zymo_Ori, P~lacUV5~::opmCherry, Spe^R^* | This work |
| pEZ-Dual | *P15A_ori, Zymo_Ori, P~tet~, Spe^R^, P~lacUV5~::opmCherry, P~tet~::EGFP* | This work |

Yang *et al. Biotechnol Biofuels*     (2019) 12:52

Page 10 of 13

were PCR amplified from *Z. mobilis* genomic DNA using primer sets listed in Additional file 1: Table S2. Each primer contains about 15–20 nucleotides overlapping region of vector without P*tet* promoter. Assembly was operated as described below. Recombinant vectors were selected by colony PCR with primer pairs of Pdual-F and Pdual-R. The schematic diagram of dual-reporter system is shown in Fig. 2a. To construct RBSs into the dual reporter-gene system, the candidate RBSs replace the original RBS in the P*tet* dual reporter-gene system by same operation as described above.

For plasmid construction, the protocol used was based on Gibson et al. [65]. Briefly, primers were designed to contain 15–20 nucleotides overlapping regions with adjacent DNA fragments. PCR products amplified by primer pairs were separated by gel electrophoresis, followed by gel purification, and subsequently quantified using NanoDrop 2000 (Thermo Fisher Scientific, USA). Fragments and vector were allocated in a molar ratio of 3:1, 0.5 U T5 exonuclease (NEB, USA), 0.5 μL buffer 4 (NEB, USA), and the final volume was set to 5 μL with ddH$_2$O.

All regents were mixed and reacted on the ice for 5 min; *E. coli* chemical competent cells were subsequently added. After incubation on the ice for 30 min, the mixture above was heat-shocked for 45 s at 42 °C, and then held on the ice for 2 min. Subsequently, 100 μL NZY was added into the mixture and incubated at least 1 h at 37 °C with shaking (250 rpm).

Cells were plated on LB agar plates containing spectinomycin, and recombinants were selected by colony PCR and confirmed by Sanger sequencing (Tsingke, China). The correct recombinant plasmids were transformed into *Z. mobilis* ZM4 competent cells, which were prepared as described previously [44], via electroporation (0.1-cm electrode gap, 1600 V, 200 Ω, 25 μF) using a Gene Pulser® (Bio-Rad, USA). Colonies with correct PCR product sizes were selected as candidate strains.

### Flow cytometry analysis

The protocol used for flow cytometry analysis of the promoter strength in terms of fluorescence intensity was modified slightly from a previous study [69]. Briefly, cells were washed with phosphate-buffered saline (PBS) twice and then resuspended into PBS to a concentration of 10$^7$ cells/mL. Cells were analyzed by flow cytometry using Beckman CytoFLEX FCM (Beckman Coulter, USA) with the PBS as the sheath fluid. The fluorescence of EGFP was excited with the 488 nm and detected with FITC; opmCherry was excited with the 561 nm and detected with PC5.5 [59, 70, 71].

As recommended by the manufacturer, compensation was applied to ensure that the EGFP has minimal affection on the detection of opmCherry. To avoid rare events which could affect the population distribution, at least 20,000 events of each sample were analyzed. Data were processed via FlowJo software (FlowJo, LLC, USA) based on the user manual with the recommended parameters. The mean fluorescence intensity of triplicates was calculated; then the ratio of average EGFP/average opmCherry was used to quantify the strength of each promoter. In addition, the standard deviation (STDEV) was set as the error bar.

### Quantitative real-time PCR analysis

The transcription levels of EGFP and opmCherry in the dual reporter-gene system were estimated at the same thermal cycling conditions. The cell extract samples were taken at exponential phase and used for quantitative real-time PCR (qPCR) analysis. Total RNA from *Z. mobilis* ZM4 strains harboring the inducible dual reporter-gene system was extracted using TRIzol reagent (Invitrogen, USA), and RNA quality was examined by NanoDrop 8000 (Thermo-Fisher, USA). The residual DNA removal and the reverse transcription was performed using the iScript™ gDNA Clear cDNA Synthesis Kits according to the manufacturer's instructions (Bio-Rad, USA).

The qPCR reactions were carried out using iTaq™ Universal SYBR® Green Supermix (Bio-Rad, USA) on a CFX96 Real-Time System (Bio-Rad, USA) as described previously [13, 15, 21, 23]. Briefly, highly purified salt-free primer pairs for EGFP and opmCherry were synthesized (Genscript, China) with an similar annealing temperature of 60 °C (Additional file 1: Table S1). PCR product specificity was confirmed through melting curve analysis. The following run protocol was used: 95 °C 5 min for denaturation; (95 °C 15 s, 60 °C 10 s, and 72 °C 30 s) 40 times for amplification and quantification with a single florescence measurement. A melting curve program (60–95 °C with heating rate of 0.1 °C per second and a continuous fluorescence measurement) was used to confirm the specificity of the primer pairs. An absolute quantification based on an internal calibration curve was applied for qPCR data analysis.

### Western blot analysis

Log phase and stationary phase cells were harvested to conduct the flow cytometry analysis. Cells were lysed and protein was extracted using Protein Extraction Kit (Zomanbio, China). Total protein concentrations of total cellular lysates were measured by the Bradford method with 200 ng of total protein loaded for each sample. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was performed with a 5% stacking and a 12% running gel, followed by stained with Coomassie Brilliant Blue R-250. Molecular weight was estimated

Yang *et al. Biotechnol Biofuels*    (2019) 12:52

Page 11 of 13

using a pre-stained protein ladder (10–170 kDa, Thermo, Lithuania) [72].

For Western blot analysis, after the electrophoresis, gels were transferred to methanol-activated PVDF membranes using the Trans-Blot® Semi-Dry Electrophoretic Transfer Cell (Bio-Rad, USA) and run for 20 min at 25 V. PVDF membranes was then blocked with 5% non-fat milk in phosphate-buffered saline with Tween 20 (PBST) for 1 h at room temperature, and subsequently EGFP or opmCherry were probed with primary antibody (1:5000, Proteintech, China), respectively. Peroxidase-conjugated goat anti-Mouse IgG (1:5000, Proteintech, China) was used as secondary antibodies. Color development was performed by West Dure Extended Duration Substrate Kit (AntGene, China). All images were visualized using AI600 Imaging System (GE, USA), and analyzed by ImageJ (National Institutes of Health, USA).

## Additional files

**Additional file 1: Table S1.** Nucleotide sequences of promoters with different strengths and their gene name and function. **Table S2.** Primers used in this study. The lowercases in each primer are the homologous arms designed to assembly with the vector. The primers prefixed with Q are used in qPCR.

**Additional file 2: Figure S1.** Determination of the concentration of tetracycline based on the effect for cellular growth with a broad range of 0–50 μg/mL (A), and a narrow range of 0–1.0 μg/mL (B). **Figure S2.** Flow cytometry results of RBSs with different strengths in *Z. mobilis*. The concentrations of tetracycline are 0, 0.2, 0.4, 0.6, 0.8, or 1.0 μg/mL, respectively. The Ptet-RBS-ori represents the original Ptet-RBS sequence, and the Ptet-RBS-10, -1K, -10K and –Max represent the RBS sequences with different translation initiation rates of 10, 1000, 10000, and 15000 based on bioinformatics server RBS calculator. **Figure S3.** The fluorescence intensity of EGFP (A), and opmCherry (B), and the corresponding correlation between two repeat experiments by two individual researchers (C).

## Abbreviations
AFM: atomic force microscope; CFP: cyan fluorescent protein; EGFP: enhanced green fluorescent protein; EMSA: electrophoresis mobility shift assay; FCM: flow cytometry; NGS: next-generation sequencing; opCFP: codon-optimized CFP; opEGFP: codon-optimized EGFP; opmCherry: codon-optimized mCherry; PBS: phosphate-buffered saline; PBST: phosphate-buffered saline with Tween 20; qPCR: quantitative real-time PCR; RBS: ribosomal binding sites; RFP: red fluorescent protein; SDS-PAGE: sodium dodecyl sulphate polyacrylamide gel electrophoresis; UTR: untranslated region.

## Authors' contributions
SY conceived and designed the experiments with inputs from all authors. YY and WS constructed the dual reporter-gene vector and plasmids, conducted qPCR, Western blot, and FACS experiments with help from JH, RL, and YX. YY and SY wrote the manuscript with help from WS, and MEH and HW conducted extensive review. All authors contributed to data analyses. All authors revised the final manuscript. All authors read and approved the final manuscript.

## Author details
¹ State Key Laboratory of Biocatalysis and Enzyme Engineering, Hubei Collaborative Innovation Center for Green Transformation of Bio-resources, Environmental Microbial Technology Center of Hubei Province, and School of Life Sciences, Hubei University, Wuhan 430062, China. ² Biosciences Center, National Renewable Energy Laboratory, Golden, CO 80401, USA. ³ National Bioenergy Center, National Renewable Energy Laboratory, Golden, CO 80401, USA.

## Competing interests
The authors declare that they have no competing interests.

## Availability of data and materials
The datasets used for this study are available from NCBI GEO and SRA databases based on previous publication.

## Consent for publication
All authors read and approved the final manuscript.

## Ethics approval and consent to participate
The authors declare that this study does not involve human subjects, human material and human data.

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

## References
1. Wang X, He Q, Yang Y, Wang J, Haning K, Hu Y, Wu B, He M, Zhang Y, Bao J, et al. Advances and prospects in metabolic engineering of *Zymomonas mobilis*. Metab Eng. 2018;50:57–73.
2. Yang S, Fei Q, Zhang Y, Contreras LM, Utturkar SM, Brown SD, Himmel ME, Zhang M. *Zymomonas mobilis* as a model system for production of biofuels and biochemicals. Microb Biotechnol. 2016;9(6):699–717.
3. He MX, Wu B, Qin H, Ruan ZY, Tan FR, Wang JL, Shui ZX, Dai LC, Zhu QL, Pan K, et al. *Zymomonas mobilis*: a novel platform for future biorefineries. Biotechnol Biofuels. 2014;7(101):101.
4. Yang Y, Hu M, Tang Y, Geng B, Qiu M, He Q, Chen S, Wang X, Yang S. Progress and perspective on lignocellulosic hydrolysate inhibitor tolerance improvement in *Zymomonas mobilis*. Bioresour Bioprocess. 2018;5(1):6.
5. Khalil AS, Collins JJ. Synthetic biology: applications come of age. Nat Rev Genet. 2010;11(5):367–79.
6. Deaner M, Alper HS. Promoter and terminator discovery and engineering. Adv Biochem Eng Biotechnol. 2016;162:21–44.
7. Cho SH, Haning K, Shen W, Blome C, Li R, Yang S, Contreras LM. Identification and characterization of 5′ untranslated regions (5′ UTRs) in *Zymomonas mobilis* as regulatory biological parts. Front Microbiol. 2017;8:2432.

Yang *et al. Biotechnol Biofuels*     (2019) 12:52

Page 12 of 13

8.   Repoila F, Darfeuille F. Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects. Biol Cell. 2009;101(2):117–31.
9.   Urban JH, Vogel J. Translational control and target recognition by *Escherichia coli* small RNAs in vivo. Nucleic Acids Res. 2007;35(3):1018–37.
10.  Cho SH, Lei R, Henninger TD, Contreras LM. Discovery of ethanol-responsive small RNAs in *Zymomonas mobilis*. Appl Environ Microbiol. 2014;80(14):4189–98.
11.  Moreno JM, Sorensen HP, Mortensen KK, Sperling-Petersen HU. Macromolecular mimicry in translation initiation: a model for the initiation factor IF2 on the ribosome. IUBMB Life. 2000;50(6):347–54.
12.  Hardesty B, Kramer G. Folding of a nascent peptide on the ribosome. Prog Nucleic Acid Res Mol Biol. 2001;66:41–66.
13.  Yang S, Pan C, Hurst GB, Dice L, Davison BH, Brown SD. Elucidation of *Zymomonas mobilis* physiology and stress responses by quantitative proteomics and transcriptomics. Front Microbiol. 2014;5:246.
14.  Yang S, Pappas KM, Hauser LJ, Land ML, Chen GL, Hurst GB, Pan C, Kouvelis VN, Typas MA, Pelletier DA, et al. Improved genome annotation for *Zymomonas mobilis*. Nat Biotechnol. 2009;27(10):893–4.
15.  Yang S, Tschaplinski TJ, Engle NL, Carroll SL, Martin SL, Davison BH, Palumbo AV, Rodriguez M Jr, Brown SD. Transcriptomic and metabolomic profiling of *Zymomonas mobilis* during aerobic and anaerobic fermentations. BMC Genom. 2009;10:34.
16.  Yang S, Land ML, Klingeman DM, Pelletier DA, Lu TY, Martin SL, Guo HB, Smith JC, Brown SD. Paradigm for industrial strain improvement identifies sodium acetate tolerance loci in *Zymomonas mobilis* and *Saccharomyces cerevisiae*. Proc Natl Acad Sci USA. 2010;107(23):10395–400.
17.  He MX, Wu B, Shui ZX, Hu QC, Wang WG, Tan FR, Tang XY, Zhu QL, Pan K, Li Q, et al. Transcriptome profiling of *Zymomonas mobilis* under furfural stress. Appl Microbiol Biotechnol. 2012;95(1):189–99.
18.  He MX, Wu B, Shui ZX, Hu QC, Wang WG, Tan FR, Tang XY, Zhu QL, Pan K, Li Q, et al. Transcriptome profiling of *Zymomonas mobilis* under ethanol stress. Biotechnol Biofuels. 2012;5(1):75.
19.  Jeon YJ, Xun Z, Su P, Rogers PL. Genome-wide transcriptomic analysis of a flocculent strain of *Zymomonas mobilis*. Appl Microbiol Biotechnol. 2012;93(6):2513–8.
20.  Skerker JM, Leon D, Price MN, Mar JS, Tarjan DR, Wetmore KM, Deutschbauer AM, Baumohl JK, Bauer S, Ibanez AB, et al. Dissecting a complex chemical stress: chemogenomic profiling of plant hydrolysates. Mol Syst Biol. 2013;9:674.
21.  Yang S, Pan C, Tschaplinski TJ, Hurst GB, Engle NL, Zhou W, Dam P, Xu Y, Rodriguez M Jr, Dice L, et al. Systems biology analysis of *Zymomonas mobilis* ZM4 ethanol stress responses. PLoS ONE. 2013;8(7):e68886.
22.  Deutschbauer A, Price MN, Wetmore KM, Tarjan DR, Xu Z, Shao W, Leon D, Arkin AP, Skerker JM. Towards an informative mutant phenotype for every bacterial gene. J Bacteriol. 2014;196(20):3643–55.
23.  Yang S, Franden MA, Brown SD, Chou YC, Pienkos PT, Zhang M. Insights into acetate toxicity in *Zymomonas mobilis* 8b using different substrates. Biotechnol Biofuels. 2014;7(1):140.
24.  Dunn KL, Rao CV. High-throughput sequencing reveals adaptation-induced mutations in pentose-fermenting strains of *Zymomonas mobilis*. Biotechnol Bioeng. 2015;112(11):2228–40.
25.  Yi X, Gu H, Gao Q, Liu ZL, Bao J. Transcriptome analysis of *Zymomonas mobilis* ZM4 reveals mechanisms of tolerance and detoxification of phenolic aldehyde inhibitors from lignocellulose pretreatment. Biotechnol Biofuels. 2015;8:153.
26.  Zhang K, Shao H, Cao Q, He MX, Wu B, Feng H. Transcriptional analysis of adaptation to high glucose concentrations in *Zymomonas mobilis*. Appl Microbiol Biotechnol. 2015;99(4):2009–22.
27.  Charoensuk K, Sakurada T, Tokiyama A, Murata M, Kosaka T, Thanonkeo P, Yamada M. Thermotolerant genes essential for survival at a critical high temperature in thermotolerant ethanologenic *Zymomonas mobilis* TISTR 548. Biotechnol Biofuels. 2017;10:204.
28.  Widiastuti H, Kim JY, Selvarasu S, Karimi IA, Kim H, Seo JS, Lee DY. Genome-scale modeling and in silico analysis of ethanologenic bacteria *Zymomonas mobilis*. Biotechnol Bioeng. 2011;108(3):655–65.
29.  Tsantili IC, Karim MN, Klapa MI. Quantifying the metabolic capabilities of engineered *Zymomonas mobilis* using linear programming analysis. Microb Cell Fact. 2007;6:8.
30.  Zhou S, Ding R, Chen J, Du G, Li H, Zhou J. Obtaining a panel of cascade promoter-5′-UTR complexes in *Escherichia coli*. ACS Synth Biol. 2017;6(6):1065–75.
31.  Yang S, Vera JM, Grass J, Savvakis G, Moskvin OV, Yang Y, McIlwain SJ, Lyu Y, Zinonos I, Hebert AS, et al. Complete genome sequence and the expression pattern of plasmids of the model ethanologen *Zymomonas mobilis* ZM4 and its xylose-utilizing derivatives 8b and 2032. Biotechnol Biofuels. 2018;11:125.
32.  Seo JS, Chong H, Park HS, Yoon KO, Jung C, Kim JJ, Hong JH, Kim H, Kim JH, Kil JI, et al. The genome sequence of the ethanologenic bacterium *Zymomonas mobilis* ZM4. Nat Biotechnol. 2005;23(1):63–8.
33.  Gao X, Gao Q, Bao J. Improving cellulosic ethanol fermentability of *Zymomonas mobilis* by overexpression of sodium ion tolerance gene ZMO0119. J Biotechnol. 2018;282:32–7.
34.  Conway T, Sewell GW, Ingram LO. Glyceraldehyde-3-phosphate dehydrogenase gene from *Zymomonas mobilis*: cloning, sequencing, and identification of promoter region. J Bacteriol. 1987;169(12):5653–62.
35.  Conway T, Ingram LO. Phosphoglycerate kinase gene from *Zymomonas mobilis*: cloning, sequencing, and localization within the *gap* operon. J Bacteriol. 1988;170(4):1926–33.
36.  Conway T, Sewell GW, Osman YA, Ingram LO. Cloning and sequencing of the alcohol dehydrogenase II gene from *Zymomonas mobilis*. J Bacteriol. 1987;169(6):2591–7.
37.  Barnell WO, Liu J, Hesman TL, O'Neill MC, Conway T. The *Zymomonas mobilis glf, zwf, edd,* and *glk* genes form an operon: localization of the promoter and identification of a conserved sequence in the regulatory region. J Bacteriol. 1992;174(9):2816–23.
38.  Conway T, Fliege R, Jones-Kilpatrick D, Liu J, Barnell WO, Egan SE. Cloning, characterization and expression of the *Zymononas mobilis eda* gene that encodes 2-keto-3-deoxy-6-phosphogluconate aldolase of the Entner–Doudoroff pathway. Mol Microbiol. 1991;5(12):2901–11.
39.  Gunasekaran P, Karunakaran T, Cami B, Mukundan AG, Preziosi L, Baratti J. Cloning and sequencing of the *sacA* gene: characterization of a sucrase from *Zymomonas mobilis*. J Bacteriol. 1990;172(12):6727–35.
40.  Pond JL, Eddy CK, Mackenzie KF, Conway T, Borecky DJ, Ingram LO. Cloning, sequencing, and characterization of the principal acid phosphatase, the *phoC*+ product, from *Zymomonas mobilis*. J Bacteriol. 1989;171(2):767–74.
41.  Keshav KF, Yomano LP, An HJ, Ingram LO. Cloning of the *Zymomonas mobilis* structural gene encoding alcohol dehydrogenase I (*adhA*): sequence comparison and expression in *Escherichia coli*. J Bacteriol. 1990;172(5):2491–7.
42.  Zhang M, Eddy C, Deanda K, Finkelstein M, Picataggio S. Metabolic engineering of a pentose metabolism pathway in ethanologenic *Zymomonas mobilis*. Science. 1995;267(5195):240–3.
43.  Mackenzie KF, Conway T, Aldrich HC, Ingram LO. Expression of *Zymomonas mobilis adhB* (encoding alcohol dehydrogenase II) and *adhB-lacZ* operon fusions in recombinant *Z. mobilis*. J Bacteriol. 1989;171(9):4577–82.
44.  Yang S, Mohagheghi A, Franden MA, Chou YC, Chen X, Dowe N, Himmel ME, Zhang M. Metabolic engineering of *Zymomonas mobilis* for 2,3-butanediol production from lignocellulosic biomass sugars. Biotechnol Biofuels. 2016;9(1):189.
45.  Yu T, Zhou YJ, Huang M, Liu Q, Pereira R, David F, Nielsen J. Reprogramming yeast metabolism from alcoholic fermentation to lipogenesis. Cell. 2018;174(6):1549–58.
46.  Anish R, Hossain MB, Jacobson RH, Takada S. Characterization of transcription from TATA-less promoters: identification of a new core promoter element XCPE2 and analysis of factor requirements. PLoS ONE. 2009;4(4):e5103.
47.  Zhang X, Yao Z, Duan Y, Zhang X, Shi J, Xu Z. Investigation of specific interactions between T7 promoter and T7 RNA polymerase by force spectroscopy using atomic force microscope. Biochem J. 2018;475(1):319–28.
48.  Hammer K, Mijakovic I, Jensen PR. Synthetic promoter libraries–tuning of gene expression. Trends Biotechnol. 2006;24(2):53–5.
49.  Subramanian S. Srienc F: quantitative analysis of transient gene expression in mammalian cells using the green fluorescent protein. J Biotechnol. 1996;49:137–51.
50.  Leveau JH, Lindow SE. Predictive and interpretive simulation of green fluorescent protein expression in reporter bacteria. J Bacteriol. 2001;183(23):6752–62.
51.  Lu C, Bentley WE, Rao G. A high-throughput approach to promoter study using green fluorescent protein. Biotechnol Prog. 2004;20(6):1634–40.

Yang *et al. Biotechnol Biofuels*     (2019) 12:52

Page 13 of 13

52. Delvigne F, Pecheux H, Tarayre C. Fluorescent reporter libraries as useful tools for optimizing microbial cell factories: a review of the current methods and applications. Front Bioeng Biotechnol. 2015;3:147.

53. Segall-Shapiro TH, Sontag ED, Voigt CA. Engineered promoters enable constant gene expression at any copy number in bacteria. Nat Biotechnol. 2018;36(4):352–8.

54. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002;297(5584):1183–6.

55. Rudge TJ, Brown JR, Federici F, Dalchau N, Phillips A, Ajioka JW, Haseloff J. Characterization of intrinsic properties of promoters. ACS Synth Biol. 2016;5(1):89–98.

56. Ou J, Furusawa C, Yomo T, Shimizu H. Analysis of stochasticity in promoter activation by using a dual-fluorescence reporter system. Biosystems. 2009;97(3):160–4.

57. Yang G, Jia D, Jin L, Jiang Y, Wang Y, Jiang W, Gu Y. Rapid generation of universal synthetic promoters for controlled gene expression in both gas-fermenting and saccharolytic *Clostridium* species. ACS Synth Biol. 2017;6(9):1672–8.

58. Liu J, Wei Y, Huang L, Wang Y, Chen D, Wu H, Liu C. Functional characterization of a new promoter isolated from torque teno sus virus 1. Arch Virol. 2016;161(2):303–6.

59. Shaner NC, Campbell RE, Steinbach PA, Giepmans BN, Palmer AE, Tsien RY. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. Nat Biotechnol. 2004;22(12):1567–72.

60. Kremers GJ, Gilbert SG, Cranfill PJ, Davidson MW, Piston DW. Fluorescent proteins at a glance. J Cell Sci. 2011;124(Pt 2):157–60.

61. Yang TT, Cheng L, Kain SR. Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein. Nucleic Acids Res. 1996;24(22):4592–3.

62. Gordon A, Colman-Lerner A, Chin TE, Benjamin KR, Yu RC, Brent R. Single-cell quantification of molecules and rates using open-source microscope-based cytometry. Nat Methods. 2007;4(2):175–81.

63. Dam P, Olman V, Harris K, Su Z, Xu Y. Operon prediction using both genome-specific and general genomic information. Nucleic Acids Res. 2007;35(1):288–98.

64. Mao F, Dam P, Chou J, Olman V, Xu Y. DOOR: a database for prokaryotic operons. Nucleic Acids Res. 2009;37(Database issue):D459–63.

65. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009;6(5):343–5.

66. Reynen M, Reipen I, Sahm H, Sprenger GA. Construction of expression vectors for the gram-negative bacterium *Zymomonas mobilis*. Mol Gen Genet. 1990;223(2):335–41.

67. Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: a laboratory manual, 2nd edn. Gene cloning & DNA analysis an introduction. Oxford: Wiley; 1989.

68. Noel RJ Jr, Reznikoff WS. Structural studies of *lac*UV5-RNA polymerase interactions in vitro Ethylation interference and missing nucleoside analysis. J Biol Chem. 2000;275(11):7708–12.

69. Hakkila K, Maksimow M, Rosengren A, Karp M, Virta M. Monitoring promoter activity in a single bacterial cell by using green and red fluorescent proteins. J Microbiol Methods. 2003;54(1):75–9.

70. Chudakov DM, Matz MV, Lukyanov S, Lukyanov KA. Fluorescent proteins and their applications in imaging living cells and tissues. Physiol Rev. 2010;90(3):1103–63.

71. Telford WG, Hawley T, Subach F, Verkhusha V, Hawley RG. Flow cytometry of fluorescent proteins. Methods. 2012;57(3):318–30.

72. Shu M, Shen W, Yang S, Wang X, Wang F, Wang Y, Ma L. High-level expression and characterization of a novel serine protease in *Pichia pastoris* by multi-copy integration. Enzyme Microb Technol. 2016;92:56–66.