

Article

Genome-Wide Landscape of Position Effect Variegation in *Pichia pastoris*

Xueyan Yao ^{1,2}, Baosong Chen ³, Zhihui Zhang ¹, Wenyan Gao ¹, Qian Zhang ¹, Yongpeng Yao ¹, Yuanyuan Pan ¹ and Gang Liu ^{1,4,*} 

- ¹ State Key Laboratory of Microbial Diversity and Innovative Utilization, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China; xueyan.yao@pku-iaas.edu.cn (X.Y.); 19816322213@163.com (Z.Z.); gaowu@im.ac.cn (W.G.); zhangqian226@mails.ucas.ac.cn (Q.Z.); yaoyu@im.ac.cn (Y.Y.); panyu@im.ac.cn (Y.P.)
- ² State Key Laboratory of Wheat Improvement, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Weifang 261325, China
- ³ College of Bioengineering, Tianjin University of Science and Technology, Tianjin 300457, China; chenbs@im.ac.cn
- ⁴ University of Chinese Academy of Sciences, Beijing 100049, China
- * Correspondence: liug@im.ac.cn; Tel.: +86-010-64806017

Abstract

Exogenous genes are generally expressed by integration into the chromosomes of *Pichia pastoris*. However, systematic studies on the chromosomal position effect are lacking, and locations that are conducive to the high expression of foreign genes are rarely reported. In this study, a genomic random insertion mutagenesis library for *P. pastoris* was successfully constructed using the *piggyBac* (PB) transposon system. Through sequencing, the sequence TTAA was identified as the major recognition site of the PB transposon, which exhibited relatively high coverage on *P. pastoris* chromosomes, making it a valuable tool for studying position effect variegation in *P. pastoris*. Using the enhanced green fluorescent protein gene (*eGFP*) as a reporter, two libraries including low-expression positions and high-expression positions were obtained by flow cytometry. The low-expression sites were mainly located upstream of ORFs around the promoter region and downstream near the terminator region, while the high-expression sites were predominantly located at the gene interior. KEGG and GO analyses showed that genes in high-expression positions were significantly enriched in the ATP-dependent chromatin remodeling and histone binding pathways, and genes in low-expression positions were significantly enriched in the MAPK signaling pathway, autophagy, mitochondrial autophagy, ABC transporters, and the arginine synthesis pathway. This study has clarified the genome-wide landscape of position effect variegation in *P. pastoris*. Additionally, it has provided novel insights into high-throughput screening strategies for strains with high exogenous gene expression.

Keywords: *Pichia pastoris*; *piggyBac* transposon; position effect variegation



Academic Editor: Aaron Neiman

Received: 11 December 2025

Revised: 5 February 2026

Accepted: 11 February 2026

Published: 12 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Pichia pastoris (also known as *Komagataella phaffii*) is a single-cell eukaryotic microorganism that can grow using methanol as the sole carbon and energy source [1]. Due to its convenient genetic manipulation, simple culture requirements, and high-density fermentation, *P. pastoris* is one of the most widely used eukaryotic expression systems for producing foreign proteins [2–5]. The absence of endotoxins and endogenous viruses makes *P. pastoris* suitable for food-grade production. It has been granted Generally Recognized As Safe

(GRAS) status by the U.S. FDA and is also approved for food use by the Chinese National Medical Products Administration. The success of the *P. pastoris* expression system is mainly attributed to its diverse inducible and constitutive promoters, high-density fermentation on specific media, efficient secretion, and capacity to undergo complex post-translational modifications, such as proper protein folding, disulfide bond formation, and glycosylation. However, the expression of exogenous genes in *P. pastoris* is often influenced by multiple factors, including promoter selection, codon usage, gene dosages, protein secretion pathways, and so on [6]. Among these factors, the integration site of the exogenous gene in the *P. pastoris* genome is a critical determinant. This phenomenon, known as the position effect, has been observed in various organisms and can lead to substantial variations in foreign protein production.

The position effect has been reported in *Escherichia coli*, *Salmonella typhimurium*, *Bacillus subtilis*, *Lactobacillus lactis*, *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and mouse and human cells [7–17]. A classic example of the position effect is the translocation of the white gene in *Drosophila melanogaster* into heterochromatin, resulting in the mottled appearance of the originally pure red eyes with white and red patches [18]. Recent genome-wide studies have provided more mechanistic details of the position effect from multiple aspects, including the regulatory role of enhancers, gene order, various epigenetic modifications, chromatin domains, and 3D positioning [7,9,10,19]. The position effect has significant impacts on the evolution of chromosomal organization [20], improvements in genetic engineering [21], and some genetic diseases [22,23]. In *S. cerevisiae*, the impact of the position effect on gene expression has attracted attention for decades due to its widespread application. *P. pastoris*, as one of the most widely used eukaryotic expression systems, has over 5000 genes distributed across four chromosomes. With glucose or methanol as the carbon source, the essential genes were identified in *P. pastoris* by transposon mutagenesis [24]. The *P. pastoris* genome contains several common integration loci (e.g., *AOX1*, *GAP*, *FLD1*), and the chromatin accessibility, transcriptional activity, and epigenetic regulation of these sites significantly affect the recombinant protein yields [25]. Studies have shown that integrating exogenous genes into the *AOX1* locus typically results in high protein production due to the strong inducible promoter (P_{AOX1}) [26]. However, alternative sites (e.g., *GAP* or *PEX8*) provide more stable constitutive expression [27]. Recent studies have highlighted the importance of understanding and optimizing the insertion site for foreign protein production in *P. pastoris*. For example, it was demonstrated that the expression of the recombinant human BiP gene *rhBiP* in *P. pastoris* was influenced by its integration site in the genome [28]. It was found that some integration sites led to higher expression levels compared to others, emphasizing the requirement for the careful selection of insertion sites in foreign protein production. Therefore, systematically evaluating the impacts of different integration sites on exogenous gene expression is essential in *P. pastoris*.

Transposable elements (TEs) are DNA fragments that use TE-encoded proteins to move or copy themselves from a donor to a target site within the host genome. The insertion of TEs can lead to gene inactivation, and the residuals of newly exposed ends at the donor site after TE excision can result in chromosomal aberrations. Therefore, TEs have a profound impact on gene expression and participate in genome evolution [29]. TEs can be divided into two major classes, namely retrotransposons (Class I) and DNA transposons (Class II). Retrotransposons employ replicative transposition via a ‘copy-and-paste’ mechanism, in which an RNA intermediate for transposition is carried out, while DNA transposons move via a non-replicative ‘cut-and-paste’ mechanism [30]. The *piggyBac* (PB) transposon belongs to the family of DNA transposons and was originally discovered and isolated from *Trichoplusia ni* cells [31]. About a decade later, the PB transposase was identified, and it was found to be able to excise and integrate non-autonomous DNA elements flanking

the transposon ends [32]. The original *PB* transposon is 2475 bp and consists of a gene coding the *PB* transposase and two terminal inverted repeats (5'TIR and 3'TIR) [33]. *PB* transposition occurs through a 'cut-and-paste' mechanism [34]. In most applications, the *PB* transposase and *PB* transposon are constructed on two separate plasmids. When the *PB* transposase is expressed, it binds to the inverted repeats of the transposon, creating nicks in the DNA and releasing the 3' hydroxyl groups at both ends of the transposon. The 3' hydroxyl groups then hydrolytically attack the flanking TTAA sequence, forming hairpin structures that release the transposon from the plasmid [35]. The *PB* transposase recognizes the TTAA sequence in the host genome and unwinds the hairpin structure of the transposon, which subsequently hydrolytically attacks the genomic DNA with its 3' hydroxyl groups, creating a staggered 4 bp cut in the genomic DNA and finally inserting the transposon into the TTAA site of the genome. The transposase can precisely excise the transposon from the original TTAA site and reinsert it into other TTAA sites without leaving a footprint. Due to its high transposition efficiency and large cargo capacity (>100 kb), the *PB* transposon has become an effective tool for gene manipulation and analysis [36]. It can efficiently integrate target genes into the genomes of various invertebrates and vertebrates through transposition [37–40]. Moreover, the *PB* transposon system has been used to construct mutagenesis libraries, and the phenotypes of mutants can be screened and analyzed to explore positional functions. Although *PB* originated from insects, it has been proven to be active in various organisms, including in *Saccharomyces cerevisiae* [35], *Schizosaccharomyces pombe* [41], plants [42], mammals [37], and human cells [39]. In *P. pastoris*, the *PB* transposon has been used for random mutagenesis [43]. However, compared with model strains such as *S. cerevisiae*, the *PB* transposon system has rarely been reported in *P. pastoris*.

Inspired by a study on the screening of essential genes by the transposons *TcBuster* (TcB) and *Sleeping Beauty* (SB) in *P. pastoris* [24], position effect variegation in the *P. pastoris* chromosome was investigated using the *PB* transposon system in our study. This study provides new ideas and a reference for the rapid and high-throughput screening of high-producing strains of foreign proteins in *P. pastoris*.

2. Materials and Methods

2.1. Strains and Cultivation

P. pastoris GS115 and its derivatives (Table S1) were grown in YPD medium, which contained 20 g/L glucose, 20 g/L peptone, and 10 g/L yeast extract. To screen the transformants, Minimal Dextrose (MD) medium was used, containing 20 g/L glucose, 1.34% (*w/v*) Yeast Nitrogen Base (YNB, containing biotin), and appropriate supplements (zeocin, histidine, or adenine) as needed [44]. For fermentation, Buffered Minimal Methanol (BMM) medium [44], containing 200 µM potassium phosphate buffer (pH 6.0) and 1.34% YNB (*w/v*), with 10 g/L methanol as a carbon source, was used. A preculture was grown in 3 mL of YPD medium supplemented with zeocin in a 15 mL glass tube for 10 h. The cells were then harvested by centrifugation, washed twice with BMM medium, and resuspended in 100 mL of fresh BMM medium in a 500 mL flask for the main fermentation. All *P. pastoris* strains were cultivated at 30 °C, 220 rpm in a shake incubator. *Escherichia coli* DH5α was grown in LB medium (10 g/L tryptone, 10 g/L NaCl, and 5 g/L yeast extract). LB medium supplemented with 100 mg/L of ampicillin was used for *E. coli* transformant selection and growth.

2.2. Construction of the *PB* Binary Transposition System

To use the *PB* binary transposition system in *P. pastoris*, we firstly constructed the PBL-eGFP-BleoR-PBR cassette. PBL and PBR were amplified from the plasmid pPB[ura4] [41] with the primers PBL-F/PBL-R and PBR-F/PBR-R; the zeocin resistance gene *BleoR* was

amplified from pPICZ α with the primers BleoR-F/BleoR-R. The reporter gene *eGFP* was expressed under the control of a weak promoter P_{ypt1} , which was amplified from *P. pastoris* GS115 with the primers Pypt1-F/Pypt1-R. Overlap PCR was performed with the above amplified fragments to generate the PBL-eGFP-BleoR-PBR cassette. The DNA fragments containing the flanking sequences of *ADE2* were amplified from *P. pastoris* GS115 with the primers ADE2-L-F/ADE2-L-R and ADE2-R-F/ADE2-R-R and were fused on both sides of the PBL-eGFP-BleoR-PBR cassette. The PBL-eGFP-BleoR-PBR cassette with the flanking arms and the autonomously replicating sequence (PARS) amplified from *P. pastoris* GS115 with the primers PARS-F/PARS-R were cloned into the backbone plasmid pUC19 (linearization with *Zra*I and *Pci*I) to generate pIMY015. The plasmid pIMY029 was constructed based on pPIC9K by inserting the PARS and using P_{AOX1} to drive *piggyBac* Transposase (PBase) gene expression. Then, pIMY015 was linearized by the restriction enzyme *Ahd*I. The *P. pastoris* strain P.pg0015 was obtained by inserting the PBL-eGFP-BleoR-PBR cassette at the TTAA site (n.t. 125–128) within the open reading frame (ORF) of *ADE2*. pIMY029 was then transformed into P.pg0015 to generate P.pg0015P. The strains and plasmids used in this study are listed in Table S1. The primers used in this study are listed in Table S2. The plasmid maps of pIMY015 and pIMY029 are presented in Figure 1.

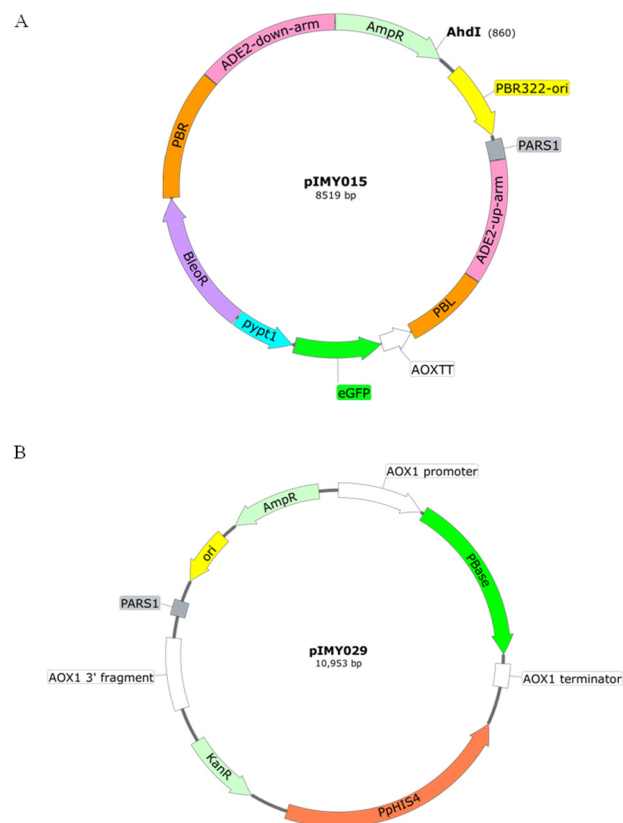


Figure 1. The PB binary transposition system. (A) The plasmid pIMY015 carrying the PBL-eGFP-BleoR-PBR cassette; (B) the plasmid pIMY029 for PBase gene expression under the control of the methanol-induced promoter P_{AOX1} .

2.3. Transposition Induction in *P. pastoris*

P.pg0015P was streaked onto a YPDZ (YPD agar containing zeocin) plate and incubated at 30 °C to obtain a single colony. One single colony was picked and inoculated in BMG liquid medium (200 μ M potassium phosphate buffer, 1.34% YNB, 10 g/L glycerol) supplemented with adenine and cultured at 30 °C with shaking at 220 rpm until it reached the logarithmic growth phase. The cells were harvested by centrifugation, washed, and adjusted to a final OD₆₀₀ of 0.001 in BMM medium, followed by daily methanol

supplementation. In our design, the PBase gene expression is under the control of the methanol-inducible P_{AOX1} promoter. Hence, 1% (v/v) methanol was utilized to induce PBase gene expression and then initiate *piggyBac* transposition. Samples were taken every 24 h, serially diluted, and spread onto MDH medium (MD supplemented with histidine), MDHA medium (MDH with adenine), and MDHZ medium (MDH with zeocin) for colony counting. The *PB* excision efficiency was calculated as (colony count on MDH/colony count on MDHA) $\times$ 100%. The *PB* reinsertion rate was determined as (colony count on MDHZ/colony count on MDHA) $\times$ 100%.

2.4. Flow Cytometry

P.pg0015P was induced by adding methanol and spread on MDHZ medium for growth until small white colonies emerged. Then, cells were collected, washed, and resuspended in PBS buffer before sorting. A BD FACSaria III (BD Biosciences, San Jose, CA, USA) equipped with 488 nm lasers was used for fluorescence-activated cell sorting (FACS). The strains with low expression or high expression of *eGFP* were sorted and named G15P-P4 and G15P-P67, respectively.

2.5. Microscopy

After transposition was induced with methanol, cells were collected by centrifugation, washed and resuspended, and subsequently spread on MDHZ and YPDZ plates, respectively. The plates were incubated at 30 °C for two days. The fluorescence of *eGFP* was excited at 488 nm and detected at 491–535 nm on a Leica SP8 confocal laser microscope.

2.6. Profiling the Insertion Sites of *PB* with High-Throughput Sequencing

The genomic DNA of G15P, G15P-P4, and G15P-P67 was extracted using the TIANamp Yeast DNA Kit (TIANGEN, Beijing, China). Tagmentation was carried out using the TruePrep DNA Library Prep Kit V2 for Illumina (Vazyme Biotech, Nanjing, China). The first-round PCR was carried out using PBLseq and N7XX to amplify the *PB* element together with its flanking sequences. The second-round PCR was performed using Illumine-R and DM5XX (Vazyme Biotech, Nanjing, China) to add sequences required for Illumina sequencing. The size selection and purification of the PCR products were carried out with VAHTS DNA Clean Beads (Vazyme Biotech, Nanjing, China). The *PB* insertion site mapping was performed using the processing and analysis pipeline previously described [45].

3. Results

3.1. Construction of the *piggyBac* Transposition System in *P. pastoris*

To establish a *PB* binary transposition system in *P. pastoris*, we constructed a plasmid pIMY015 (Figure 1A) carrying the PBL-*eGFP*-BleoR-PBR cassette. After linearization by the restriction enzyme *AhdI*, pIMY015 was transformed into *P. pastoris* GS115. Finally, the strain P.pg0015 was obtained by inserting the PBL-*eGFP*-BleoR-PBR cassette of pIMY015 at the TTAA site (n.t. 125–128) within the open reading frame (ORF) of *ADE2*. The insertion of *BleoR* endowed P.pg0015 with resistance to zeocin. The disruption of *ADE2* led to the adenine auxotrophy of P.pg0015 and caused the accumulation of purine precursors in the cellular vacuole, which gave the cells a pink color. The *eGFP* gene was designed to be driven by a weak constitutive promoter P_{ypt1} to mitigate the confounding influence of endogenous promoters at the transposon insertion sites.

Subsequently, the plasmid pIMY029 (Figure 1B) was constructed based on pPIC9K, in which P_{AOX1} was used to drive the expression of the *piggyBac* transposase (PBase) gene. Due to lacking a self-replicating component, we amplified the *P. pastoris* autonomous replication sequence (PARS) and inserted it in the *NdeI* site of pPIC9K. The resulting pIMY029 was

transformed into P.pg0015 to generate the strain P.pg0015P. The precise excision of *PB* would convert P.pg0015P cells from adenine-auxotrophic to adenine-prototrophic, restoring the cell color to white. Zeocin resistance was retained only when the PBL-eGFP-BleoR-PBR cassette was excised from the donor site and successfully reintegrated into a new genomic locus, thereby maintaining the resistance cassette in the genome. When *PB* excision occurred without a productive reintegration event (i.e., the transposon was lost), the zeocin resistance cassette was no longer present in the genome and the zeocin resistance was lost (Figure 2).

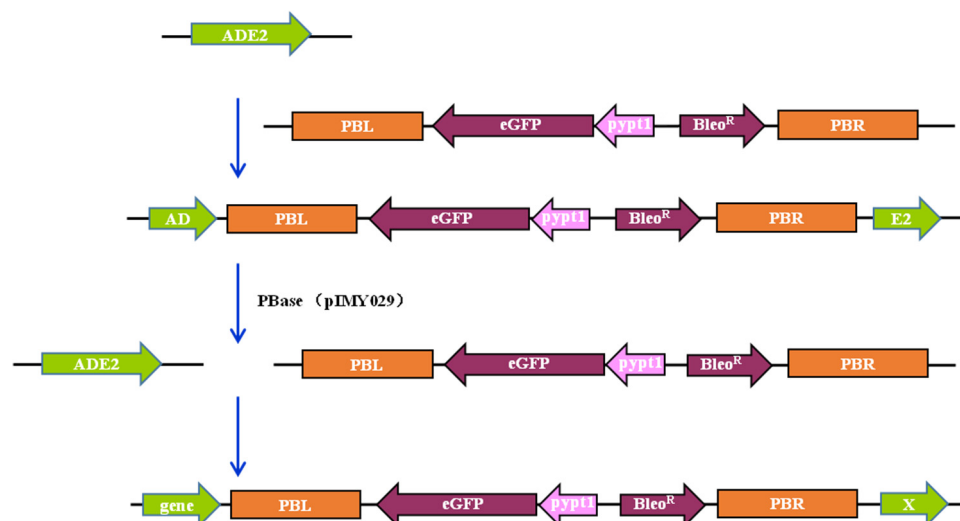


Figure 2. Diagram of transposition via the *PB* binary system in *P. pastoris*.

With methanol as the carbon source, the expression of the PBase gene was induced in P.pg0015P and the cells were spread on adenine-deficient (-Ade) plates with/without zeocin. When the methanol-induced cells were spread on MDH or MDHZ plates, white colonies were obtained (Figure 3), while the non-induced cells could not grow. This result indicated that the methanol-induced PBase successfully excised the *PB* cassette from the original locus and reinserted it into a new locus.

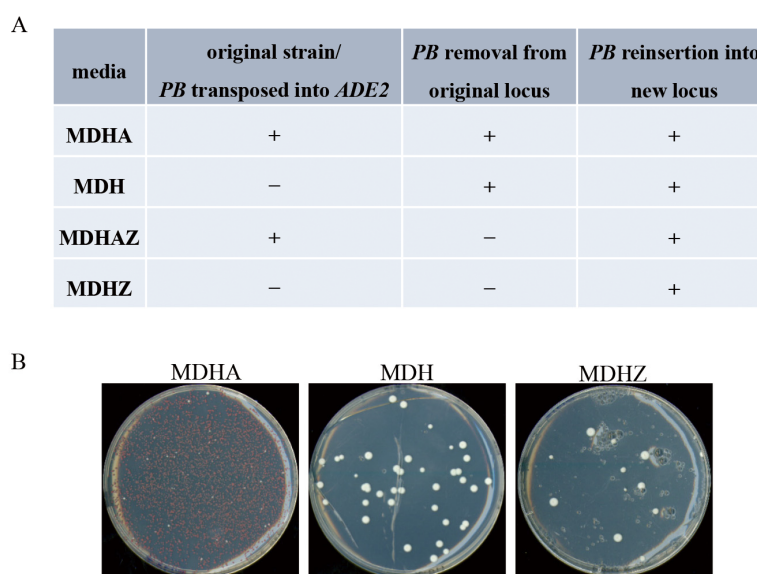


Figure 3. The growth of the transposable strains at different stages. (A) Theoretical growth profile of the transposable strains at different stages on MD medium supplemented with/without adenine or zeocin. “–”, no visible growth; “+”, normal growth. (B) Growth of the transposable strains after methanol induction on MDHA, MDH, and MDHZ.

3.2. PB Transposition Efficiency in *P. pastoris*

Aliquots of methanol-induced cells were harvested at intervals, and each was divided into three equal parts. One part was spread on MDHA medium, in which all cells could normally grow; another part was spread on MDH medium, in which the cells where the *PB* cassette was excised could grow; and the third part was spread on MDHZ medium, in which only the *PB*-reinserted cells could grow. Finally, the *PB* removal rate was calculated as 1.24% through dividing the number of colonies on MDH by the total number of colonies on MDHA. The *PB* transposition frequency was calculated through dividing the number of colonies on MDHZ by the total number of colonies on MDHA. The results demonstrated that the *PB* transposition frequency increased with the extension of the methanol induction time, and it reached about 0.39% at 70 h (Figure S1).

3.3. Position Effect Variegation in *P. pastoris*

To initially evaluate the fluorescence intensity, the strain P.pg0015P was induced to undergo transposition and then incubated on YPDZ medium (YPD medium supplemented with zeocin) and MDHZ medium. Upon observation under confocal fluorescence microscopy, the majority of the strains displayed negligible fluorescence, a smaller subset exhibited faint fluorescence, and only a minute fraction demonstrated marked fluorescence (Figure 4). These variations in the fluorescence intensity possibly arose from the position effect due to the insertion of the eGFP-carrying *PB* element in diverse genomic loci. In the strains grown on MDHZ, the *PB* element was successfully transposed from its original site—specifically, the TTAA insertion site of the *ADE2* gene—to a novel locus. Subsequently, these strains, designated as G15P, were collected. One aliquot of the collected strains was used to construct a library for sequencing, while the remaining aliquot was preserved for subsequent flow cytometry-based cell sorting.

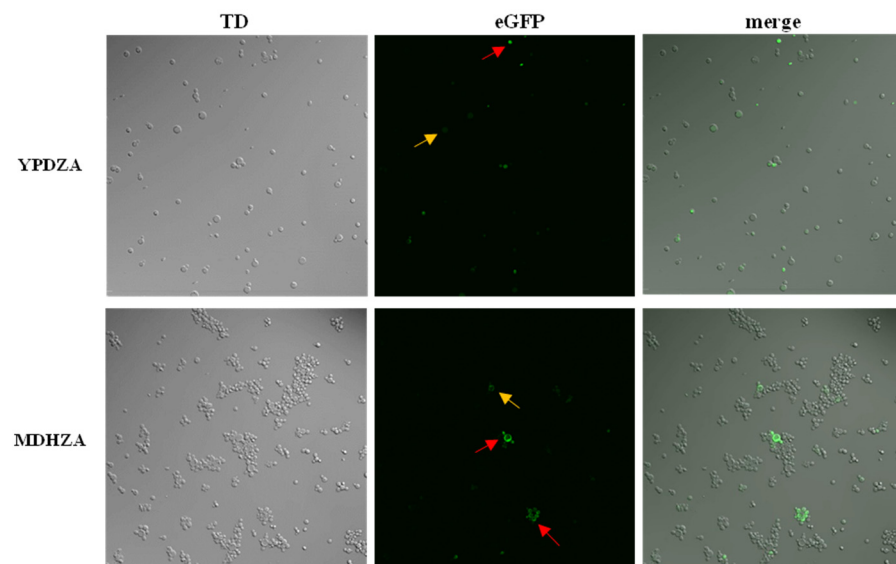


Figure 4. Fluorescence intensity of G15P observed under confocal fluorescence microscopy. YPDZA, the G15P strains were cultured in YPD medium supplemented with adenine and zeocin; MDHZA, the G15P strains were cultured in MDH medium supplemented with adenine and zeocin. TD, transmitted detector image; eGFP, fluorescence of the G15P strains expressing *eGFP* (Ex 488 nm/Em 500–550 nm); merge, merge of eGFP (green) fluorescence and transmitted light image (TD, gray). Orange arrow, low-*eGFP*-expression strains; red arrow, high-*eGFP*-expression strains.

3.4. Genome-Wide Analysis of the *PB* Transposon Distribution in *P. pastoris*

The *PB* distribution in the *P. pastoris* genome was analyzed using next-generation sequencing (NGS). The genomic DNA of transposable strains was extracted from the G15P

pool and subjected to NGS. A total of 30,092,056 raw reads and 15,452,566 clean reads were acquired. Through sequencing analysis, 14,064 loci that aligned with the *P. pastoris* GS115 genomic sequence were identified, exhibiting a notably uniform distribution across distinct chromosomal positions (Figure 5A). Notably, the *PB* transposon was found to predominantly target TTAA sites, a preference that corroborates previous reports (Figure 5C). Nucleotide motif analysis revealed that 77.87% of the insertion sites were located at TTAA sites, 1.17% at CTAA sites, 1.14% at TTAG sites, 0.88% at ATAA sites, and 0.87% at TTAT sites (Figure 5B).

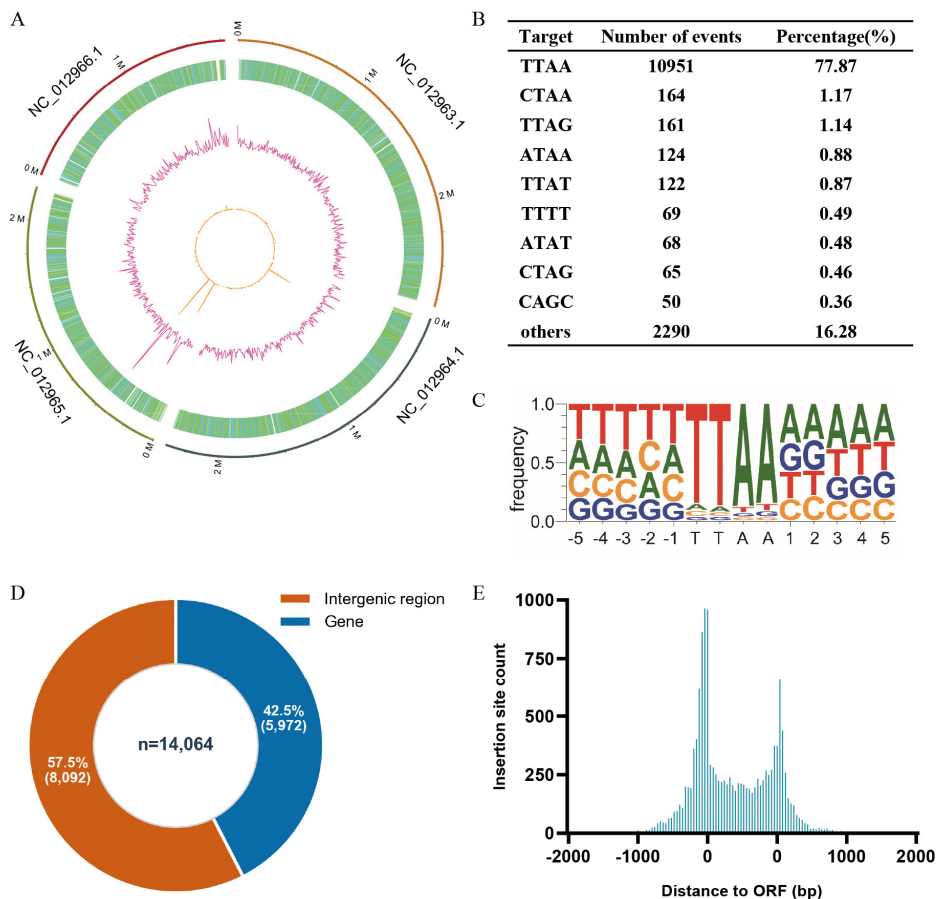


Figure 5. Insertion distribution and target site bias of the *PB* transposon in the *P. pastoris* chromosome. (A) Distribution of the ORF and transposon insertions in the *P. pastoris* GS115 genome. From the outermost to innermost, the first circle represents the genomic position in megabases, the second circle represents ORFs, and the third circle indicates the insertion number per 10 kb, while the innermost circle indicates the read number per 10 kb. (B) Number and proportion of *PB* transposon insertion target sequences. (C) WebLogos generated from the insertion sites containing 5 bp upstream and downstream sequences of the *PB* transposon using www.weblogo.berkeley.edu. (D) Distribution of *PB* transposon insertions in the genes and their intergenic regions. (E) The *PB* transposon insertion sites are plotted relative to the ORF positions. Each ORF is divided into 25 equal-sized segments and the number of insertions in each segment is displayed. Insertion sites in the intergenic regions closer to the 5' or 3' ends of an ORF are plotted upstream or downstream of the ORF, respectively. X-axis indicates distance to ORF, while Y-axis indicates integration events.

Genome annotation and the detailed analysis of the distribution of the *PB* insertions indicated that 58% of these insertions were located in intergenic regions (Figure 5D). This is very similar to previous studies conducted using the transposons *TcBuster* (TcB) and *Sleeping Beauty* (SB) in *P. pastoris* [24]. This preference of *PB* insertion for intergenic regions was possibly influenced by the distribution of nucleosomes, as reported in *S. cerevisiae*,

in which the insertion of *Hermes* transposons from the hAT family was significantly affected by the nucleosome distribution, with higher nucleosome occupancy correlating with a lower transposon insertion density and the lowest nucleosome occupancy in the intergenic regions, particularly near the start and end sites of genes [46]. In fact, the *PB* transposon has been used to reveal gene functions in *P. pastoris* [47]. The distances of the insertion sites from open reading frames (ORFs) were analyzed, and the results demonstrated that *PB* transposons exhibited a high insertion density in the promoter regions upstream of ORFs and the terminator regions downstream of ORFs, with a sharp decrease in the insertion density within the ORFs and an extremely low insertion density at sites far from the ORFs (Figure 5E).

3.5. Distribution of High- and Low-Expression Loci in the *P. pastoris* Genome

To investigate positional effect variegation in *P. pastoris* using the *PB* system, G15P was subjected to flow cytometry cell sorting. A low-fluorescence strain pool named G15P-P4 and a high-fluorescence strain pool named G15P-P67 were obtained, from which the genomic DNA was extracted and used for sequencing (Figure S2). A total of 21,033,168 clean reads were obtained, including 5.26 G clean data, for G15P-P4, and 16,654,590 clean reads were obtained, including 4.16 G clean data, for G15P-P67 (Table 1).

Table 1. Sequencing data from G15P-P4 and G15P-P67.

Sample	Low Expression (G15P-P4)	High Expression (G15P-P67)
Sorting cell number	1.4×10^7	1.8×10^6
Clean reads	21,033,168	16,654,590
Clean base (G)	5.26	4.16
GC content (%)	37.91	39.38

Sequencing analysis revealed that 8863 sites from G15P-P4 were aligned to the *P. pastoris* GS115 genome, and the insertion sites covered different loci in the four chromosomes, whereas only 796 sites from G15P-P67 were aligned to the *P. pastoris* GS115 genome, and the insertion sites were sparsely distributed across different loci in the four chromosomes (Figure 6A–C). The analysis revealed that the proportion of high-expression loci in each chromosome was significantly lower compared to the low-expression sites (Figure 6B). Concurrently, an analysis of the distances of low-expression loci from open reading frames (ORFs) revealed an exceptionally high insertion density in the promoter region and the terminator region, a sharp reduction in the insertion density within ORFs, and a very low insertion density at locations far from ORFs (Figure 6F). However, the high-expression loci exhibited a different pattern in terms of their distances from ORFs, appearing as an extremely high insertion density within ORFs, followed by the promoter region upstream of ORFs and the terminator region downstream of ORFs (Figure 6G). Additionally, a statistical analysis of the distribution of high- and low-expression loci within and between genes was conducted. The proportion of low-expression loci in intergenic regions was found to be as high as 63% (Figure 6D). In contrast, the proportion of high-expression loci within genes was as high as 77% (Figure 6E).

We performed KEGG and GO analyses on the genes containing insertion sites (Figure 7). Based on the KEGG analysis, we found that the low-expression genes were significantly enriched in the MAPK signaling pathway, autophagy, mitochondrial autophagy, ABC transporters, and the arginine synthesis pathway, while the high-expression genes were significantly enriched in ATP-dependent chromatin remodeling and the arginine synthesis pathway. The results of the GO analysis indicated that the low-expression genes were primarily enriched in the transmembrane, phosphorylation, membrane formation, zinc ion binding, and ATPase activity pathways, whereas the high-expression genes were

mainly enriched in the DNA replication, plasma membrane formation, cytoplasm, and histone binding pathways. It has been reported that ATP-dependent remodeling complexes can move nucleosomes along DNA, facilitate histone exchange, or completely replace nucleosomes in DNA, thereby affecting gene expression [48]. In our study, the high-expression genes were significantly enriched in the ATP-dependent chromatin remodeling and histone binding pathways (Table S3). This observation led us to postulate that the insertion of the transposon in such genes for the ATP-dependent chromatin remodeling machinery and histone binding factors could potentially alter nucleosome positioning and histone binding, thereby exerting a profound influence on gene expression profiles.

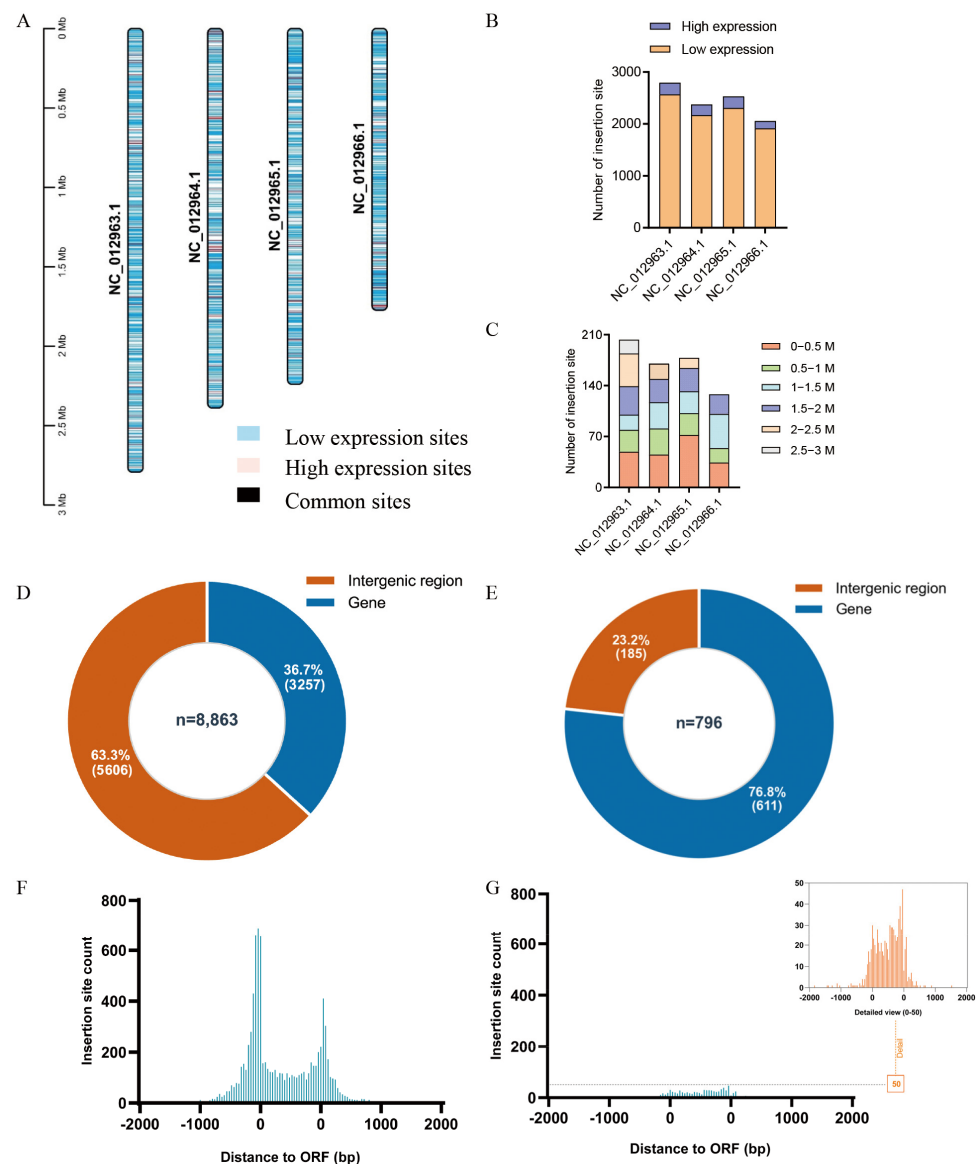


Figure 6. The insertion distributions and target site biases of high- and low-expression loci. (A) Distribution of the insertion sites with high and low expression levels, respectively, in the *P. pastoris* GS115 genome. (B) The number of high- and low-expression loci in each chromosome. (C) The number of high-expression loci in each chromosome per 0.5 Mb. (D,E) Distribution of low-expression loci and high-expression loci in genes and intergenic regions. (F,G) Low- and high-expression loci are plotted relative to ORF positions. Each ORF is divided into 25 equal-sized segments and the number of insertions in each segment is displayed. Insertion sites in intergenic regions closer to the 5' or 3' end of an ORF are plotted upstream or downstream of the ORF, respectively. X-axis indicates distance to ORF, while Y-axis indicates integration events.

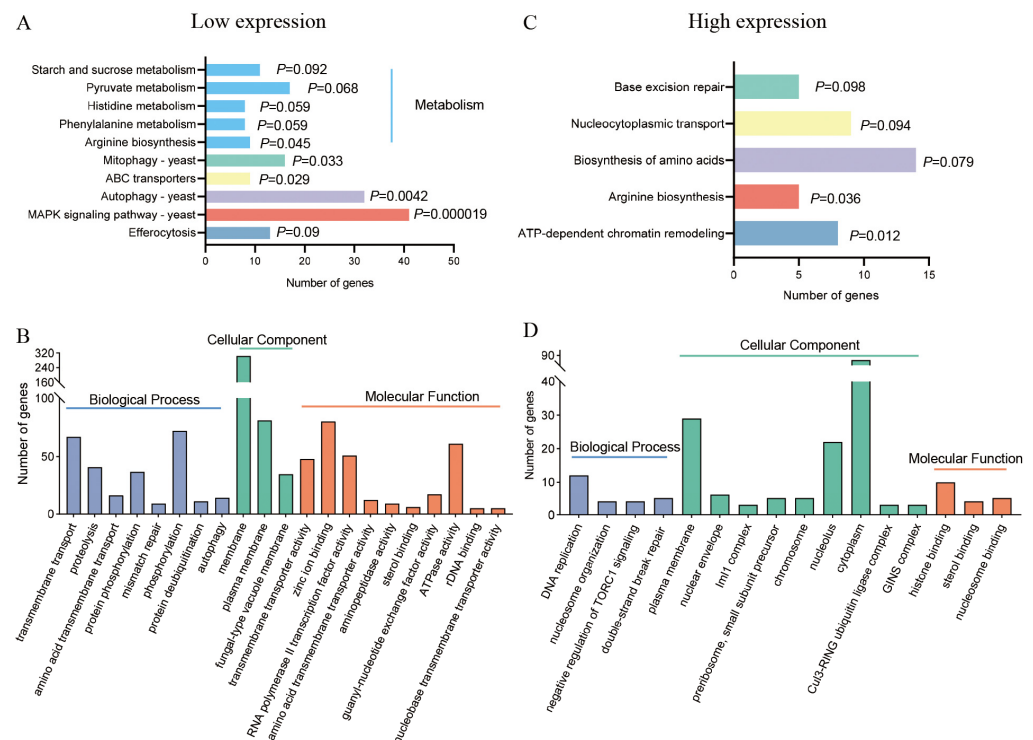


Figure 7. KEGG and GO analyses of the inserted genes with high or low expression levels. **(A)** KEGG analysis of the inserted genes with low expression levels; **(B)** GO analysis of the inserted genes with low expression levels; **(C)** KEGG analysis of the inserted genes with high expression levels; **(D)** GO analysis of the inserted genes with high expression levels.

4. Discussion

We constructed a binary transposition system based on the *piggyBac* (PB) transposon and utilized this system to create a random insertional mutagenesis library in which the *eGFP* expression cassette was randomly integrated at different chromosomal loci within *P. pastoris*.

Next-generation sequencing revealed that the primary recognition site for the PB transposon in *P. pastoris* is TTAA, which is highly represented in the chromosomes, making it a useful tool for studying the position effect in *P. pastoris*. Despite the low proportion (19.8%) of intergenic regions in *P. pastoris*, a significant proportion (58%) of PB transposon insertion sites was distributed in regions possibly related to nucleosome positioning; alternatively, insertions at intergenic regions may be favored because they are less lethal and harmful for the cell. This finding is similar to those of previous studies conducted using TcB and SB [24], indicating that transposons other than PB could be used to study gene function and screen high-expression loci in *P. pastoris*. By employing flow cytometry to sort the transposon pool named G15P, we obtained the pool of low-*eGFP*-expression strains G15P-P4 and the pool of high-*eGFP*-expression strains G15P-P67. Sequencing analysis confirmed the existence of chromosomal position effects in *P. pastoris*. The high-expression sites were predominantly located within ORFs, while the low-expression sites were primarily found in the promoter regions and the terminator regions. The proportion of low-expression sites in the intergenic regions was found to be as high as 63%, closely resembling the 58% observed in G15P. This may be due to the low-expression sites being primarily concentrated in the promoter regions and the terminator regions, where nucleosome occupancy is diminished, making them more susceptible to insertion, but not in actively transcribed regions, resulting in the reduced expression of foreign proteins. In contrast, the proportion of high-expression sites within genes was as high as 77%, which may have been due to the fact the gene regions

where the high-expression sites were located were transcriptionally active areas, allowing for the robust expression of foreign proteins. These loci with high expression levels could be used for exogenous gene expression in *P. pastoris*.

In our study, the observed transposition frequency of *PB* was notably low at approximately 0.39%. This limited efficiency could be caused by several factors. Primarily, the use of wild-type PBase, rather than a hyperactive variant, likely constitutes a major constraint. Previous studies have demonstrated that hyperactive PBase can enhance excision activity several-fold in yeast [49]. Furthermore, the recent development of highly efficient, AI-engineered ‘mega-active’ synthetic transposases using protein language models (e.g., Progen2) underscores the potential for substantial gains in transposition efficiency through molecular optimization [50,51]. Beyond the transposase itself, the host cell’s DNA repair machinery presents another critical barrier. The low frequency of homologous recombination (HR) in *P. pastoris* may hinder the resolution of the transposition intermediate. To favor homology-directed repair over the error-prone non-homologous end joining (NHEJ) pathway, future efforts could involve the genetic engineering of the host, such as generating a Δ ku70 knockout to disrupt NHEJ and/or overexpressing key HR proteins like RAD52 and RAD59 [52,53].

This study has, to some extent, revealed the chromosomal position effects in *P. pastoris* and provides a novel research approach for the high-throughput screening of strains with high expression of foreign proteins in *P. pastoris*. Based on this work, we will implement this approach—using *eGFP* as a selectable reporter and incorporating expression cassettes for heterologous genes—to establish a rapid, high-throughput screening platform for the isolation of microbial strains that achieve the exceptionally high-level production of the target protein in the future.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/jof12020136/s1>, Figure S1: Construction of the plasmids pIMY015 and pIMY029; Figure S2: Efficiency of the *PB* transposon in *P. pastoris*; Figure S3: Flow cytometry sorting different fluorescence intensity strains; Table S1: Strains and plasmids used in this study; Table S2: Primers used in this study; Table S3: The high-expression loci in G15P-P67.

Author Contributions: Conceptualization, G.L. and X.Y.; methodology, G.L. and X.Y.; software, X.Y.; validation, G.L. and X.Y.; formal analysis, X.Y.; investigation, X.Y., Z.Z., W.G. and Q.Z.; resources, G.L.; data curation, X.Y. and B.C.; writing—original draft preparation, X.Y.; writing—review and editing, G.L., Y.P. and Y.Y.; visualization, X.Y.; supervision, G.L.; project administration, G.L.; funding acquisition, G.L. and X.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China (Grant No. 2021YFA0910600), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA28030000), and the National Natural Science Foundation of China (Grant No. 32200058).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The *PB* transposon sequencing data have been submitted to GenBank SRA (PRJNA1376236).

Acknowledgments: We thank Lilin Du (National Institute of Biological Sciences, Beijing) for the generous donation of the plasmid pPB[ura4].

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analysis, or interpretation of the data; in the writing of the manuscript; or in the decision to publish the results.

References

- Wegner, G.H. Emerging applications of the methylotrophic yeasts. *FEMS Microbiol. Rev.* **1990**, *7*, 279–283. [\[CrossRef\]](#)
- De Schutter, K.; Lin, Y.C.; Tiels, P.; Van Hecke, A.; Glinka, S.; Weber-Lehmann, J.; Rouze, P.; Van de Peer, Y.; Callewaert, N. Genome sequence of the recombinant protein production host *Pichia pastoris*. *Nat. Biotechnol.* **2009**, *27*, 561–566. [\[CrossRef\]](#)
- Nocon, J.; Steiger, M.G.; Pfeffer, M.; Sohn, S.B.; Kim, T.Y.; Maurer, M.; Russmayer, H.; Pflugl, S.; Ask, M.; Haberhauer-Troyer, C.; et al. Model based engineering of *Pichia pastoris* central metabolism enhances recombinant protein production. *Metab. Eng.* **2014**, *24*, 129–138. [\[CrossRef\]](#) [\[PubMed\]](#)
- Schwarzshans, J.P.; Luttermann, T.; Geier, M.; Kalinowski, J.; Friehs, K. Towards systems metabolic engineering in *Pichia pastoris*. *Biotechnol. Adv.* **2017**, *35*, 681–710. [\[CrossRef\]](#) [\[PubMed\]](#)
- Walsh, G. Biopharmaceutical benchmarks 2018. *Nat. Biotechnol.* **2018**, *36*, 1136–1145. [\[CrossRef\]](#)
- Ahmad, M.; Hirz, M.; Pichler, H.; Schwab, H. Protein expression in *Pichia pastoris*: Recent achievements and perspectives for heterologous protein production. *Appl. Microbiol. Biotechnol.* **2014**, *98*, 5301–5317. [\[CrossRef\]](#) [\[PubMed\]](#)
- Akhtar, W.; de Jong, J.; Pindyurin, A.V.; Pagie, L.; Meuleman, W.; de Ridder, J.; Berns, A.; Wessels, L.F.; van Lohuizen, M.; van Steensel, B. Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell* **2013**, *154*, 914–927. [\[CrossRef\]](#)
- Bryant, J.A.; Sellars, L.E.; Busby, S.J.; Lee, D.J. Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res.* **2014**, *42*, 11383–11392. [\[CrossRef\]](#)
- Chen, M.; Licon, K.; Otsuka, R.; Pillus, L.; Ideker, T. Decoupling epigenetic and genetic effects through systematic analysis of gene position. *Cell Rep.* **2013**, *3*, 128–137. [\[CrossRef\]](#)
- Chen, X.; Zhang, J. The genomic landscape of position effects on protein expression level and noise in yeast. *Cell Syst.* **2016**, *2*, 347–354. [\[CrossRef\]](#)
- Gierman, H.J.; Indemans, M.H.; Koster, J.; Goetze, S.; Seppen, J.; Geerts, D.; van Driel, R.; Versteeg, R. Domain-wide regulation of gene expression in the human genome. *Genome Res.* **2007**, *17*, 1286–1295. [\[CrossRef\]](#) [\[PubMed\]](#)
- Markstein, M.; Pitsouli, C.; Villalta, C.; Celniker, S.E.; Perrimon, N. Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nat. Genet.* **2008**, *40*, 476–483. [\[CrossRef\]](#)
- Ogawa, H.; Miyazaki, H.; Kimura, M. Isolation and characterization of human skin lysozyme. *J. Investig. Dermatol.* **1971**, *57*, 111–116. [\[CrossRef\]](#)
- Pavitt, G.D.; Higgins, C.F. Chromosomal domains of supercoiling in *Salmonella typhimurium*. *Mol. Microbiol.* **1993**, *10*, 685–696. [\[CrossRef\]](#)
- Sauer, C.; Syvertsson, S.; Bohorquez, L.C.; Cruz, R.; Harwood, C.R.; van Rij, T.; Hamoen, L.W. Effect of genome position on heterologous gene expression in *Bacillus subtilis*: An unbiased analysis. *ACS Synth. Biol.* **2016**, *5*, 942–947. [\[CrossRef\]](#)
- Sousa, C.; de Lorenzo, V.; Cebolla, A. Modulation of gene expression through chromosomal positioning in *Escherichia coli*. *Microbiology* **1997**, *143*, 2071–2078. [\[CrossRef\]](#)
- Thompson, A.; Gasson, M.J. Location effects of a reporter gene on expression levels and on native protein synthesis in *Lactococcus lactis* and *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **2001**, *67*, 3434–3439. [\[CrossRef\]](#)
- Grewal, S.I.; Jia, S. Heterochromatin revisited. *Nat. Rev. Genet.* **2007**, *8*, 35–46. [\[CrossRef\]](#) [\[PubMed\]](#)
- Dey, S.S.; Foley, J.E.; Limsirichai, P.; Schaffer, D.V.; Arkin, A.P. Orthogonal control of expression mean and variance by epigenetic features at different genomic loci. *Mol. Syst. Biol.* **2015**, *11*, 806. [\[CrossRef\]](#)
- Batada, N.N.; Hurst, L.D. Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat. Genet.* **2007**, *39*, 945–949. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wilson, C.; Bellen, H.J.; Gehring, W.J. Position effects on eukaryotic gene expression. *Annu. Rev. Cell Biol.* **1990**, *6*, 679–714. [\[CrossRef\]](#)
- Kleinjan, D.J.; van Heyningen, V. Position effect in human genetic disease. *Hum. Mol. Genet.* **1998**, *7*, 1611–1618. [\[CrossRef\]](#)
- Milot, E.; Fraser, P.; Grosveld, F. Position effects and genetic disease. *Trends Genet.* **1996**, *12*, 123–126. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zhu, J.X.; Gong, R.Q.; Zhu, Q.Y.; He, Q.; Xu, N.; Xu, Y.; Cai, M.; Zou, X.; Zhang, Y.; Zhou, M. Genome-wide determination of gene essentiality by transposon insertion sequencing in yeast *Pichia pastoris*. *Sci. Rep.* **2018**, *8*, 10223. [\[CrossRef\]](#) [\[PubMed\]](#)
- Sturmberger, L.; Chappell, T.; Geier, M.; Krainer, F.; Day, K.J.; Vide, U.; Trstenjak, S.; Schiefer, A.; Richardson, T.; Soriaga, L.; et al. Refined *Pichia pastoris* reference genome sequence. *J. Biotechnol.* **2016**, *235*, 121–131. [\[CrossRef\]](#)
- Cereghino, J.L.; Cregg, J.M. Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol. Rev.* **2000**, *24*, 45–66. [\[CrossRef\]](#) [\[PubMed\]](#)
- Waterham, H.R.; Digan, M.E.; Koutz, P.J.; Lair, S.V.; Cregg, J.M. Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. *Gene* **1997**, *186*, 37–44. [\[CrossRef\]](#)
- Žitkus, E.; Čiplys, E.; Žiaunys, M.; Sakalauskas, A.; Slibinskas, R. Development of an efficient expression system for human chaperone BiP in *Pichia pastoris*: Production optimization and functional validation. *Microb. Cell Fact.* **2025**, *24*, 66. [\[CrossRef\]](#)
- Feschotte, C.; Pritham, E.J. DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **2007**, *41*, 331–368. [\[CrossRef\]](#)
- Hayward, A.; Gilbert, C. Transposable elements. *Curr. Biol.* **2022**, *32*, R904–R909. [\[CrossRef\]](#)

31. Fraser, M.J.; Smith, G.E.; Summers, M.D. Acquisition of host cell DNA sequences by baculoviruses: Relationship between host DNA insertions and FP mutants of *Autographa californica* and *Galleria mellonella* nuclear polyhedrosis viruses. *J. Virol.* **1983**, *47*, 287–300. [CrossRef] [PubMed]
32. Elick, T.A.; Bauser, C.A.; Fraser, M.J. Excision of the piggyBac transposable element in vitro is a precise event that is enhanced by the expression of its encoded transposase. *Genetica* **1996**, *98*, 33–41. [CrossRef]
33. Cary, L.C.; Goebel, M.; Corsaro, B.G.; Wang, H.G.; Rosen, E.; Fraser, M.J. Transposon mutagenesis of baculoviruses: Analysis of *Trichoplusia ni* transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology* **1989**, *172*, 156–169. [CrossRef]
34. Woodard, L.E.; Wilson, M.H. piggyBac-ing models and new therapeutic strategies. *Trends Biotechnol.* **2015**, *33*, 525–533. [CrossRef] [PubMed]
35. Mitra, R.; Fain-Thornton, J.; Craig, N.L. piggyBac can bypass DNA synthesis during cut and paste transposition. *EMBO J.* **2008**, *27*, 1097–1109. [CrossRef]
36. Li, M.A.; Turner, D.J.; Ning, Z.; Yusa, K.; Liang, Q.; Eckert, S.; Rad, L.; Fitzgerald, T.W.; Craig, N.L.; Bradley, A. Mobilization of giant piggyBac transposons in the mouse genome. *Nucleic Acids Res.* **2011**, *39*, e148. [CrossRef]
37. Ding, S.; Wu, X.; Li, G.; Han, M.; Zhuang, Y.; Xu, T. Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell* **2005**, *122*, 473–483. [CrossRef]
38. Lu, Y.; Lin, C.; Wang, X. PiggyBac transgenic strategies in the developing chicken spinal cord. *Nucleic Acids Res.* **2009**, *37*, e141. [CrossRef]
39. Wilson, M.H.; Coates, C.J.; George, A.L., Jr. PiggyBac transposon-mediated gene transfer in human cells. *Mol. Ther.* **2007**, *15*, 139–145. [CrossRef]
40. Yusa, K.; Rad, R.; Takeda, J.; Bradley, A. Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon. *Nat. Methods* **2009**, *6*, 363–369. [CrossRef] [PubMed]
41. Li, J.; Zhang, J.M.; Li, X.; Suo, F.; Zhang, M.J.; Hou, W.; Han, J.; Du, L.L. A piggyBac transposon-based mutagenesis system for the fission yeast *Schizosaccharomyces pombe*. *Nucleic Acids Res.* **2011**, *39*, e40. [CrossRef] [PubMed]
42. Johnson, E.T.; Dowd, P.F. A non-autonomous insect piggyBac transposable element is mobile in tobacco. *Mol. Genet. Genomics* **2014**, *289*, 895–902. [CrossRef]
43. Jiao, J.; Wang, S.; Liang, M.; Zhang, Y.; Xu, X.; Zhang, W.; Liu, B. Basal transcription profiles of the rhamnose-inducible promoter PLRA3 and the development of efficient PLRA3-based systems for markerless gene deletion and a mutant library in *Pichia pastoris*. *Curr. Genet.* **2019**, *65*, 785–798. [CrossRef] [PubMed]
44. Thermo Fisher Scientific. *Pichia* Expression Kit: User Guide for Expression of Recombinant Proteins in *Pichia pastoris* [User Manual]. (Publication No. MAN0000012, Rev. B., Catalog No. K171001). Available online: <https://www.thermofisher.com/order/catalog/product/K171001> (accessed on 15 September 2020).
45. Li, Z.; Wang, H.; Cai, C.; Wong, A.H.; Wang, J.; Gao, J.; Wang, Y. Genome-wide piggyBac transposon-based mutagenesis and quantitative insertion-site analysis in haploid *Candida* species. *Nat. Protoc.* **2020**, *15*, 2705–2727. [CrossRef]
46. Gangadharan, S.; Mularoni, L.; Fain-Thornton, J.; Wheelan, S.J.; Craig, N.L. DNA transposon hermes inserts into DNA in nucleosome-free regions in vivo. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 21966–21972. [CrossRef]
47. Zhu, J.; Zhu, Q.; Gong, R.; Xu, Q.; Cai, M.; Jiang, T.; Zhou, X.; Zhou, M.; Zhang, Y. PiggyBac transposon-mediated mutagenesis and application in yeast *Komagataella phaffii*. *Biotechnol. Lett.* **2018**, *40*, 1365–1376. [CrossRef] [PubMed]
48. Gangaraju, V.K.; Bartholomew, B. Mechanisms of ATP dependent chromatin remodeling. *Mutat. Res.* **2007**, *618*, 3–17. [CrossRef]
49. Yusa, K.; Zhou, L.; Li, M.A.; Bradley, A.; Craig, N.L. A hyperactive piggyBac transposase for mammalian applications. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 1531–1536. [CrossRef]
50. Nijkamp, E.; Ruffolo, J.A.; Weinstein, E.N.; Naik, N.; Madani, A. ProGen2: Exploring the boundaries of protein language models. *Cell Syst.* **2023**, *14*, 968–978.e963. [CrossRef]
51. Ivanci, D.; Agudelo, A.; Lindstrom-Vautrin, J.; Jaraba-Wallace, J.; Gallo, M.; Das, R.; Ragel, A.; Herrero-Vicente, J.; Higuera, I.; Billeci, E.; et al. Discovery and protein language model-guided design of hyperactive transposases. *Nat. Biotechnol.* **2025**. [CrossRef]
52. Wang, X.; Li, Y.; Jin, Z.H.; Liu, X.J.; Gao, X.; Guo, S.Y.; Yu, T. A novel CRISPR/Cas9 system with high genomic editing efficiency and recyclable auxotrophic selective marker for multiple-step metabolic rewriting in *Pichia pastoris*. *Synth. Syst. Biotechnol.* **2023**, *8*, 445–451. [CrossRef] [PubMed]
53. Zhang, K.; Duan, X.P.; Cai, P.; Gao, L.H.; Wu, X.Y.; Yao, L.; Zhou, Y.J. Fusing an exonuclease with Cas9 enhances homologous recombination in *Pichia pastoris*. *Microb. Cell Fact.* **2022**, *21*, 182. [CrossRef] [PubMed]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.