Microbial Cell Factories

RESEARCH Open Access

Characterization of MAP c21873-1 as a new counter-selectable marker for unmarked genetic modification of *Pichia pastoris*

Minzhi Liu[1,2†], Sihan Zhou[2†], Yunsong Cao[2], Keqin Yang[2], Yao Xiao[2] and Wei Wang[1,2*]

Abstract

Background Selection markers are useful in genetic modification of yeast *Pichia pastoris*. However, the leakage of the promoter caused undesired expression of selection markers especially those toxic proteins like MazF, halting the cell growth and hampering the genetic manipulation in procaryotic system. In this study, a new counter-selectable marker-based strategy has been established for seamless modification with high efficiency and low toxicity.

Results At first, the leaky expression of the enhanced green fluorescent protein (EGFP) as a reporter gene under the control of six inducible promoters of *P. pastoris* was investigated in two hosts *Escherichia coli* and *P. pastoris*, respectively. The results demonstrated that the DAS1 and FDH1 promoters (P_{DAS1} and P_{FDH1}) had the highest leakage expression activities in procaryotes and eukaryotes, and the DAS2 promoter (P_{DAS2}) was inducible with medium strength but low leakage expression activity, all of which were selected for further investigation. Next, *Mirabilis* antiviral proteins (MAPs) c21873-1, c21873-1T (truncated form of c21873-1) and c23467 were mined as the new counter-selectable markers, and hygromycin B (Hyg B) resistance gene was used as the positive-selectable marker, respectively. Then, modular plasmids with MAP-target gene-Hyg B cassettes were constructed and used to transform into *P. pastoris* cells after linearization, and the target genes were integrated into its genome at the BmT1 locus through single-crossover homologous recombination (HR). After counter-selection induced by methanol medium, the markers c21873-1 and c21873-1T were recycled efficiently. But c23467 failed to be recycled due to its toxic effect on the *P. pastoris* cells. At last, the counter-selectable marker c21873-1 under the tightly regulated P_{DAS2} enabled the encoding genes of reporter EGFP and tested proteins to be integrated into the target locus and expressed successfully.

Conclusions We have developed MAP c21873-1 as a novel counter-selectable marker which could perform efficient gene knock-in by site-directed HR. Upon counter-selection, the marker could be recycled for repeated use, and no undesirable sequences were introduced except for the target gene. This unmarked genetic modification strategy may be extended to other genetic modification including but not limited to gene knock-out and site-directed mutagenesis in future.

†Minzhi Liu and Sihan Zhou contributed equally to this work and should be considered co-first authors.

*Correspondence:
Wei Wang
wwang@imm.ac.cn

Full list of author information is available at the end of the article

Keywords Counter-selectable markers, *Mirabilis* antiviral protein, Site-directed homologous recombination, *Pichia pastoris*, Protein expression

Background

As a non-conventional methylotrophic yeast, *Pichia pastoris* has been harnessed extensively as a eukaryotic host for heterologous protein expression [1]. In contrast to prokaryotic expression system, it has many advantages, including possessing various inducible promoters under the strict regulation, enabling high-density fermentation and easy purification due to secretion of products to the cultivation medium, together with the capability of post-translational modifications to keep physical functions of proteins, etc. Moreover, a great breakthrough of it is that, the Food and Drug Administration (FDA) acknowledges *P. pastoris* as a generally recognized as safe (GRAS) strain [2] and approves to market biopharmaceutical products [3–5], making it a potential cell factory in the industrial field for high level recombination protein production.

The expression of heterologous proteins needs the insertion of exogenous genes into the genome of *P. pastoris*, since it lacks of stable natural plasmid systems [1]. As is known, site-directed homologous recombination (HR) and non-homologous end-joining (NHEJ) are two mechanisms involved in the integration process in *P. pastoris* [6]. However, the NHEJ pathway is the major mechanism with a relatively high strength which often affects the availability of gene integration to the target position, and leads to the low efficiency of gene knock-out as well. Therefore, it needs the help of genetic manipulations to reduce the strength of the NHEJ pathway and increase the probability of HR pathway. With the development of modern molecular biology and biotechnology, the emergence of a variety of genetic techniques makes them more convenient. One of the common methods is based on the use of site-specific recombinase systems, such as the Cre recombinase/*lox*P sequence (Cre/*lox*P) system [7, 8] and the flippase (FLP) and flippase recognition target (FLP/*FRT*) system [9]. However, a scar of *lox*P or *FRT* is left in the genome of *P. pastoris* after the gene recombination, which is redundant for native-site gene replacement or mutagenesis. The scar is also detrimental for multiple gene modifications when the latter modification locus is close to the former one, because the unexpected recombination may occur between the old scar and the repeated *lox*P or *FRT* sequence [10]. The other method is the utilization of selection markers, such as antibiotic resistant genes, which are very limited and can only be used once. Therefore, exploitation of counter-selectable markers and scarless rescue of them are of great importance. As a classical counter-selectable marker, uracil biosynthetic gene *URA3* has been commonly used in many yeast strains including *P. pastoris* [11–13]. During the selection, orotidine-5'-phosphate decarboxylase encoded by *ura*3 can catalyze 5-fluoroorotic acid (5-FOA) into the toxic product 5-fluorouracil, causing the death of host cells [14]. Unfortunately, its application has been limited for the reason that only auxotrophic host strains with *ura3* mutant can be screened in the presence of 5-FOA. Besides, *ura3* auxotroph of *P. pastoris* grows slowly even after supplementing uracil in the medium, making its downstream manipulation inconvenient [13].

The occurrence of another counter-selectable marker, MazF, overcomes the shortcomings mentioned above. MazF is a stable toxin protein derived from stress-induced toxin-antitoxin (TA) modules in *Escherichia coli* [15, 16]. It functions as an endoribonuclease with ACA cleavage site specificity in a manner independent of ribosome, cleaves cellular mRNA and inhibits cell growth under stress conditions [17]. The application of MazF as a counter-selectable marker can be attributed to its programmed cell death effect on cells, not only in prokaryotic systems such as *E. coli* [15, 18] and *Bacillus* species [19, 20], but also in eukaryotic systems such as *P. pastoris* [10], *Saccharomyces cerevisiae* [21] and *Hansenula polymorpha* [22]. Our research group previously used MazF as the counter-selectable marker for gene knock-out in *P. pastoris*. However, the standard cloning procedure often failed. Usually, the target DNA fragment produced by polymerase chain reaction (PCR) amplification was incorrectly joined with the vector DNA, or lost. Consequently, neither a few of transformants grew on the selective plates, nor did they propagate in culture medium any more. It was suspected that undesirable promoter leakage might cause MazF protein expression at the cloning stage, the latter of whose toxicity arrested the normal cell growth. To confirm this perspective, the first objective of our study was to investigate the leakage expression of promoters in *E. coli*, in which six inducible promoters (alcohol oxidase 1, P_{AOX1}; dihydroxyacetone synthase 1, P_{DAS1}; dihydroxyacetone synthase 2, P_{DAS2}; formate dehydrogenase, P_{FDH1}; formaldehyde dehydrogenase, P_{FLD1}; Protein PXR1, GenBank NO. CAY67289, P_{PXR1}) commonly used in *P. pastoris* were selected.

Based on that, the other goal of our study was to find a novel counter-selectable marker to replace MazF. On the one hand, its deleterious effect on the normal growth of bacteria seriously interfered with routine genetic manipulation. On the other hand, most of the studies have only achieved the proof of concept since the integrated cassettes are very short [6, 23, 24] when using MazF as the counter-selectable marker. Ribosome-inactivating proteins (RIPs) are a group of plant enzymes

with N-glycosidase activity which mainly depurinate an adenine residue from 23S/25S/28S rRNA, inactivating ribosomes and inhibiting protein synthesis [25]. RIPs are primarily found in plants, bacteria and fungi, and even in some insects by recent evidence [26, 27]. Generally, they can be divided into three categories according to their different structures [25]. Type I RIPs comprise a single domain protein with N-glycosylase activity, such as the classical pokeweed antiviral protein (PAP) [28] and saporin [29]. Type II RIPs contain a domain functionally equivalent to Type I RIP linked via a disulfide bond to a lectin chain. Traditionally, type II RIPs have higher toxicity than type I RIPs because of the lectin chain of the latter proteins, which attribute to binding of membrane glycoproteins and facilitating their entrance into cells. Ricin is a prototypic type II RIP which was first discovered in *Ricinus communis* [30]. Type III RIPs are either single RIP domains formed from post-translational proteolytic processing, or RIP domains fused with unknown functional domains, such as the early characterized members maize B-32 and jasmonate-induced RIP (JIP60). They have drawn much attention for their broad-spectrum activities including insecticide, anti-bacteria, anti-fungi, anti-virus and anti-cancer [31–33]. Except for experiments in the lab, some RIPs have even been tested in the clinical trials. For example, Trichosanthin (TCS) was the first RIP used in phase I clinical trials to treat human immunodeficiency virus (HIV) infection [34]. Besides, PAP for HIV therapy [28], ricin for leukemia therapy [30], and gelonin for myeloid malignancies therapy [35] have also been carried out in clinical trials. However, the short plasma half-life, non-selective cytotoxicity and antigenicity of most RIPs impeded their use as therapeutic proteins [36]. Due to the disturbance of the final protein synthesis, it was hypothesized that RIPs might be taken as counter-selectable markers similar to MazF, which had never been reported before. Hence, the availabilities of three *Mirabilis* antiviral proteins (MAPs) as counter-selectable markers were explored. Among them, one was discovered by transcriptome sequencing in our research (designated as c21873-1) for the first time, the other was its mutant with a truncated open reading frame (ORF) at the C-terminus (designated as c21873-1T), and the third one (Swiss-Prot NO. P21326.2, designated as c23467) was reported before [37]. The counter-selection functions of c21873-1, c21873-1T and c23467 were characterized, respectively. In our new strategy, using *pcsk2* and *kex2* as target genes, the long integration cassettes (about 14 kb) were successfully inserted into the BmT1 locus of *P. pastoris*' genome with high efficiencies. Finally, the target gene of 8.8 kb was inserted into the target site after scarless marker recycling of c21873-1. Furthermore, this strategy was also successfully applicated to the integration and expressions of heterologous enhanced green fluorescent protein (EGFP), endoplasmic reticulum protein 46 (ERp46) and prolyl 4-hydroxylase beta polypeptide (P4HB) in *P. pastoris*.

Results

The leaky expression of EGFP under the control of different promoters in *E. coli*

In order to implement a genetic manipulation for site-directed HR of exogenous genes into the BmT1 site of *P. pastoris*, a plasmid vector was constructed using the toxic gene *mazF* as the counter-selectable marker under the control of P_{AOX1} of *P. pastoris*. A failure scenario happened that the DNA elements used for HR were always incorrectly assembled in the host strain *E. coli*. So, it was suspected that the leakage of the promoter resulted in the undesirable production of the toxic MazF protein. As is known to all, MazF mediates programmed cell death of *E. coli* when cells suffer from adverse external conditions. Conversely, in normal conditions, one MazE dimer conjugates two MazF dimers to form a stable TA complex [38, 39], maintaining a homeostatic steady state of the cells. Hence, to counteract its adverse effect, a counterpart *mazE* gene was inserted into the plasmids to form mazEF TA system, which could neither remedy the observed problems. Then the fusion gene *mazF51-egfp* encoding N-terminal fifty-one amino acids of MazF and EGFP was constructed and used to investigate the leakage effect of P_{AOX1}. Regretfully, its inherent toxicity of truncated MazF still affected the normal growth of *E. coli* cells, so the verification assay was given up (data not shown). At last, the toxic gene *mazF* was deleted and the expression of EGFP under the control of different promoters was analyzed. As shown in Fig. 1, after 24-h culture, the leaky expressions of EGFP were detected in the cell pellet (Fig. 1a) and in the supernatant (Fig. 1b), respectively. Compared to the control group, the leaky expressions of the reporter protein EGFP under the control of P_{AOX1}, P_{DAS1} and P_{FLD1} were significantly increased in the cell pellet ($p<0.01$, $p<0.001$ and $p<0.05$, respectively). Among six promoters, mean fluorescence value of the experimental group under the control of P_{DAS1} was the strongest, about 2.6-fold of that of control group, indicating the most leakage severity of this promoter. In the supernatant, the leaky expressions of EGFP also existed in the experimental groups under the control of P_{FDH1} and P_{FLD1} ($p<0.05$ and $p<0.01$, respectively).

The leaky expression of EGFP under the control of different promoters in *P. pastoris*

To achieve the uniform expression of EGFP under the control of different promoters, the expression plasmids containing a *P. pastoris*-specific autonomous replicating sequence (PARS) and the encoding *egfp* gene were constructed and then used to transform into *P. pastoris* cells.

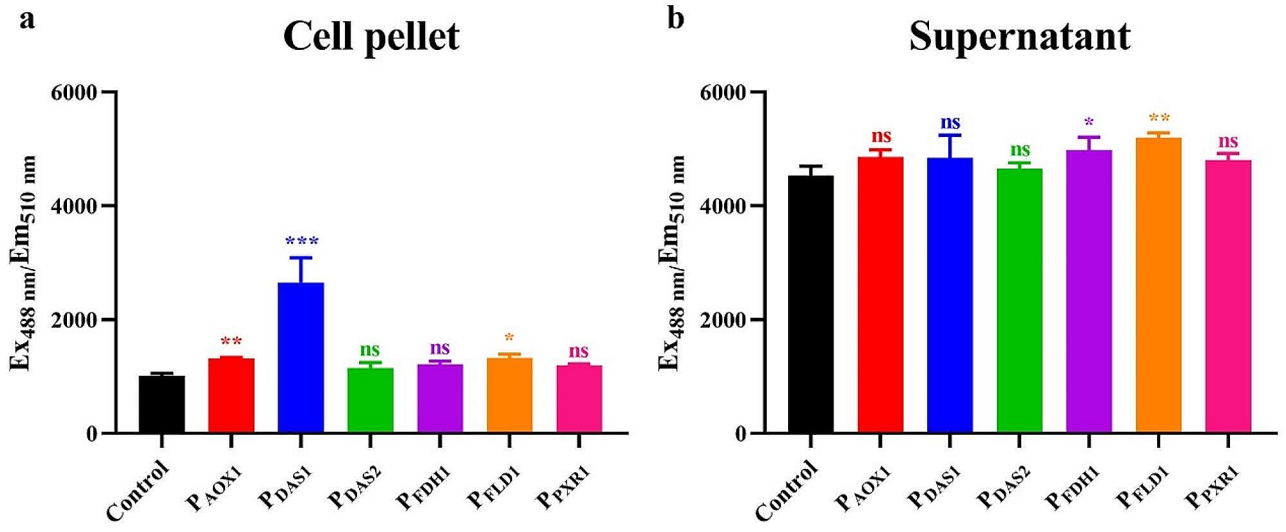

Fig. 1 The leaky expression of EGFP under the control of different promoters in *E. coli*. The fluorescence values in the cell pellet (**a**) and in the supernatant (**b**) were analyzed, respectively. By statistical analyses, the significant differences between the experimental group and the control group were compared, respectively

The cell growth curves and expressions of EGFP under the control of different promoters were shown in Fig. 2. Maximum optical density at 600 nm (OD_{600}) of the cell was achieved at 48 h for control group (OD_{600}=72.2±6.7) cultured in extract-peptone-dextrose (YPD) medium, then decreased a little and kept constant (OD_{600}≈65) for additional three days (Fig. 2a). The cells transformed with plasmids containing EGFP could increase steadily during five days' culture, but the growth rates of them were a little bit slower than that of control strain. However, maximum cell density was reached at 120 h for all groups when *P. pastoris* grew in BMMY medium replenishing methanol as a carbon source every 24 h (Fig. 2b). Among them, the OD_{600} values of the groups expressing EGFP under the control of P_{FDH1} and P_{PXR1} could reach about 75. Therefore, nutritional supplementation caused continuous cell growth in BMMY medium. The expression of EGFP in two media were quite different under the control of different promoters. In YPD medium, leaky expressions of EGFP under the control of P_{FDH1} and P_{FLD1} were stronger than that of other promoters (Fig. 2c). The fluorescence intensities were strongest at 72 h, whose values were 53-fold and 30-fold of control group, respectively, and decreased a little bit in the last two days. The strength of leaky capacity of promoters were P_{FDH1} > P_{FLD1} > P_{PXR1} > P_{AOX1} > P_{DAS1} > P_{DAS2}. Then the effects of different promoters on the expressions of EGFP induced by methanol in BMMY medium were investigated. The fluorescence intensities increased continuously in five days (Fig. 2d). EGFP almost did not express under the control of P_{PXR1}, since its fluorescence value was nearly the same as that of control group. Under the control of P_{FDH1}, the expression of EGFP was almost 58-fold of that of control group, indicating the strong initiating effect of this promoter on the downstream element. The strength of initiating effects of promoters were P_{FDH1} > P_{DAS1} > P_{AOX1} > P_{DAS2} > P_{FLD1} > P_{PXR1}. This result was a little bit different from that reported by Vogl et al., in which P_{DAS2} was the strongest promoter for the expression of EGFP. This might be the reason that the carbons (glucose vs. glycerol) for biomass accumulation and the concentration of methanol (5% vs. 1%) for induction were different [40]. In general, for gene manipulation, the promoter with the lowest leaky toxicity and the highest induction activity was an ideal candidate. As a result, P_{DAS1}, P_{DAS2} and P_{FDH1} (as a negative control) were chosen to examine the influences on HR of target genes.

The discovery of novel counter-selectable markers from transcriptome sequencing analysis of *Mirabilis jalapa*

Based on transcriptome sequencing analysis of *Mirabilis jalapa*, our team found out a novel sequence named c21873-1. It possesses 49.17% sequence identity with the reported MAP c23467, whose activity may be similar to it. Through sequence blasting analysis in NCBI database, another twenty-four representative RIPs from plants, bacteria and fungi were chosen, and their identities compared to c21873-1 were from 25.19 to 36.06% (Table S1). Then, MazF, c21873-1, c23467 and twenty-four RIPs were selected to construct a phylogenetic tree to investigate their relationships in the evolutionary history. As shown in Fig. 3, the phylogenetic analysis demonstrates that c21873-1 and c23467 are closely relative to type I RIPs such as PAP-S and saporin, but far from type II RIPs such as abrin-a-like protein and ricin in the evolutionary process. Besides, from a structural point of view, c21873-1

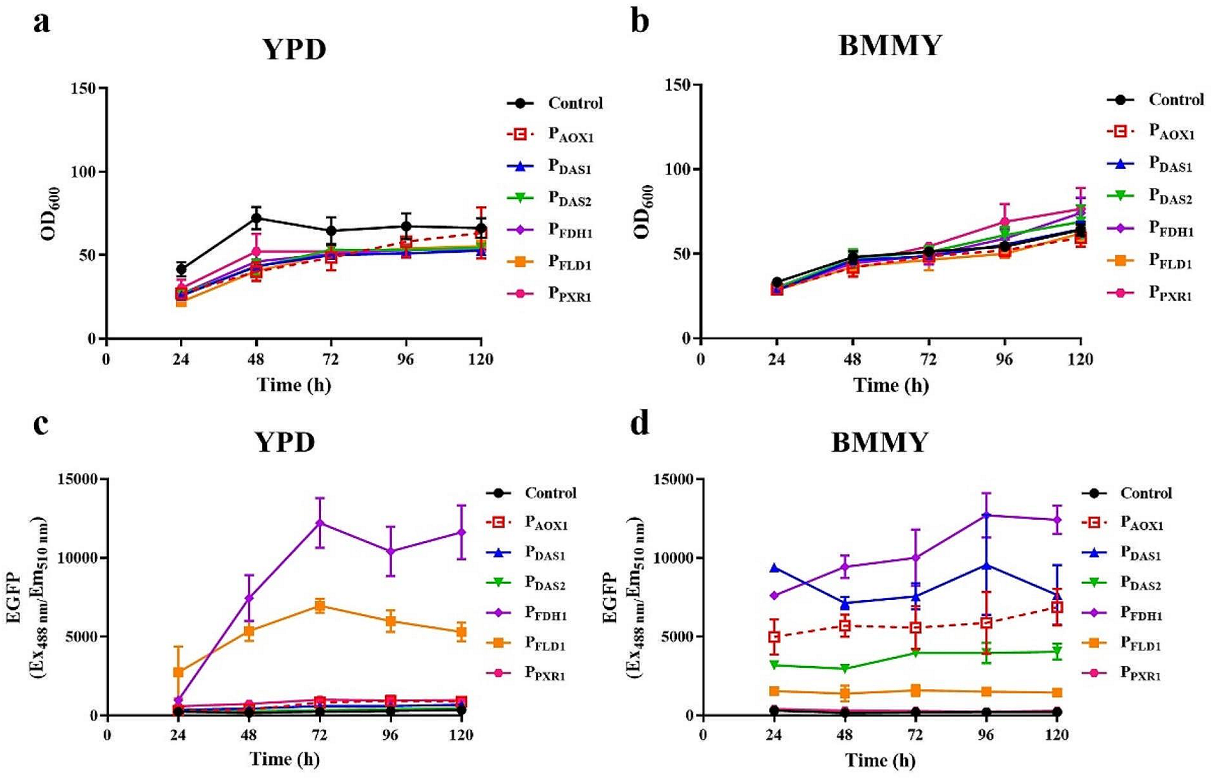

Fig. 2 The leaky expression of EGFP under the control of different promoters in *P. pastoris*. The yeast cells were cultured in in YPD and BMMY media for 120 h, respectively. Every 24 h, the OD_{600} values of them were monitored and the cell growth curves (**a**, **b**) were depicted. Meanwhile, the same quantity of the cells (OD_{600} = 8.0) was removed out and the fluorescence values of expressed EGFP (**c**, **d**) were tested

is also different from MazF and barley JIP60 (type III RIP), which is consistent with the fact that there are no significant sequence similarities among them (Table S1). Considering that the lethal effect of MazF seriously hampered the genetic manipulation in our previous work, and RIPs have the similar functions of MazF in blocking protein synthesis, it prompted the idea of creating new counter-selectable markers from these RIPs. Naturally, c21873-1 and c23467 were inferred to be less cytotoxic and proper to be the candidates. Therefore, the investigation of c21873-1, c21873-1T and c23467 as new counter-selectable markers was performed in our next research.

The test of feasibility of the mined MAP c21873-1 as a new counter-selectable marker

The vectors containing MAP, the target gene (*pcsk2* or *kex2*) and BmT1 homologous region were constructed to investigate the availability of counter-selection (Fig. 4). To avoid the influence of leakage toxicity of the promoter on the counter-selectable marker such as MazF, an appropriate promoter should be selected carefully. Therefore, P_{DAS1} and P_{DAS2} were chosen because of their lowest leaky expression and higher induction activity. Besides, P_{FDH1} with the highest leakage toxicity was chosen as a negative control. The target gene was designed to insert into BmT1 homologous region of *P. pastoris*, so the disruption of the BmT1 locus by the integrated expression cassette would cause reduced hypermannosylation. The vector was linearized by restriction enzyme *EcoR* I, transformed into the yeast cell, and followed by HR at BmT1-3′ homologous region through the single-crossover event. The positive colonies were screened by analysis of PCR amplification, and then cultured in corresponding induction medium to perform counter-selection. During this process, the MAPs or MAPs- gene of interest (GOI)-Hyg B cassettes would be removed by internal HR, generating MAPs deletion-type or wild-type strain, respectively.

Screening and characterization of site-directed HR of GOI

As mentioned above, *pcsk2* and *kex2* were chosen as target genes, and the effects of different promoters and selectable markers on site-desired HR were investigated. The ORF sequence of BmT1 in *P. pastoris*′ genome was about 2.5 kb, and MAPs-GOI-Hyg B cassettes were about 11.2 kb. From hundreds of Hyg B resistant transformants obtained, twenty colonies were selected and verified by colony PCR analysis. The amplification of the integration

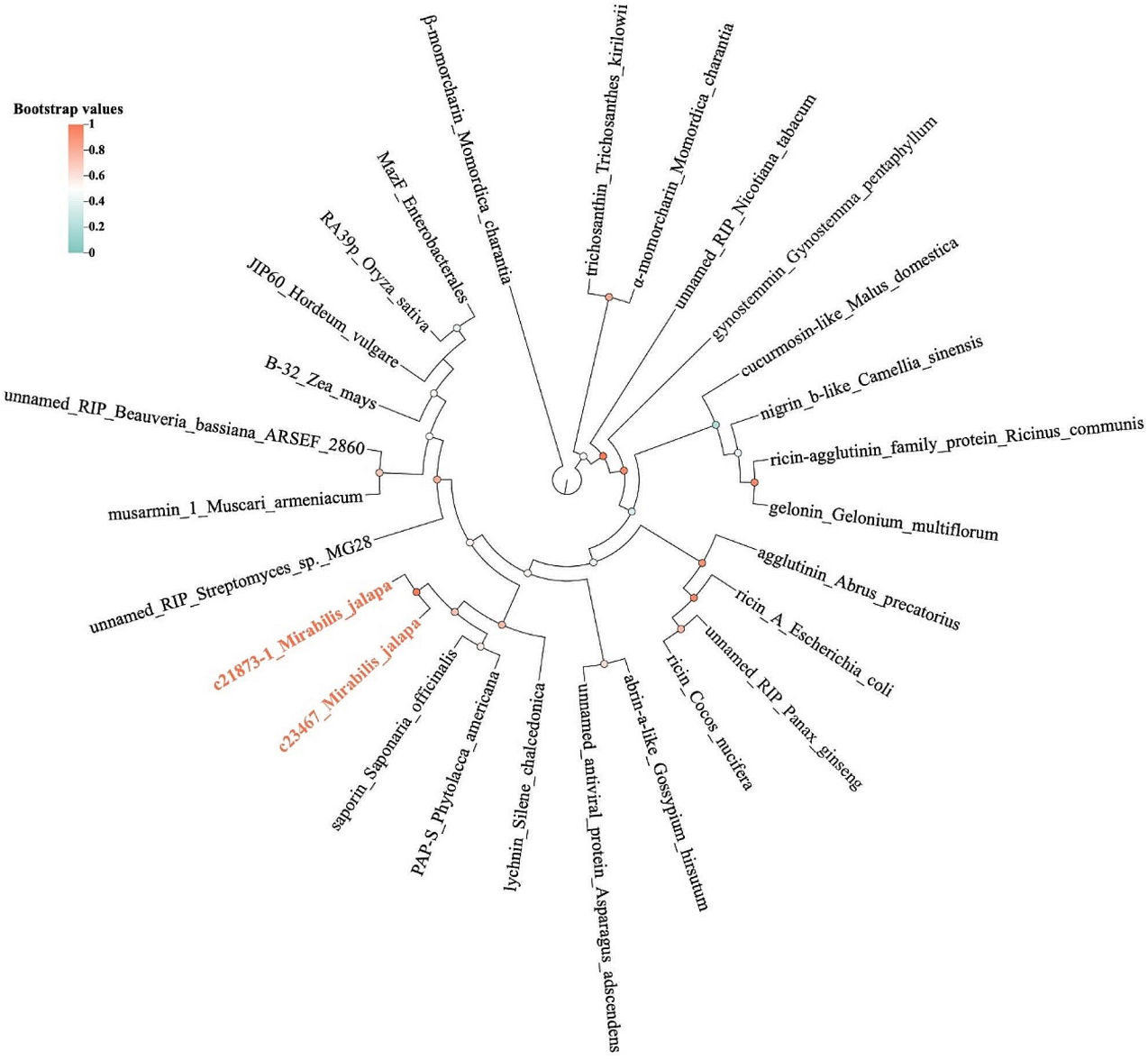

Fig. 3 The phylogenetic tree of c21873-1, c23467, twenty-four representative RIPs and MazF protein. By sequence blasting analysis with c21873-1 in NCBI database, c23467 and twenty-four representative RIPs were selected. Then, the amino acid sequences of c21873-1, c23467, twenty-four RIPs and MazF protein were aligned using Clustal W. The phylogenetic tree was constructed by MEGA11 software [41], using the neighbor-joining method with the amino acid substitution model and bootstrapping with 990 iterations. The tree was annotated and displayed using an online service software tvBOT [42]. The name, species and the accession numbers are listed in Table S1

cassette with a size of about 14 kb indicated the successful site-directed recombination of target gene into the homologous BmT1 locus. However, it failed when only about 2.5-kb DNA fragment was amplified (Fig. S1 and Fig. S2). It was speculated to be the reason of the random integration of the MAPs-GOI-Hyg B cassettes in other genomic loci. As shown in Table 1, site-target integration efficiency is the proportion of colonies bearing correct integration cassettes (about 14 kb). On the one hand, different promoters had different effects on the same selectable marker. When c23467 was used as the selectable marker, neither *pcsk2* nor *kex2* could inserted into the BmT1 locus in case of selecting P_{FDH1}, but the efficiency of HR was within 20% under the control of P_{DAS1} and P_{DAS2}. When c21873-1 was taken as the selectable marker, the efficiency of HR of both genes was 20% in case of using P_{FDH1}, and instead of which was 25–45% under the control of P_{DAS1} and P_{DAS2}. When c21873-1T was selected as the selectable marker, the efficiency of HR was 0 for *kex2* or 50% for *pcsk2* in case of choosing P_{FDH1}, of which was 15–35% under the control of P_{DAS1} and P_{DAS2}. As a result, when the selectable marker was

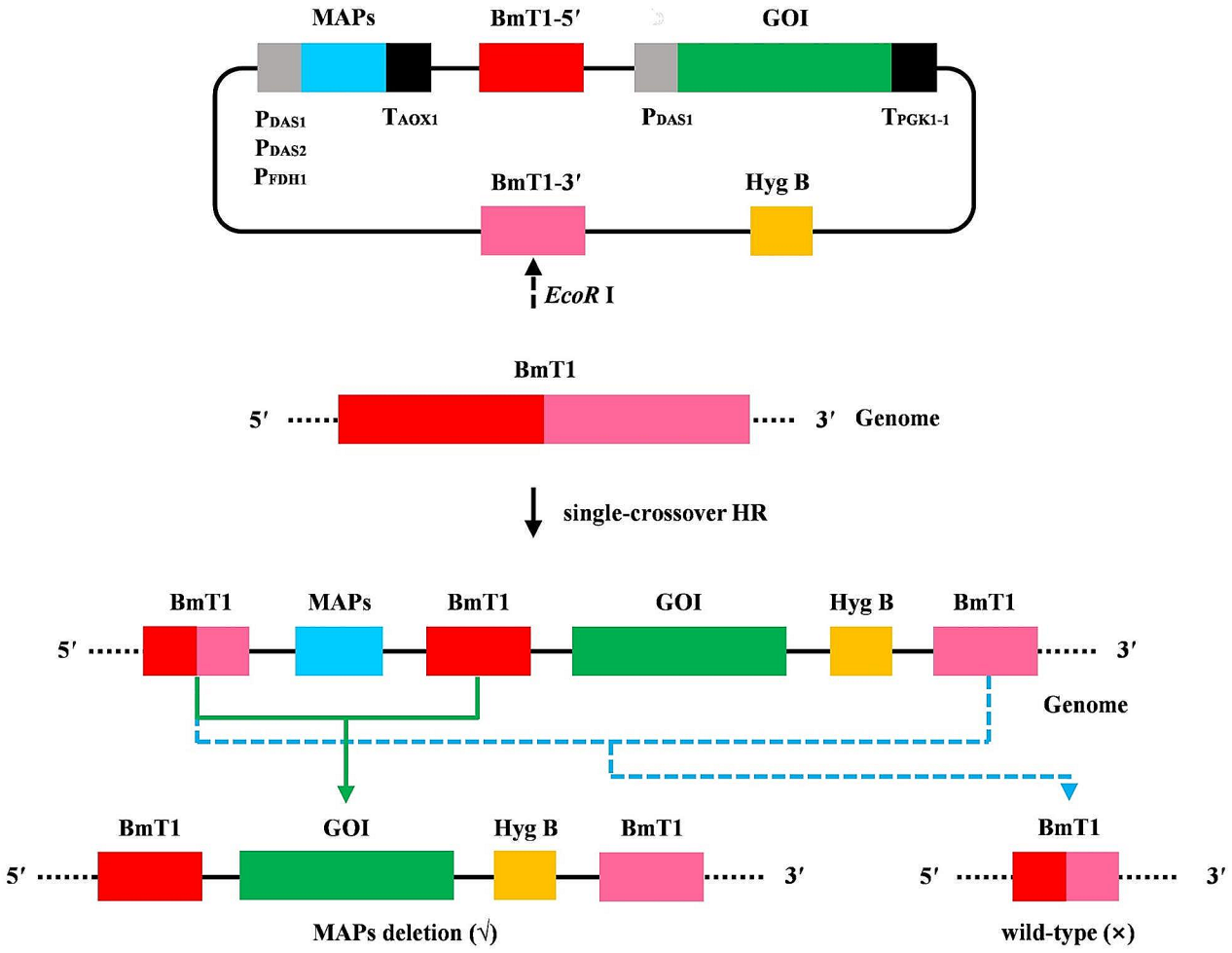

Fig. 4 Schematic diagram of application of the counter-selectable markers MAPs. The modular vector mainly contains the counter-selectable marker (MAPs: c21873-1, c21873-1T or c23467), the target gene or GOI, resistance gene (Hyg B) and BmT1 homologous region. After linearizing by restriction endonuclease *EcoR* I and transformation, the MAPs and GOI sequences are integrated into the *P. pastoris'* genome by single-crossover HR at the BmT1-3′ homologous region. To perform counter-selection, the positive colony was cultured in BMMY induction medium for three days. When the counter-selection was carried out, MAPs would be deleted by internal HR between BmT1-5′ or BmT1-3′ homologous fragments, generating MAPs deletion type at BmT1 locus or wild-type strain, respectively

confirmed, the leaky toxicity of P_{FDH1} was the strongest, which was not beneficial for gene modification due to the increase of the effort of screening. This result was in accordance with leaky expression of reporter gene *egfp* in *P. pastoris*. On the other hand, different selectable marker had different effects on efficiency of site-directed HR even using the same promoter. When selecting P_{FDH1} and the selectable marker c23467, the exogenous genes could not be integrated into the *P. pastoris'* genome, but using the selectable marker c21873-1, they had different efficiencies of site-target integration. When implementing P_{DAS1}, the efficiency of HR of exogeneous genes was 10–45% (c21873-1 > c21873-1T > c23467), of which P_{DAS2} was 5–45% (c21873-1, c21873-1T > c23467). All the results demonstrated that c21873-1 and c21873-1T were better candidates than c23467 as the selectable marker while utilizing the same promoter.

In further investigation, the counter-selection capability using different selectable markers under the control of P_{DAS1}, P_{FDH1} and P_{DAS2} were compared. Every single colony of correctly site-directed HR was picked up and inoculated into the culture medium under the induction of methanol for three days. The cells were spread on the YPD plates with corresponding antibiotics, counted and then identified by PCR amplification screening. The about 1.0-kb DNA fragment amplified indicated that the strains might occur the wrong deletion of MAPs (Fig. S3). Conversely, the 8.8-kb DNA fragment amplified showed the successful loss of genetic markers and integration of target sequences at the BmT1 locus in *P. pastoris'* genome (Fig. 5a). The amplified DNA fragments of positive colonies were then confirmed by the nucleotide sequencing (Fig. 5b). As shown in Table 2, under the control of P_{FDH1}, no yeast cells grew which might be the reason of

Table 1 The integration efficiencies of assembled MAPs-GOI-Hyg B cassettes with BmT1 homologous region

Name	Total colony NO. screened	Integration colony	Efficiency (%)[a]
P_{DAS1}-c23467-P_{DAS1}-*pcsk2*	20	4	20
P_{DAS1}-c23467-P_{DAS1}-*kex2*	20	2	10
P_{DAS1}-c21873-1-P_{DAS1}-*pcsk2*	20	7	35
P_{DAS1}-c21873-1-P_{DAS1}-*kex2*	20	9	45
P_{DAS1}-c21873-1T-P_{DAS1}-*pcsk2*	20	3	15
P_{DAS1}-c21873-1T-P_{DAS1}-*kex2*	20	5	25
P_{FDH1}-c23467-P_{DAS1}-*pcsk2*	19	0	0
P_{FDH1}-c23467-P_{DAS1}-*kex2*	20	0	0
P_{FDH1}-c21873-1-P_{DAS1}-*pcsk2*	20	4	20
P_{FDH1}-c21873-1-P_{DAS1}-*kex2*	20	4	20
P_{FDH1}-c21873-1T-P_{DAS1}-*pcsk2*	20	10	50
P_{FDH1}-c21873-1T-P_{DAS1}-*kex2*	20	0	0
P_{DAS2}-c23467-P_{DAS1}-*pcsk2*	20	1	5
P_{DAS2}-c23467-P_{DAS1}-*kex2*	20	3	15
P_{DAS2}-c21873-1-P_{DAS1}-*pcsk2*	20	9	45
P_{DAS2}-c21873-1-P_{DAS1}-*kex2*	20	5	25
P_{DAS2}-c21873-1T-P_{DAS1}-*pcsk2*	15	3	20
P_{DAS2}-c21873-1T-P_{DAS1}-*kex2*	20	7	35

a, After the site-directed integration of MAPs-GOI-Hyg B cassettes into the BmT1 locus, the positive colonies had the inserted DNA fragments of about 14 kb by analysis of PCR amplification. Then the integration efficiencies of every group were calculated via the number of the positive colonies divided by the number of the total screened colonies

its serious leakage expression, giving rise to the induction of the expression of MAPs. Under the control of P_{DAS1}, all the colonies might subject to the wrong deletion of MAPs since no bands with correct sizes were detected. When selecting P_{DAS2}, yeast cells could not grow or no colonies with the correct MAPs deletion were obtained using c23467 as the selectable marker. Under the same circumstance, however, the correct deletants using c21873-1or c21873-1T selection system were produced and verified by the DNA sequencing of PCR-amplified DNA fragments of the positive colonies. These results demonstrated that as the counter-selection marker, c21873-1 or c21873-1T was the better choice for seamless genetic modification due to the lower toxicity and higher integration efficiency.

The expression and analysis of EGFP in *P. pastoris*

During the genetic transformation process, the high efficiency of directed HR of heterologous gene at specific target site is of great importance, reducing the workload and cost of screening. At the protein expression stage, the fast and correct recycling of counter-selectable markers is also very important, which shortens the experimental period. Therefore, the expression of the reporter gene *egfp* was investigated to examine the effect of the most superior counter-selectable marker c21873-1 induced by P_{DAS2}, using the similar integration method mentioned above. To compare the expression level under the different circumstances, modular reporter proteins EGFP were expressed under the control of the inducible P_{AOX1} and the constitutive promoter GAP (P_{GAP}), with or without α-factor secretion signal, respectively. Similar to the site-target integration of the gene *kex2* constructed aforementioned, *egfp* could also be integrated into the BmT1 site, after popping-out c21873-1 marker. Consistent with most literature reported, EGFP was intracellularly expressed [10], which was detected in the cell pellet after ultrasonication in our experiment, as shown in Fig. 6a. Meanwhile, constitutive P_{GAP} was superior to inducible P_{AOX1} for the expression of EGFP. The fluorescence value of the experimental group under the control of P_{GAP} was significantly higher than the control group (GS115, $p<0.01$). However, the flanked α-factor secretion signal could not help the secretion of protein to extracellular medium (Fig. 6b). The ratio of positive cells expressing P_{AOX1}-EGFP was 12.7±2.6 (%), much less than that of P_{GAP}-EGFP of 52.6±27.6 (%) in flow cytometry (FCM) analysis (Fig. 7), which was consistent with the result in fluorescence detection.

Discussion

P. pastoris has been developed into an efficient workhorse for the production of gram amounts of heterologous proteins per liter, possessing great potential in pharmaceutical and industrial market [43, 44]. To realize the production, it is necessary to manipulate genetic modification of the yeast such as integration of heterologous genes, knock-in/knock-out the yield-relative target genes in the metabolic pathway, etc.

Among the genetic tools, the unmarked genetic modification by means of recyclable counter-selectable markers are valuable. MazF has been used as the counter-selectable marker in many kinds of organisms because its expression resulted in host cell growth arrest. In the early stage of this study, the toxicity of MazF killed all the *E. coli* cells. Therefore, the cloning procedure had to be interrupted. It was deduced to be the leaky expression of the eukaryotic promoters in prokaryotic systems. Not like prokaryotic expression vectors possessing promoters activated by an inducing agent, such as commonly used isopropyl β-D-1-thiogalactopyranoside (IPTG), the promoters of eukaryotic expression vectors are not under the control of active regulatory elements [45]. Therefore, the promoters of eukaryotic expression vectors are usually not functional in bacteria [46, 47]. Nevertheless, eukaryotic promoters and eukaryotic DNA are reported having a high probability of transcription initiation after transfer into bacteria [48]. It is because of the read-through effect mediated by some bacterial RNA polymerases, causing the leaky expression of recombination proteins to some extent at the cloning procedure. In our study, the result

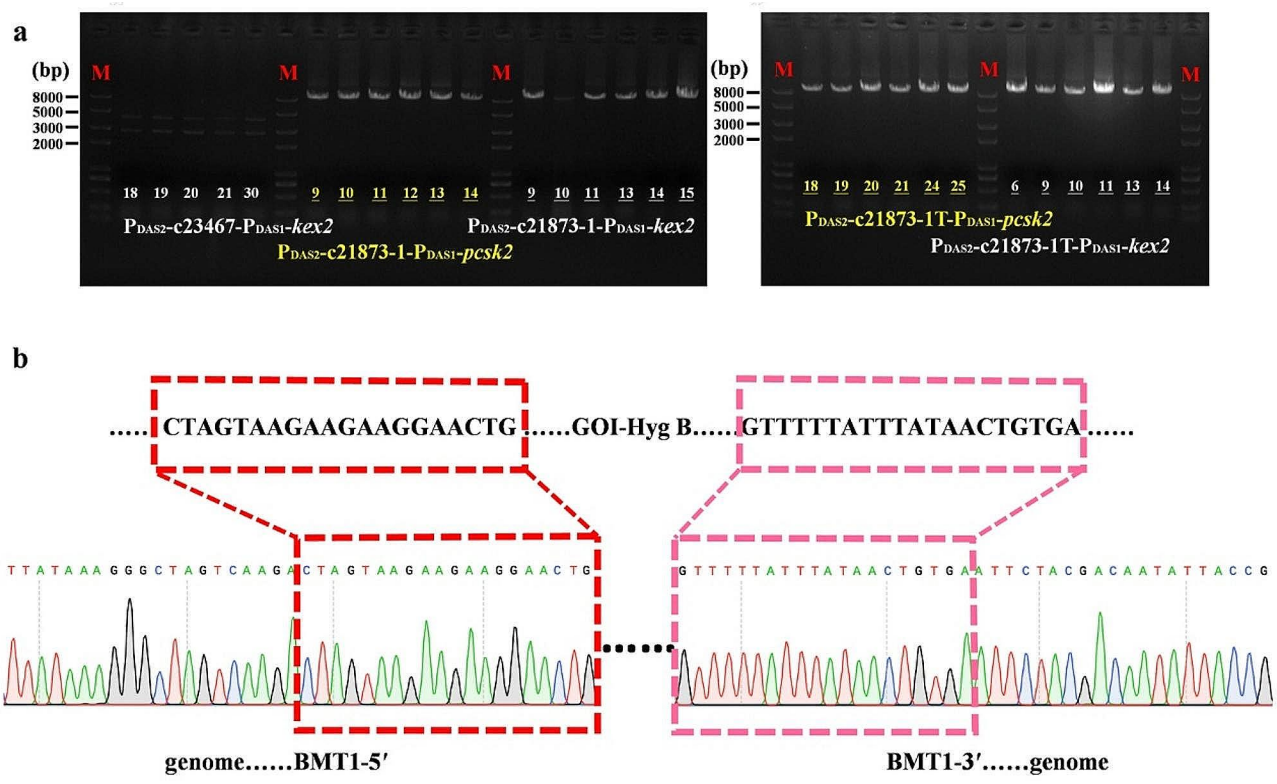

Fig. 5 HR of GOI (*pcsk2* or *kex2*) at the BmT1 locus. (**a**) The transformants confirmed by PCR amplification. After counter-selection, directed integration transformant presented an 8.8 kb of DNA band. M: DNA molecular weight markers. (**b**) Partial nucleotide sequencing of the colonies with correct deletion of selectable markers

Table 2 Comparison of counter-selection efficiencies of different MAPs

Name	Total colony NO. screened	Positive colony with expected deletion	Colony NO. verified by DNA Sequencing	Rescue ratio (%)[a]
P_{DAS1}-c23467-P_{DAS1}-*pcsk2*	30	2	2	0 (0/2)
P_{DAS1}-c23467-P_{DAS1}-*kex2*	30	11	3	0 (0/3)
P_{DAS1}-c21873-1-P_{DAS1}-*pcsk2*	30	14	3	33.3 (1/3)
P_{DAS1}-c21873-1-P_{DAS1}-*kex2*	30	29	3	0 (0/3)
P_{DAS1}-c21873-1T-P_{DAS1}-*pcsk2*[b]	-	-	-	-
P_{DAS1}-c21873-1T-P_{DAS1}-*kex2*	30	28	3	0 (0/3)
P_{FDH1}-c21873-1-P_{DAS1}-*pcsk2*[b]	-	-	-	-
P_{FDH1}-c21873-1-P_{DAS1}-*kex2* [b]	-	-	-	-
P_{FDH1}-c21873-1T-P_{DAS1}-*pcsk2*[b]	-	-	-	-
P_{DAS2}-c23467-P_{DAS1}-*pcsk2*[b]	-	-	-	-
P_{DAS2}-c23467-P_{DAS1}-*kex2*	30	8	5	0 (0/5)
P_{DAS2}-c21873-1-P_{DAS1}-*pcsk2*	30	22	6	100 (6/6)
P_{DAS2}-c21873-1-P_{DAS1}-*kex2*	30	16	6	100 (6/6)
P_{DAS2}-c21873-1T-P_{DAS1}-*pcsk2*	30	11	6	100 (6/6)
P_{DAS2}-c21873-1T-P_{DAS1}-*kex2*	30	20	6	100 (6/6)

a, to calculate the rescue ratio, the number of the positive colonies with expected deletion of MAPs was counted at first. Then the genomic DNAs of three positive colonies were isolated for DNA sequencing. For the colonies with correct deletion of selectable markers, another three positive colonies were picked up to repeat the procedure and perform the DNA sequencing, in order to guarantee its accuracy. The rescue ratio was calculated via the number of the colonies with the correct DNA sequence information divided by the number of the total screened colonies. **b**, due to the no growth of *P. pastoris* cells spread on the YPD-Hyg B (100 μg/mL) plates after the counter-selection, five groups were of no subsequent statistics

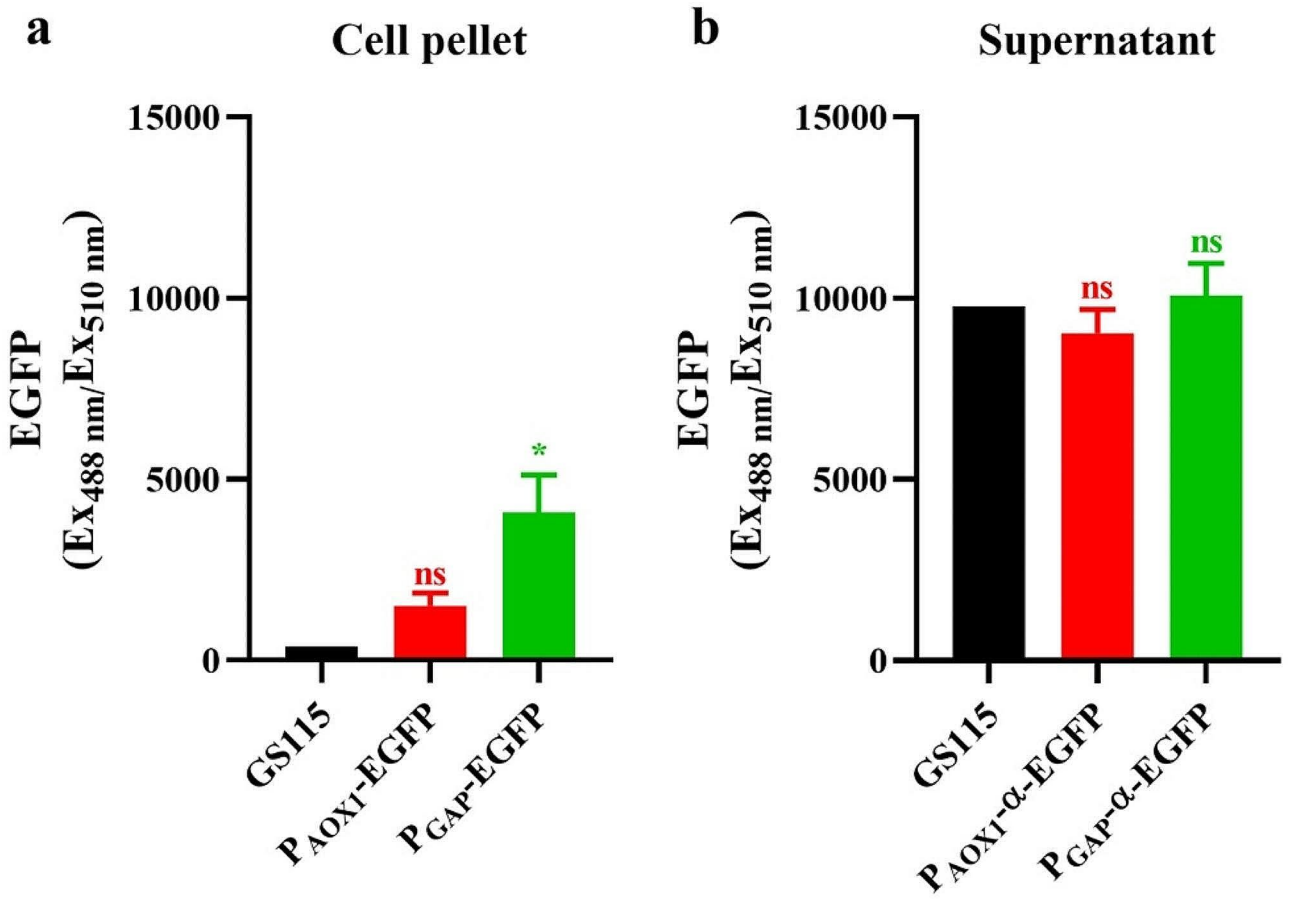

Fig. 6 The expression of EGFP in *P. pastoris*. The fluorescence of EGFP protein was detected in the cell pellet (**a**) and in the supernatant (**b**), respectively. The effects of promoters' type and with or without α-factor secretion signal on the expression and the secretion of EGFP were investigated. The significant differences by statistical analyses were compared between the experimental group and the control group (GS115), respectively

in bacteria experiment demonstrated that the leakage of P_{DAS1} was the most serious among six promoters with different strength (Fig. 1). Then the leaky expression of them in *P. pastoris* when cultured in YPD medium without inducing agent and in BMMY medium induced by methanol were studied, respectively. The P_{FDH1} showed the most serious leakage among them, while P_{DAS2} showed the weakest leaky expression but very strong induced expression (Fig. 2). It was not difficult to understand that the low leakage of eukaryotic promoters produced the unwanted proteins such as MazF which were noxious to the viability of bacteria, and cell growth arrest appeared when selecting correct constructs. Therefore, P_{DAS2} was an ideal promoter for strict expression regulation of toxic protein.

Then, it was urgent to construct a novel counter-selection system to resolve this problem. RIPs are a group of plant enzyme characterized by inactivating ribosomes and arresting protein synthesis. The characteristics of plant RIPs imply the possibility to act as a counter-selectable marker. A recently characterized MAP c21873-1 in this work has 49.17% of the identity with the reported MAP c23467 [37], both of which were alike type I RIPs through phylogenetic tree analysis (Table S1). Therefore, the capabilities of c21873-1, c21873-1T and c23467 to act as counter-selectable markers were further investigated. The results showed that this new strategy may overcome the problems encountered in the early stage of this study. First, none of them expressed in bacteria, not disturbing the cloning procedure. The HR of MAP-GOI-Hyg B cassette in the genome of *P. pastoris* did neither interfere with its growth under the normal condition without counter-selection stress. As reported, RIPs eliminate a single adenine residue located in the conserved GAGA-tetraloop present in 23S rRNA of the bacteria, inhibiting the protein translation [25]. However, except for direct cleavage of mRNA at specific ACA site, Bezrukov et al. found that MazF also indirectly decreased translation factors and promoted ribosome hibernation in *Staphylococcus aureus* [49]. It perhaps implies the stronger toxicity of MazF which has both effects on transcription and translation while RIPs only inhibit translation events. Actually,

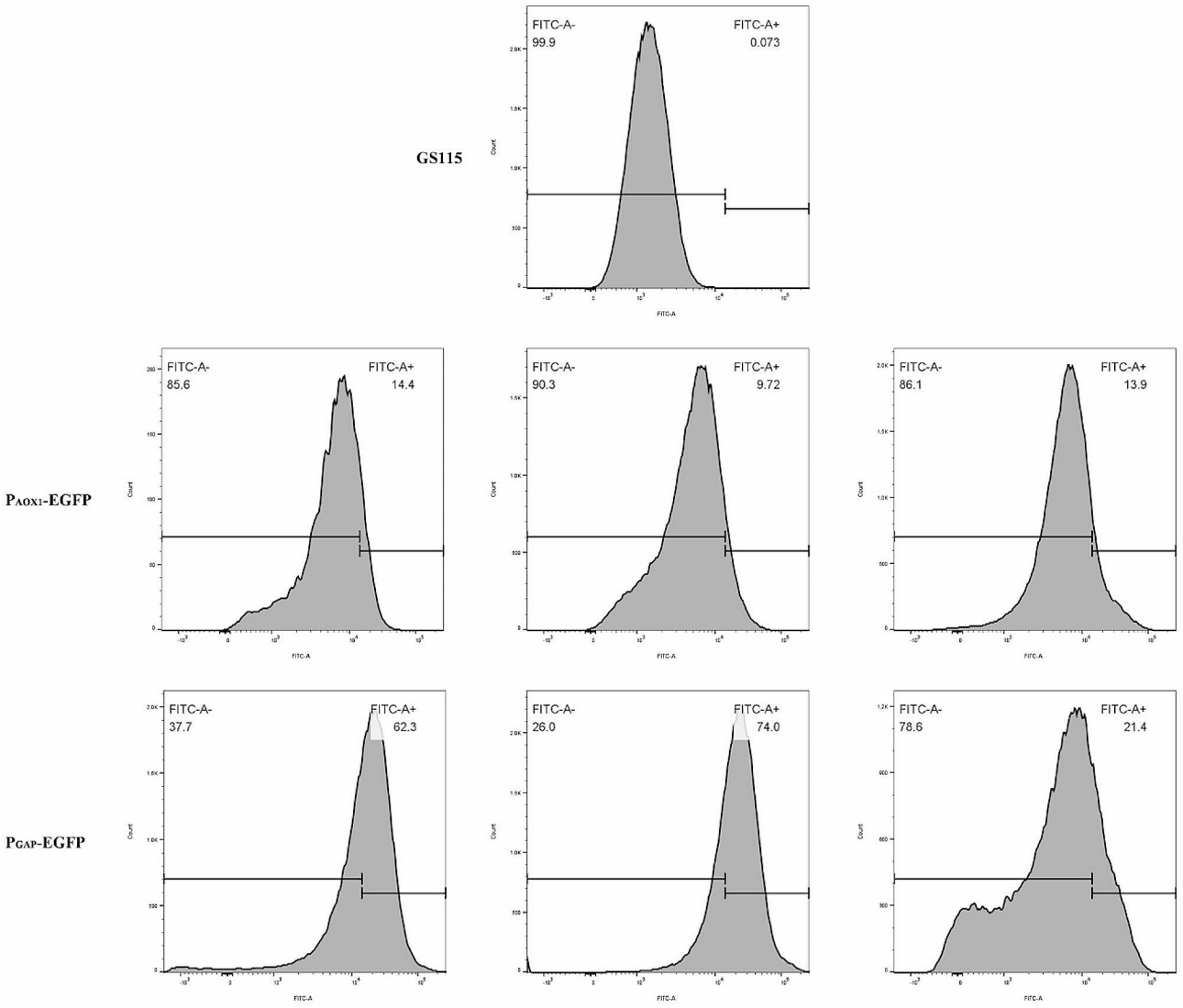

Fig. 7 Histogram of fluorescence for the yeast cells expressed EGFP. The whole population of the cells were harvested and the percentage of EGFP expression cells in *P. pastoris* was analyzed by FCM. The cells in the control group were considered as the negative cells having no fluorescence

in another research of our lab, only one positive colony of a thousand colonies with GOI was inserted into the genome before MazF recycling by methanol induction in *P. pastoris*, demonstrating its fierce virulence on cells (data not shown). Second, this system has the advantage of HR of long fragments comparing to other counter-selection systems. In previous studies, MazF was utilized as the counter-selectable marker only to realize the proof of concept since the integration fragments were not very long. Zhang et al. engineered *P. pastoris* for increasing *myo*-inositol production by replacement of the promoter *IPS* gene with P_{GAP} and deletion of two possible inositol transporters, in which the integration of "*mazF-zeocin*-short-arm" fragment was less than 8 kb [23]. Jiao et al. designed an improved method by directed HR in *P. pastoris* to delete *Och1* gene in which *mazF-zeocin* cassette was less than 5 kb [6]. Jiao et al. developed an efficient rhamnose-inducible promoter P_{LRA3}-based system for markerless deletion of *his4* gene in *P. pastoris* in which the expression cassette was about 4.46 kb [24]. However, using MAPs as the novel counter-selectable markers, the integration cassette (including BmT1 homologous region and MAP-GOI-Hyg B cassette) was as long as about 14 kb in this research. When using P_{DAS2} and the counter-selectable marker c21873-1 (or c21873-1T), the integration efficiency was 20–45% (Table 1). Third, we achieved the seamless modification in *P. pastoris* by means of getting the selection marker rescue. After counter-selection, the correct rescue ratios of c21873-1 and c21873-1T were 100% (Table 2). Marker-recycling and confirmation can be accomplished in 3–4 days if the integrated heterologous fragment is short, and in 4–5

days if it is long (>8000 bp fragment needs the extraction of the genome for further characterization). It saves both time and labor since the same procedure using the *URA3* marker takes a week or longer [11]. It also avoids the additional fragments introduced in the genome after marker recycling, leaving no scar which exists in the Cre/*lox*P system or FLP/*FRT* system. Although integration of the target genes (*pcsk2* and *kex2*) in the genome of *P. pastoris* was successful, unfortunately, no secretory expressions in extracellular medium were detected (data not shown). It was speculated to account for the inappropriate integration site which possibly affected the secretion signal, endoplasmic reticulum folding and Golgi apparatus transportation. Differently, the progress went well when some non-secretory proteins were also integrated at the BmT1 locus. The results of fluorescence detection (Fig. 6) and FCM analysis (Fig. 7) confirmed the intracellular expression of EGFP. Besides, under this circumstance, P_{GAP} was superior to P_{AOX1} for its expression. Two chaperone proteins, ERp46 and P4HB, also expressed successfully, for whose bands' sizes were consistent to their theoretical molecular weights (46.9 and 57.7 kDa, respectively) in western blot analysis (Fig. S4). All these results demonstrated the good applicability of this strategy.

Conclusions

In the present study, we established a new counter-selection system utilizing MAP c21873-1 as a novel counter-selectable marker. This system is simple, efficient and timesaving which can be applicable to the seamless integration by single-crossover HR in *P. pastoris*. This is the first time MAP c21873-1 from plant RIPs family reported as the counter-selectable marker, which might be exploited in other yeasts and higher eukaryotes. This unmarked genetic modification strategy may be expanded to other genetic modifications including but not limited to gene knock-out and site-directed mutagenesis in future.

Materials and methods

Strains, plasmids and culture conditions

The strains, plasmids and primers used in this research are listed in Table 3 and Table S2. *E. coli* DH10B was used

Table 3 Strains and plasmids used in the study

Strains or plasmids	Description	Source
E. coli DH10B	F$^-$*mcrA*Δ(*mrr-hsdRMS-mcrBC*) φ80*lacZ*ΔM15 Δ*lacX74 recA1 endA1 araD139*Δ (*ara-leu*)7697 *galU galK*λ–*rpsL*(Str^R) *nupG*	Lab storage
P. pastoris GS115	*his4*, *Mut*$^+$	Lab storage
pUG6-*egfp*	pUG6 derivative with gene *egfp*	Lab storage
pUG6-PARS2	pUG6 derivative with replicon PARS2	Lab storage
pUG6-P_{DAS1}-T_{PGK1}-SBP	pUG6 derivative with DAS1 promoter, α-factor, SBP-tag and PGK1 terminator	Lab storage
pδGAPh	pBluescript II KS(+) derivative with homologous δ region, GAP promoter, PGK1 terminator and HygB^r	Lab storage [50]
pδGAPh-P_{PXR1}-MazF	pδGAPh derivative with PXR1 promoter, gene *mazF*, and AOX1 terminator	This study
pδGAPh-Promoter-MazF-*egfp*-P_{FDH1}-*pcsk2*	pδGAPh-P_{PXR1}-MazF derivative with different promoters including AOX1, DAS1, DAS2, FDH1 or FLD1 promotes, and with gene *egfp*, FDH1 promoter and gene *pcsk2*	This study
pδGAPh-Promoter-*egfp*	pδGAPh-P_{PXR1}-MazF derivative with different promoters including AOX1, DAS1, DAS2, FDH1 or FLD1 promotes, and gene *egfp*	This study
pUG6-Promoter-*egfp*	pUG6-PARS2 derivative with different promoters including AOX1, DAS1, DAS2, FDH1, FLD1 or PXR1 promotes, and gene *egfp*	This study
pδGAPh-Promoter-MAP-P_{DAS1}-*pcsk2* (or *kex2*)	pδGAPh-Promoter-MAP derivative with a construct of DAS1 promoter, α-factor and gene *pcsk2* (or *kex2*); promoter selecting from DAS1, DAS2 or FDH1 promoters, and MAP selecting from c23467, c21873-1 or c21873-1T, respectively	This study
pδGAPh-P_{DAS2}-MAP c21873-1-P_{AOX1}-His6-*kex2*	pδGAPh-P_{DAS2}-MAP derivative with AOX1 promoter, α-factor and gene *kex2* with his-tag	This study
pδGAPh-P_{DAS2}-MAP c21873-1-P_{GAP}-His6-*kex2*	pδGAPh-P_{DAS2}-MAP derivative with GAP promoter, α-factor and gene *kex2* with his-tag	This study
pδGAPh-P_{DAS2}-MAP c21873-1-P_{AOX1}-α-factor-*egfp*	pδGAPh-P_{DAS2}-MAP derivative with AOX1 promoter, α-factor and gene *egfp*	This study
pδGAPh-P_{DAS2}-MAP c21873-1-P_{GAP}-α-factor-*egfp*	pδGAPh-P_{DAS2}-MAP derivative with GAP promoter, α-factor and gene *egfp*	This study
pδGAPh-P_{DAS2}-MAP c21873-1-P_{AOX1}-*egfp*	pδGAPh-P_{DAS2}-MAP derivative with AOX1 promoter and gene *egfp*	This study
pδGAPh-P_{DAS2}-MAP c21873-1-P_{GAP}-*egfp*	pδGAPh-P_{DAS2}-MAP derivative with GAP promoter and gene *egfp*	This study

for plasmid amplification and assayed for leaky expression of promoters, *P. pastoris* GS115 was used to test the leakage of promoters and general protein expression, respectively. KAPA HiFi Hotstart Readymix (Roche Diagnostics GmbH) was used for routine PCR amplification, and LongAmp Hot Start *Taq* 2× Master Mix (NEB, USA) was used for amplification of the fragments longer than 8 kb. The transformation of *P. pastoris* GS115 was performed on the electroporation device MicroPluser (Bio-Rad, USA) according to the instruction.

E. coli DH10B was cultured at 37 °C with agitation at 220 rpm in Luria-Bertani (LB) medium (5 g/L yeast extract, 10 g/L tryptone, and 10 g/L NaCl) supplemented with ampicillin (Amp, 100 μg/mL) where appropriate. For transformant screening, *P. pastoris* GS115 was cultured at 30 °C in yeast YPD medium (10 g/L yeast extract, 20 g/L peptone, and 20 g/L glucose). For protein expression, *P. pastoris* GS115 was cultured at 30 °C with agitation at 220 rpm in BMGY (10 g/L yeast extract, 20 g/L peptone, 10 g/L glycerin, 13.4 g/L YNB, 4×10^{-4} g/L biotin, and 0.1 M K_2HPO_4/KH_2PO_4; pH 6.0) or BMMY (5 mL methanol instead of glycerol in BMGY) medium. Geneticin (G418, 100 μg/mL) and Hyg B (100 μg/mL) were added where appropriate when *P. pastoris* GS115 was cultured.

Construction of EGFP expression plasmids under the control of different promoters in *E. coli*

To investigate the leaky expression of different promoters, *egfp* was taken as a reporter gene. Taking the construction of pδGAPh-P_{PXR1}-*egfp* as an example: at first, in order to delete the encoding region of MazF, the *egfp* and P_{PXR1} fragments were amplified by PCR with pδGAPh-P_{PXR1}-MazF-*egfp*-P_{FDH1}-*pcsk2* as the template, respectively. Then a P_{PXR1}-*egfp* gene construct was generated by overlap PCR amplification. Finally, it was subcloned into the pδGAPh-P_{PXR1}-MazF vector to obtain the resulting plasmid pδGAPh-P_{PXR1}-*egfp*. pδGAPh-P_{AOX1}-*egfp*, pδGAPh-P_{DAS1}-*egfp*, pδGAPh-P_{DAS2}-*egfp*, pδGAPh-P_{FDH1}-*egfp*, and pδGAPh-P_{FLD1}-*egfp* were also constructed only instead of corresponding promoters.

Detection of the leaky expression of EGFP under the different promoters in *E. Coli*

100 ng pδGAPh-Promoter-*egfp* (promoter including AOX1, DAS1, DAS2, FDH1, FLD1 or PXR1) was transformed into *E. coli* DH10B competent cells. The cells were inoculated in 10 mL of LB medium supplemented with 100 μg/mL Amp in 100-mL shake flask, and incubated at 37 °C for 24 h with agitation at 220 rpm. The OD_{600} value of the bacteria was monitored. When the OD_{600} reached to 8.0, the cells were collected and centrifugated at 12,000 rpm for 5 min at 4 °C. The supernatant was collected and 200 μL of it was taken out for fluorescence detection at $Ex_{488\ nm}/Em_{510\ nm}$. The cell pellet was washed twice with sterile ddH_2O to remove the cell debris. Then it was resuspended in 220 μL of PBS buffer (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 2 mM KH_2PO_4; pH 7.4), and the intracellular protein was obtained by ultrasonication in ice bath for 90 s (3 s on, 3 s off). After centrifugation at 12,000 rpm for 5 min at 4 °C, 200 μL of cell-free sample was also taken out for fluorescence detection at $Ex_{488\ nm}/Em_{510\ nm}$.

Construction of EGFP expression plasmids under the control of different promoters in *P. pastoris*

To investigate the leaky expression of EGFP in *P. pastoris*, the plasmids under the control of different promoters were constructed. To create pUG6-P_{FDH1}-*egfp* construct, the FDH1-*egfp* fragment was amplified by PCR from pδGAPh-P_{FDH1}-*egfp*. Then it was subcloned into pUG6-PARS2, and designated as pUG6-P_{FDH1}-*egfp*. The other plasmids pUG6-Promoter-*egfp* were obtained with the similar method only instead of corresponding promoters.

The leaky expression of EGFP under the control of different promoters in *P. pastoris*

pUG6-Promoter-*egfp* (promoter including P_{AOX1}, P_{DAS1}, P_{DAS2}, P_{FDH1}, P_{FLD1} or P_{PXR1}) was transformed into *P. pastoris* GS115 by electroporation, and the cells were spread on the YPD plates containing G418 (100 μg/mL). The positive colonies were confirmed by colony PCR identification.

The positive colony was inoculated into 5 mL of BMGY medium containing G418 (100 μg/mL) and cultured for 24 h at 30 °C and 220 rpm. The culture was then inoculated in a 100-mL shake flask with an initial OD_{600} of 1.0 in 20 mL of BMMY medium containing 0.5% (v/v) methanol and G418 (100 μg/mL). They were incubated for another 120 h at 30 °C and 250 rpm, and 1% (v/v) methanol as the carbon source and protein expression inducer was added every 24 h. To measure the fluorescence, the cells of OD_{600} values of 8.0 were harvested and centrifugated at 12,000 rpm for 10 min at 4 °C. The supernatant was collected and 200 μL of it was taken out for fluorescence detection at $Ex_{488\ nm}/Em_{510\ nm}$. The cell pellet was rinsed once with sterile ddH_2O and resuspended with 500 μL of 1 M sorbitol-PBS buffer (pH 7.4), followed by digestion with addition of 8 μL of lyticase for 30 min at 30 °C and 240 rpm. The cells were centrifugated again for 10 min at 4 °C and 12,000 rpm, and the supernatant was discarded. The cell pellet was resuspended with 220 μL of PBS buffer (pH 7.4) and ultrasonicated in ice bath for 90 s (3 s on, 3 s off). After centrifugation at 12,000 rpm for 5 min at 4 °C, 200 μL of cell-free sample was taken out for fluorescence detection at $Ex_{488\ nm}/Em_{510\ nm}$.

Construction of plasmids containing *pcsk2* and *kex2* as experimental target genes with counter-selectable markers MAPs under the control of different promoters in *P. pastoris*

The genes of MAPs (c21873-1, c21873-1T and c23467) and target genes (*pcsk2* and *kex2*) were synthesized by Generay Biotech Co., Ltd (Shanghai, China). To construct pδGAPh-Promoter-MAPs vectors, an expression construct containing $P_{DAS2,}$ T_{AOX1} and c21873-1 (c21873-1T or c23467) was generated by overlapping PCR amplification, and then subcloned into pδGAPh-P_{DAS2}-*egfp* vector, the resulting plasmid was designated as pδGAPh-P_{DAS2}-c21873-1 (pδGAPh-P_{DAS2}-c21873-1T or pδGAPh-P_{DAS2}-c23467). The other plasmids pδGAPh-P_{DAS1}-c21873-1 (c21873-1T or c23467) and pδGAPh-P_{FDH1}-c21873-1 (c21873-1T or c23467) of this set of pδGAPh-Promoter-MAPs were constructed via replacing P_{DAS2} with P_{DAS1} and P_{FDH1}, respectively. Next, to construct the plasmids pδGAPh-Promoter-MAP-P_{DAS1}-target genes, another expression construct having P_{DAS1} and *pcsk2* (or *kex2*) was prepared as aforementioned and then inserted into the set plasmids of pδGAPh-Promoter-MAPs.

Screening and characterization of site-directed HR of target genes

pδGAPh-Promoter-MAP-P_{DAS1}-*pcsk2* (or *kex2*) were transformed into *P. pastoris* GS115 by electroporation, and the cells were spread on the YPD plates containing Hyg B (100 μg/mL). After 3 to 4 days, twenty colonies were picked up and reinoculated on new YPD plates containing Hyg B with the same concentration. The colonies were characterized by colony PCR analysis. The amplified DNA fragments of the negative colonies without destruction of BmT1 homologous fragment or with random integration of MAPs-GOI-Hyg B cassettes in other genomic loci were 2.5 kb. The positive colonies of site-directed HR had the inserted DNA fragments of about 14 kb.

The positive colonies were inoculated into 5 mL of BMGY medium containing Hyg B (100 μg/mL) and cultured for 24 h at 30 °C and 220 rpm. The culture was then inoculated in a 100-mL shake flask with an initial OD_{600} of 1.0 in 20 mL of BMMY medium containing Hyg B (100 μg/mL). They were incubated for another 72 h at 30 °C and 220 rpm, and 1% (v/v) methanol was added every 24 h. The cells were counted and one hundred of the cells were spread on the YPD plates containing Hyg B (100 μg/mL) after serial dilution. When accomplishing counter-selection, the MAPs would be excised and recovered by internal HR. Therefore, twenty single colonies of each group were randomly selected and subjected to characterization of the integration by colony PCR amplification. If no unexpected DNA band was detected, it was speculated that the MAPs might be removed. For further verification, the genome was extracted to perform PCR amplification and DNA sequencing. The appearance of about 8.8-kb band proved that correct HR had been realized.

Construction of EGFP expression cassettes with MAP c21873-1 as a counter-selectable marker under the control of different promoters in *P. pastoris*

A construct harboring P_{AOX1}-α-factor or P_{GAP}-α-factor and *egfp* was gotten by overlapping PCR amplification and then used to replace the P_{AOX1}-His6-*kex2* element of the pδGAPh-P_{DAS2}-MAP c21873-1-P_{AOX1}-His6-*kex2* vector, generating the resulting plasmid pδGAPh-P_{DAS2}-MAP c21873-1-P_{AOX1}-α-factor-*egfp* or pδGAPh-P_{DAS2}-MAP c21873-1-P_{GAP}-α-factor-*egfp*. Following this method, EGFP expression vectors of pδGAPh-P_{DAS2}-MAP c21873-1-P_{AOX1}-*egfp* or pδGAPh-P_{DAS2}-MAP c21873-1-P_{GAP}-*egfp* without the α-factor DNA element were also constructed.

The expression and analysis of EGFP in *P. pastoris*

The procedures of expression of EGFP with MAP c21873-1 as the counter-selectable marker and confirmation of deletion of MAP c21873-1 were the same as those of target genes in *P. pastoris* GS115. The expression of EGFP in the supernatant and the cell pellet were analyzed by fluorescence detection at $Ex_{488\ nm}/Em_{510\ nm}$. Furthermore, 10^5 cells were harvested and the whole population of them were analyzed on flow cytometer BD FACSCanto II (BD Biosciences, San Jose, CA, USA). The positive cells with the expression of EGFP were determined using single platform method on FlowJo™ 10.8 software (BD Biosciences, San Jose, CA, USA).

Statistical analysis

Each experiment was performed with triplicate samples, and the data was expressed as mean ± sd. Statistical analyses and data visualizing of growth curves of *P. pastoris* and fluorescence detection of EGFP were performed using GraphPad Prism 8.0 software (GraphPad Software, CA, USA). The comparison between the experimental group and the control group was analyzed by one-way analysis of variance (ANOVA) followed by Brown-Forsythe test, or the nonparametric ANOVA, depending on the experiments. Differences were considered statistically significant with values of $p < 0.05$.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12934-024-02496-w.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Acknowledgements
Not applicable.

Author contributions
Minzhi Liu and Sihan Zhou conceptualized the study and performed the experiments. Yunsong Cao analyzed the data. Keqin Yang and Yao Xiao drew the figures. Minzhi Liu drafted the manuscript. Sihan Zhou revised the manuscript. Wei Wang coordinated the project, and provided support and guidance. All authors read and approved the final manuscript.

Funding
This work was supported by the National Natural Science Foundation of China (No. 82073757) and Chinese Academy of Medical Sciences Initiative for Innovative Medicine (No.2021-I2M-1-029).

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Institute of Materia Medica, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100050, China
[2]Key Laboratory of Biosynthesis of Natural Products of National Health Commission of the Peoples Republic of China, Institute of Materia Medica, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100050, China

Received: 18 April 2024 / Accepted: 30 July 2024
Published online: 08 August 2024

References

1. Cereghino JL, Cregg JM. Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. FEMS Microbiol Rev. 2000;24:45–66.
2. Safder I, Khan S, Islam IU, Kazim M, Bibi Z, Waqas M. *Pichia pastoris* expression system: a potential candidate to express protein in industrial and biopharmaceutical domains. Biomed Lett. 2018;4:1–14.
3. Thompson CA. FDA approves kallikrein inhibitor to treat hereditary angioedema. Am J Health Syst Pharm. 2010;67:93–93.
4. Mullard A. 2012 FDA drug approvals. Nat Rev Drug Discov. 2013;12:87–90.
5. Syed YY, Dhillon S, Ocriplasmin. A review of its use in patients with symptomatic vitreomacular adhesion. Drugs. 2013;73:1617–25.
6. Jiao L, Zhou Q, Yan Y. Efficient gene disruption by posttransformational directed internal homologous recombination in *Pichia pastoris*. Anal Biochem. 2019;576:1–4.
7. Shibui T, Hara H. A new type of gene-disruption cassette with a rescue gene for *Pichia pastoris*. Biotechnol Prog. 2017;33:1201–08.
8. Han M, Wang W, Gong X, Zhu G, Liu X, Yu Z, Zhou J, Ma C, Ma X. A modified method of gene disruption in *Komagataella phaffii* with Cre/loxP system. J Biotechnol. 2022;347:40–8.
9. Ahmad M, Winkler CM, Kolmbauer M, Pichler H, Schwab H, Emmerstorfer-Augustin A. *Pichia pastoris* protease-deficient and auxotrophic strains generated by a novel, user-friendly vector toolbox for gene deletion. Yeast. 2019;36:557–70.
10. Yang J, Jiang W, Yang S. *mazF* as a counter-selectable marker for unmarked genetic modification of *Pichia pastoris*. FEMS Yeast Res. 2009;9:600–09.
11. Dave MN, Chattoo BB. A counter-selectable marker for genetic transformation of the yeast *Schwanniomyces Alluvius*. Appl Microbiol Biotechnol. 1997;48:204–07.
12. Lu P, Davis BP, Hendrick J, Jeffries TW. Cloning and disruption of the beta-isopropylmalate dehydrogenase gene (*LEU2*) of *Pichia stipitis* with *URA3* and recovery of the double auxotroph. Appl Microbiol Biotechnol. 1998;49:141–46.
13. Lin Cereghino GP, Lin Cereghino J, Sunga AJ, Johnson MA, Lim M, Gleeson MA, Cregg JM. New selectable marker/auxotrophic host strain combinations for molecular genetic manipulation of *Pichia pastoris*. Gene. 2001;263:159–69.
14. Boeke JD, Trueheart J, Natsoulis G, Fink GR. [10] 5-Fluoroorotic acid as a selective Agent in yeast Molecular Genetics. Methods Enzymol. 1987;154:164–75.
15. Aizenman E, Engelberg-Kulka H, Glaser G. An *Escherichia coli* chromosomal addiction module regulated by guanosine [corrected] 3',5'-bispyrophosphate: a model for programmed bacterial cell death. Proc Natl Acad Sci U S A. 1996;93:6059–63.
16. Yamaguchi Y, Inouye M. Regulation of growth and death in *Escherichia coli* by toxin–antitoxin systems. Nat Rev Microbiol. 2011;9:779–90.
17. Zhang Y, Zhang J, Hoeflich KP, Ikura M, Qing G, Inouye M. MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in *Escherichia coli*. Mol Cell. 2003;12:913–23.
18. Kolodkin-Gal I, Hazan R, Gaathon A, Carmeli S, Engelberg-Kulka H. A Linear Pentapeptide is a quorum-sensing factor required for *mazEF*-Mediated cell death in *Escherichia coli*. Science. 2007;318:652–55.
19. Zhang XZ, Yan X, Cui ZL, Hong Q, Li SP. *mazF*, a novel counter-selectable marker for unmarked chromosomal manipulation in *Bacillus subtilis*. Nucleic Acids Res. 2006; 34:e71.
20. Zhao X, Xu J, Tan M, Yu Z, Yang S, Zheng H, Song H. Construction of a plasmid interspecific transfer system in *Bacillus* species with the counter-selectable marker *mazF*. J Ind Microbiol Biotechnol. 2018;45:417–28.
21. Liu Q, Wu Y, Yang P, Zhang X, Bai Y, Xu H, Qiao M. *Mazf*-mediated deletion system for large-scale genome engineering in *Saccharomyces cerevisiae*. Res Microbiol. 2014;165:836–40.
22. Song P, Liu S, Guo X, Bai X, He X, Zhang B. Scarless gene deletion in methylotrophic *Hansenula polymorpha* by using *mazF* as counter-selectable marker. Anal Biochem. 2015;468:66–74.
23. Zhang Q, Wang X, Luo H, Wang Y, Wang Y, Tu T, Qin X, Su X, Huang H, Yao B, et al. Metabolic engineering of *Pichia pastoris* for *myo*-inositol production by dynamic regulation of central metabolism. Microb Cell Fact. 2022;21:112.
24. Jiao J, Wang S, Liang M, Zhang Y, Xu X, Zhang W, Liu B. Basal transcription profiles of the rhamnose-inducible promoter P_{LRA3} and the development of efficient P_{LRA3}-based systems for markerless gene deletion and a mutant library in *Pichia pastoris*. Curr Genet. 2019;65:785–98.
25. Fabbrini MS, Katayama M, Nakase I, Vago R. Plant ribosome-inactivating proteins: Progesses, challenges and Biotechnological Applications (and a few digressions). Toxins (Basel). 2017;9:314.
26. Lapadula WJ, Ayub MJ. Ribosome inactivating proteins from an evolutionary perspective. Toxicon. 2017;136:6–14.
27. Lapadula WJ, Marcet PL, Mascotti ML, Sanchez-Puerta MV, Juri Ayub M. Metazoan Ribosome inactivating protein encoding genes acquired by Horizontal Gene Transfer. Sci Rep. 2017;7:1863.
28. Uckun FM, Bellomy K, O'Neill K, Messinger Y, Johnson T, Chen CL. Toxicity, biological activity, and pharmacokinetics of TXU (anti-CD7)-pokeweed antiviral protein in chimpanzees and adult patients infected with human immunodeficiency virus. J Pharmacol Exp Ther. 1999;291:1301–7.
29. Kuroda K, Liu H, Kim S, Guo M, Navarro V, Bander NH. Saporin toxin-conjugated monoclonal antibody targeting prostate-specific membrane antigen has potent anticancer activity. Prostate. 2010;70:1286–94.
30. Furman RR, Grossbard ML, Johnson JL, Pecora AL, Cassileth PA, Jung S-H, Peterson BA, Nadler LM, Freedman A, Bayer R-L, et al. A phase III study of anti-B4-blocked ricin as adjuvant therapy post-autologous bone marrow transplant: CALGB 9254. Leuk Lymphoma. 2011;52:587–96.
31. Puri M, Kaur I, Perugini MA, Gupta RC. Ribosome-inactivating proteins: current status and biomedical applications. Drug Discov Today. 2012;17:774–83.
32. Zhu F, Zhou YK, Ji ZL, Chen XR. The plant ribosome-inactivating proteins play important roles in defense against pathogens and insect pest attacks. Front Plant Sci. 2018;9:146.
33. Citores L, Iglesias R, Ferreras JM. Antiviral activity of ribosome-inactivating proteins. Toxins. 2021;13:80.

34. Byers VS, Levin AS, Waites LA, Starrett BA, Mayer RA, Clegg JA, Price MR, Robins RA, Delaney M, Baldwin RW. A phase I/II study of trichosanthin treatment of HIV disease. Aids. 1990;4:1189–96.
35. Borthakur G, Rosenblum MG, Talpaz M, Daver N, Ravandi F, Faderl S, Freireich EJ, Kadia T, Garcia-Manero G, Kantarjian H, et al. Phase 1 study of an anti-CD33 immunotoxin, humanized monoclonal antibody M195 conjugated to recombinant gelonin (HUM-195/rGEL), in patients with advanced myeloid malignancies. Haematologica. 2013;98:217–21.
36. Lu JQ, Zhu ZN, Zheng YT, Shaw PC. Engineering of Ribosome-inactivating proteins for improving Pharmacological properties. Toxins. 2020;12:167.
37. Kataoka J, Habuka N, Furuno M, Miyano M, Takanami Y, Koiwai A. DNA sequence of *Mirabilis* antiviral protein (MAP), a ribosome-inactivating protein with an antiviral property, from *mirabilis jalapa* L. and its expression in *Escherichia coli*. J Biol Chem. 1991;266:8426–30.
38. Zhang J, Zhang Y, Inouye M. Characterization of the interactions within the *mazEF* Addiction Module of *Escherichia coli*. J Biol Chem. 2003;278:32300–06.
39. Kamada K, Hanaoka F, Burley SK. Crystal structure of the MazE/MazF complex: molecular bases of antidote-toxin recognition. Mol Cell. 2003;11:875–84.
40. Vogl T, Sturmberger L, Kickenweiz T, Wasmayer R, Schmid C, Hatzl AM, Gerstmann MA, Pitzer J, Wagner M, Thallinger GG, et al. A toolbox of Diverse promoters related to methanol utilization: functionally verified parts for Heterologous Pathway expression in *Pichia pastoris*. ACS Synth Biol. 2016;5:172–86.
41. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol. 2021;38:3022–27.
42. Xie J, Chen Y, Cai G, Cai R, Hu Z, Wang H. Tree visualization by one table (tvBOT): a web application for visualizing, modifying and annotating phylogenetic trees. Nucleic Acids Res. 2023;51:W587–92.
43. Duman-Özdamar ZE, Binay B. Production of industrial enzymes via *Pichia pastoris* as a cell factory in Bioreactor: current status and future aspects. Protein J. 2021;40:367–76.
44. De S, Mattanovich D, Ferrer P, Gasser B. Established tools and emerging trends for the production of recombinant proteins and metabolites in *Pichia pastoris*. Essays Biochem. 2021;65:293–307.
45. Krela R, Poreba E, Weglewska M, Skrzypczak T, Lesniewicz K. A novel method for cloning of coding sequences of highly toxic proteins. Biochim Biophys Acta Gen Subj. 2019;1863:521–27.
46. Nielsen KM, Bones AM, Smalla K, van Elsas JD. Horizontal gene transfer from transgenic plants to terrestrial bacteria – a rare event? FEMS Microbio Rev. 1998;22:79–103.
47. Bertolla F, Simonet P. Horizontal gene transfers in the environment: natural transformation as a putative process for gene transfers between transgenic plants and microorganisms. Res Microbiol. 1999;150:375–84.
48. Lewin A, Mayer M, Chusainow J, Jacob D, Appel B. Viral promoters can initiate expression of toxin genes introduced into *Escherichia coli*. BMC Biotechnol. 2005;5:19.
49. Bezrukov F, Prados J, Renzoni A, Panasenko OO. MazF toxin causes alterations in *Staphylococcus aureus* transcriptome, translatome and proteome that underlie bacterial dormancy. Nucleic Acids Res. 2021;49:2085–101.
50. Tang L, Wang W, Zhou W, Cheng K, Yang Y, Liu M, Cheng K, Wang W. Three-pathway combination for glutathione biosynthesis in *Saccharomyces cerevisiae*. Microb Cell Fact. 2015;14:139.

Publisher's Note