Development of synthetic biology tools to engineer *Pichia pastoris* as a chassis for the production of natural products

Jucan Gao, Lihong Jiang, Jiazhang Lian

A Key Laboratory of Biomass Chemical Engineering of Ministry of Education, College of Chemical and Biological Engineering, Zhejiang University, Hangzhou, 310027, China

B Hangzhou Global Scientific and Technological Innovation Center, Zhejiang University, Hangzhou, 310027, China

**A R T I C L E  I N F O**

**Keywords:**
- Pichia pastoris
- Natural products
- Synthetic biology
- CRISPR/Cas9
- Heterologous gene expression

**A B S T R A C T**

The methylotrophic yeast *Pichia pastoris* (a.k.a. *Komagataella phaffii*) is one of the most commonly used hosts for industrial production of recombinant proteins. As a non-conventional yeast, *P. pastoris* has unique biological characteristics and its expression system has been well developed. With the advances in synthetic biology, more efforts have been devoted to developing *P. pastoris* into a chassis for the production of various high-value compounds, such as natural products. This review begins with the introduction of synthetic biology tools for the engineering of *P. pastoris*, including vectors, promoters, and terminators for heterologous gene expression as well as Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated System (CRISPR/Cas) for genome editing. This review is then followed by examples of the production of value-added natural products in metabolically engineered *P. pastoris* strains. Finally, challenges and outlooks in developing *P. pastoris* as a synthetic biology chassis are prospected.

1. Introduction

Natural products generally refer to secondary metabolites isolated from animals, plants, and microorganisms, which have special physiological activities. Many high-value natural products play an extremely important role in medicine, cosmetics, and industry [1]. However, natural products generally have complex chemical structures and are extracted from sources with long growth periods and extremely low abundance, resulting in short supply and high cost, particularly those with plant origins [2]. With the development of synthetic biology, various microorganisms including *Pichia pastoris* (a.k.a. *Komagataella phaffii*) have been developed as cell factories to produce natural products [3–5]. Compared with bacterial cell factories (i.e. *Escherichia coli*), the ability of post-translational modifications and the presence of inner membrane systems make yeasts including *P. pastoris* preferred hosts to express eukaryotic complex proteins, such as cytochrome P450s (CYPs), which are often involved in the biosynthesis of natural products [6]. When compared with the model yeast *Saccharomyces cerevisiae*, *P. pastoris* has the advantage of strong and tightly regulated promoters for high level expression of recombinant proteins [7]. For example, the expression level of the target gene can account for more than 30% of the total proteins of *P. pastoris*, which is much higher than that in *S. cerevisiae* [8]. In addition, hyperglycosylation is another concern for expressing eukaryotic proteins in the *S. cerevisiae* expression system [9]. Natural product biosynthetic pathway enzymes (i.e. polyketide synthases and CYPs) are generally found to have relatively low catalytic activities, which should be expressed at high levels to achieve efficient biosynthesis. Therefore, *P. pastoris* is a promising host for large-scale production of natural products, particularly those with eukaryotic origins.

Although generally considered as a non-conventional yeast, genetics, physiology, and cell biology of *P. pastoris* have been studied in-depth [4]. The genomes of *P. pastoris* GS115 and CBS7435 have been sequenced and annotated, and genome-scale metabolic models have been constructed by analyzing the metabolic patterns [10]. Metabolomics studies indicate that the intermediate metabolites of *P. pastoris* and *S. cerevisiae* are very similar, with identical metabolites up to 90% [11]. The exploration of the genetic background of *P. pastoris* has laid a solid foundation for customized modification of *P. pastoris*. Currently, besides therapeutic proteins and enzymes [12], *P. pastoris* has been...
engineered to produce various chemicals and value-added compounds, such as \( \alpha \)-lactic acid [13], 2,3-butanediol (BDO) [14], 2-phenylethanol [15], isobutanol and isobutyl acetate [16], carotenoids [17], lovastatin [18], and nootkatone [19] (Fig. 1).

In this review, the synthetic biology tools needed for the construction and optimization of natural product biosynthetic pathways in \( P. \) pastoris are firstly summarized, in particular the heterologous gene expression system and the CRISPR/Cas (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated System) based genome editing system. Subsequently, several examples of the establishment of \( P. \) pastoris as cell factories for the production of terpenoids, polyketides, and flavonoids are introduced. Finally, future perspectives in the development of novel synthetic biology tools for the assembly and integration of multi-gene biosynthetic pathways and high throughput genome engineering are discussed.

2. Synthetic biology toolkit for \( P. \) pastoris

2.1. Gene expression vectors

The most common way to introduce exogenous genes into \( P. \) pastoris is to construct recombinant vectors. The method of plasmid maintenance in yeast is through auxotrophic markers or resistance selection markers (Table 1) [20]. Plasmids can be divided into episomal plasmids and integrative plasmids according to the way existing in the host. Unfortunately, the episomal plasmids suffer from low stability and genome integration is generally preferred for high level expression of heterologous genes in \( P. \) pastoris. Usually the vector is linearized and integrated into the \( P. \) pastoris genome in a single copy manner. For example, the vectors pPIC5K-His and pHIL-S1 can be linearized by \( \text{Sal} \) for the integration into the \( \text{HIS4} \) locus of \( P. \) pastoris GS115, which are then screened in histidine-dropout medium to obtain single-copy integrated strains [21,22]. In addition to single-copy integration, multi-copy integration is generally demanded for high-level expression of the target genes. The pPIC series of vectors are commonly used integrative vectors in \( P. \) pastoris [23-26], which enable the screening of the multi-copy integration strains under high concentration of antibiotics, a mechanism known as post-transformation amplification [27-30]. In addition to the formation of tandem repeats via post-transformation amplification, multi-copy strains can be constructed by integrating into the repetitive sequences of the \( P. \) pastoris genome, such as the ribosomal DNA (rDNA) sequences [13].

Nevertheless, episomal expression possesses unique advantages for several applications, such as the combinatorial optimization of multi-gene biosynthetic pathways and the development of efficient CRISPR-based genome editing tools [31]. In this case, a set of episomal vectors with various autonomously replicating sequences (ARSs) have been constructed and systematically compared for their transformation efficiency, copy numbers, and reproductive stability (Table 1) [32]. Of a particular note, panARS, a broad host ARS derived from \( Kluveomyces \) lactis, was found to enable the highest plasmid stability and chosen for the development of an efficient CRISPR/Cas9 system for \( P. \) pastoris [33].

2.2. Promoters and terminators

Promoters are considered as the most important synthetic biology elements and have direct impacts on the expression of the transcription units. The selection of appropriate promoters with the desirable strength is essential to construct well-controlled synthetic biology modules and to achieve optimal expression of the target genes. The alcohol oxidase 1 promoter (pAOX1) and the glyceraldehyde 3-phosphate dehydrogenase promoter (pGAP) are two most commonly used promoters [49]. The AOX1 promoter is generally considered as the strongest promoter of \( P. \) pastoris, which is strongly induced by methanol and inhibited by glycerol, ethanol, and glucose. Under the full induction conditions, Aox1p accounted for more than 30% of the total cellular proteins (Table 2) [50]. The GAP promoter is a strong constitutive promoter, whose expression strength is relatively stable. The expression level of some foreign proteins under the control of pGAP can reach up to the level of g/L [51].

To explore metabolic engineering and synthetic biology applications, a series of promoters with different properties have been characterized and engineered [52]. Based on RNA-seq results, p0188 (Pyruvate decarboxylase isozyme, locus tag: \( \text{PAS}_{\text{chr3}} \) 0188) was determined to be the strongest constitutive promoter and Aslan et al. found that p0188 had a strong driving force and the strength could reach up to 2-fold as that of pGAP [53]. By combining RNA-seq and mRNA folding free energy, 16 promoter candidates were selected and characterized with glucose, glycerol, or methanol as the sole carbon source, respectively. The p0547 (peroxidase promoter, locus tag: \( \text{PAS}_{\text{chr1-4}} \) 0547) and p0472 (mitochondrial alcohol dehydrogenase isozyme III, locus tag: \( \text{PAS}_{\text{chr2-1}} \) 0472) were determined to be the strongest methanol-inducible and constitutive promoters to express \( \alpha \)-amylase, respectively [49]. Interestingly, in another separate study, Karaoglan et al. found that the ADH3 promoter (locus tag: \( \text{PAS}_{\text{chr2-1}} \) 0472) performed even better than the AOX1 promoter when using methanol as the sole carbon source to express \( \text{Aspergillus niger} \) xylanase [54].

In addition to the mining and characterization of endogenous

---

**Fig. 1.** \( P. \) pastoris as a cell factory for the production of natural products. \( P. \) pastoris converts various carbon sources (i.e. methanol, glycerol, and glucose) into a few central metabolites (i.e. pyruvate and acetyl-CoA), which serve as the precursors for the biosynthesis of a variety of natural products (i.e. terpenoids, polyketides, and flavonoids).
features. For example, at least 12 cis-acting elements were found in the promoters, existing promoters can be modified to possess the desirable features. For example, at least 12 cis-acting elements were found in the AOX1 promoter and a promoter library of pAOX1 is constructed by deleting and replicating the putative transcription factor binding modules. The activity range of the promoter library is 10%–160% of the wild-type pAOX1 [55]. A key transcription factor, called Mxr1p, was.

Table 1
Gene expression vectors commonly used in *P. pastoris.*

| Plasmid backbone | ARS or integration locus | Copy numbers | Selection markers | Episomal or Integrative | Reference |
|------------------|-------------------------|--------------|------------------|--------------------------|-----------|
| pHIL-S1          | HIS4                    | 1.0          | HIS4             | Integrative              | [21]      |
| pPIC9X <sup>a</sup> or pPIC9K-His <sup>a,b</sup> | HIS4/AOX1            | 1.0 or 4.0-6.0 | HIS4/G418<sup>b</sup> | Integrative              | [23,34,37]|
| pPIC3.5K         | HIS4/AOX1              | 1.0          | HIS4/G418<sup>b</sup> | Integrative              | [24]      |
| pGAPZ <sup>c</sup>         | GAP                    | 1.0 or 21   | Zeocin<sup>c</sup> | Integrative              | [14,27,38]|
| pPICZ            | AOX1                    | 1.0 or 10.0 | Zeocin<sup>c</sup> | Integrative              | [25,26,28]|
| pAOX15          | HIS4/AOX1              | 1.0          | HIS4             | Integrative              | [39]      |
| pPink HC or pPink LC | TRP2                  | 2.0-4.0     | ADE2             | Integrative              | [40,41]   |
| pGLY2664 <sup>d</sup> | TRP2                  | 2.0-12.0    | ADE2/Zeo<sup>c</sup> | Integrative              | [42]      |
| pUC19           | rDNA                   | −9.0        | Zeocin<sup>c</sup> | Integrative              | [13]      |
| pPIC9Xa <sup>b</sup> | AOX1                  | −10.0       | Zeocin<sup>c</sup> | Integrative              | [29]      |
| pMCID <sup>e</sup> | HIS4                  | 2.0-16.0    | G418<sup>b</sup> | Integrative              | [30]      |
| pGHyB           | GAP                    | −3.0        | Hygromycin<sup>c</sup> | Integrative              | [43]      |
| pPIC9K <sup>f</sup> | panARS               | −18.0       | HIS4             | Episomal                 | [32]      |
| pPIC9K <sup>f</sup> | PAR1                  | −15.0       | HIS4             | Episomal                 | [32]      |
| pPIC9K <sup>f</sup> | mitsARS              | −14.0       | HIS4             | Episomal                 | [32]      |
| pPIC9K <sup>f</sup> | PpaARS               | −4.0        | HIS4             | Episomal                 | [32]      |
| pPIC9K <sup>f</sup> | SacARS               | −4.0        | HIS4             | Episomal                 | [32]      |
| pSEC-SUMO       | panARS               | −19.0       | Zeocin<sup>c</sup> | Episomal                 | [33]      |
| pMito           | mtDNA                  | −3.0        | HIS4             | Episomal                 | [44]      |
| pPICGbnBH <sup>g</sup> | PARS1                | −7.0        | Zeocin<sup>c</sup> | Episomal                 | [65]      |
| pBGPl <sup>h</sup> | PARS1                | N.A.        | Zeocin<sup>c</sup> | Episomal                 | [46,47]   |
| pPHEn<sup>i</sup> | PARS1                | N.A.        | HIS4             | Episomal                 | [48]      |

Note.  
<sup>a</sup> These plasmids contain a signal peptide for protein secretion.  
<sup>b</sup> pPIC9X is used for the expression of recombinant α-amylase, while 22 times stronger when expressing xylanase.  
<sup>c</sup> pAOX1 promoters commonly used for the expression of heterologous genes.  
<sup>d</sup> High-copy strains (up to 21 copies) were constructed by integrating the expression cassettes into the ribosomal DNA loci and screening under high Zeocin concentration conditions. Insert the HIS4 gene into the plasmid to obtain a single copy strain.  
<sup>e</sup> With pPIC9K as the backbone, a series of episomal plasmids were constructed by adding different yeast replicons (ARSs).

Table 2
*P. pastoris* promoters commonly used for the expression of heterologous genes.

| Promoter | Type             | Relative Strength<sup>i</sup> | Locus tag | Reference |
|----------|------------------|-------------------------------|-----------|-----------|
| pAOX1    | Methanol inducible | −18                           | /         | PAS_chr4_0821 [49] |
| pAOX176  | Methanol inducible | −26                           | /         | /         | PAS_chr4_0821 [55] |
| pAOX373−<sub>D</sub>+3D | Methanol inducible | −28                           | /         | /         | PAS_chr4_0821 [56] |
| pAOX2    | Methanol inducible | 1.8 ± 0.2                     | /         | PAS_chr4_0152 [52] |
| pAOX2-mutant<sup>a</sup> | Methanol inducible | −18                           | /         | /         | PAS_chr4_0152 [58] |
| pCAT1    | Methanol inducible | −18                           | /         | PAS_chr2-2_0131 [62] |
| pCAT1-mutant<sup>a</sup> | Methanol inducible | −23                           | /         | PAS_chr2-2_0131 [62] |
| pDA52    | Methanol inducible | 24.1 ± 3.3                    | /         | PAS_chr3_0834 [53] |
| pFLD1    | Methanol inducible | 10.9 ± 1.1                    | /         | PAS_chr1_0128 [52] |
| pPM200   | Methanol inducible | 16.4 ± 3.3                    | /         | PAS_chr1-4_0547 [52] |
| pDA51    | Methanol inducible | 14.7 ± 2.7                    | /         | PAS_chr3_0832 [52] |
| pFDH1    | Methanol inducible | 14.2 ± 2.2                    | /         | PAS_chr3_0952 [52] |
| p3074    | Methanol inducible | −4                            | /         | PAS_chr3_0374 [49] |
| p3019    | Methanol inducible | −3                            | /         | PAS_chr1-1_0319 [49] |
| p5047    | Methanol inducible | −2                            | /         | PAS_chr1-4_0547 [49] |
| pDM72    | Constitutive      | −3 or 22                      | −0.5      | PAS_chr2-1_0472 [49,54] |
| pGAP     | Constitutive      | 1.0                           | −2.3      | PAS_chr2-1_0437 [49] |
| p0769    | Constitutive      | −0.4                          | −1.8      | PAS_chr2-1_0769 [49] |
| p0072    | Constitutive      | −0.3                          | −1.4      | PAS_chr1-1_0872 [49] |
| pFDC<sup>b</sup> | Constitutive      | −0.5                          | −3.5 or 4.7 | PAS_chr3_0188 [49,53] |
| pFDI<sup>c</sup> | Constitutive      | −0.5                          | −0.4      | PAS_chr3_0844 [63] |

Note.  
<sup>a</sup> Promoter strengths are normalized to that of the constitutive promoter (pGAP) under methanol conditions.  
<sup>b</sup> pAOX176 is generated by removing a small number of bases before the TATA box of pAOX1.  
<sup>c</sup> pAOX176 represents that the bases from −940 to −737 bp of pAOX1 are deleted. Region D is defined as the position of −638 to −510 bp in pAOX1. -<sub>D</sub>+3D means the deletion of region D followed by the addition of 3 copies of region D at the 5’-end of pAOX1.  
<sup>d</sup> pAOX2-mutant is obtained by mutating the bases at −255 and −456 positions of pAOX2.  
<sup>e</sup> pCAT1-mutant represents that the bases from −433 to −411 bp of pCAT1 are duplicated.  
<sup>f</sup> pDM72 is 3 times stronger than pGAP when expressing recombinant α-amylase, while 22 times stronger when expressing xylanase.  
<sup>g</sup> pFDC is 1.5 times stronger than pGAP when expressing recombinant α-amylase, while 2 times stronger when expressing human growth hormone.  
<sup>h</sup> pFDI is constructed by deleting and replicating the putative transcription factor binding modules. The activity range of the promoter library is 10%–160% of the wild-type pAOX1 [55]. A key transcription factor, called Mxr1p, was meaningless.
determined to be essential for the induction of pAOX1 by methanol [56]. Due to the toxicity of methanol, Shen et al. established a methanol-free expression system using AOX1 promoter mutants, in which the mutant promoter was induced by dihydroxyacetone and suppressed by glucose. The strength of this protein expression system can reach 50%–60% of the wild-type pAOX1 expression system [57]. Dai et al. identified a pAOX2 mutant, whose expression level was even higher than that of pAOX1. DNA sequencing revealed two point mutations at positions of −529 bp and −255 bp responsible for the dramatically improved promoter strength [58].

Terminators have been found to have important regulatory effects on transcription termination and the half-life of mRNA in S. cerevisiae [59, 60]. However, the significance of terminators is largely overlooked and little work has been done on the characterization of P. pastoris terminators. Vogl et al. tested the effect of different terminators on the expression of eGFP (enhanced green fluorescent reporter protein) under the control of AOX1 promoter and AOX1 terminator was found to enable the highest fluorescence intensity. In addition, inserting NotI restriction site into the AOX1 terminator can further increase the fluorescence intensity by 37% [52]. Ito et al. characterized 72 terminators derived from P. pastoris, S. cerevisiae, and synthetic terminators, and found that the tunable range could reach up to 17-fold. Interestingly, the S. cerevisiae terminators could maintain their function after being transferred to P. pastoris [61]. These preliminary studies indicated the significance of terminators in regulating the expression level of heterologous genes and more mechanistic studies should be carried out in the near future.

2.3. Genome editing tools

As a fundamental tool, genome editing technology is essential for establishing P. pastoris as cell factories for recombinant proteins and value-added compounds. In the very beginning, site-directed gene integration and gene knockout were achieved through homologous recombination. Construction of a selection marker-containing plasmid that is capable of gene replacement in P. pastoris is one of the first genome editing tools [64]. For example, HIS4, URA3, and URA5 genes are often used as selection markers in the corresponding defective P. pastoris strains [65]. However, these genome editing techniques usually leave selection marker expression cassettes in the host, which is not desirable for subsequent genetic manipulations and industrial applications. To enable multiple rounds of genome editing, Cre/loxP system was introduced into P. pastoris. Cre is a site-specific recombinase that specifically recognizes and recombines genes between two loxP loci. The advantage of this technology is that antibiotic resistance genes can be used for screening first and then recycled after the disruption of the target gene [66]. In addition, mazF, a toxic gene from E. coli, was used to construct a set of counter-selection techniques for marker-less genome editing in P. pastoris [67].

In recent years, emerging genome editing tools, such as ZFN (Zinc-finger nucleases), TALEN (transcription activator-like effector nucleases), and CRISPR/Cas, have revolutionized our capability of genetic manipulations of microbial cell factories (Fig. 2). These technologies use specific nucleases to create double-strand breaks (DSB) at the corresponding loci, which are repaired by homologous recombination (HR) or non-homologous end joining (NHEJ) to achieve the desirable genome editing. Particularly, the CRISPR/Cas9 system is the most widely used and most powerful genome editing technology. The CRISPR/Cas9 expression system is derived from the immune defense systems of bacteria and archaea [68], and has received in-depth research in microbial cell factories development, plant breeding, animal breeding, disease modeling, and biotherapy [69]. Weninger et al. systematically optimized the CRISPR/Cas9 expression system to achieve efficient and precise genome editing in P. pastoris, including but not limited to Cas9 coding sequences, gRNA sequences, gRNA structures (i.e. with ribozyme sequences), and

![Fig. 2. Three emerging genome editing techniques, including ZFN, TALEN and CRISPR/Cas9.](image-url)
promoters for the expression of Cas9 and gRNAs [70]. Among 95 combinations, only 6 constructs were found to be functional for genome editing, indicating the necessity for further optimization (Table 3). For example, Gu et al. found that the replacement of the origin of replication of the bearing plasmid from PARS1 to panARS increased the disruption efficiency of ADE2 locus from ~10% to ~80% [32]. In addition, Dalvie et al. developed a sequencing-based strategy for the design of host-specific cassettes for modular and efficient expression of gRNAs and achieved high genome editing efficiency up to 95% [71]. Besides gene disruption, multiplex integration of heterologous genes is another essential synthetic biology tool for establishing P. pastoris as cell factories for natural products. Simultaneous integration of multiple genes was reported in a Ku70-deficient P. pastoris strain, with an integration efficiency ranged from 57.7% to 70% and 12.5%–32.1% for double- and triple-loci, respectively [72].

3. Engineering of P. pastoris to produce natural products

3.1. Terpenoids

Terpenoids are value-added natural products derived from mevalonate and widely existed in nature, including but not limited to higher plants, fungi, and microorganisms. Many terpenoids have been found applications in medicine, food, cosmetics, animal feeds, and industry, leading to the exploration of the production of terpenoids using microbial cell factories. Bhataya et al. introduced the lycopene biosynthetic pathway into non-carotenogenic P. pastoris for the first time. Two lycopene-pathway plasmids were constructed, with plasmid pGAPZB-EbBI harboring the same set of genes with a peroxisomal targeting sequence (PTS1). Similar amount of lycopene was produced in the two yeast strains, indicating that the supply of FPP might be limited in P. pastoris. One clone expressing pGAPZB-EbBI with the highest lycopene production was identified and further optimized by investigating the effects of cultivating conditions (i.e. carbon sources and aeration). Finally, the proportion of lycopene reached up to 73.9 mg/g in the basic medium with glucose as the carbon source [77]. Later, β-carotene was synthesized by additionally integrating the lycopene-β-cyclase gene from Ficus carica into the chromosomes of the lycopene-producing strain, leading to the production of 339 μg of β-carotene per gram dry cell weight (DCW) [17]. Starting from the β-carotene-producing strain, further introduction of β-carotene ketolase gene (crtW) and β-carotene hydroxylase gene (crtZ) from Agrobacterium aurantiacum resulted in the production of 3.7 μg/g DCW of astaxanthin in P. pastoris [78]. In another study, Vogl et al. characterized a panel of promoters in the methanol utilization pathway of P. pastoris, which were further employed for combinatorial optimization of the β-carotene biosynthetic pathway. With different combinations of the methanol inducible promoters, the production of β-carotene can be varied for more than 10-fold. Via choosing appropriate promoters from the established promoter library, the yield of β-carotene reached up to 5 mg/g DCW [52].

(+)–Nootkatone, an excellent fragrance and insect repellent, have also been successfully produced in P. pastoris. The introduction of valencene synthase resulted in the biosynthesis of (+)-valencene. Followed by the co-expression of the prenaspirodiene oxygenase from Hyoscyamus muticus (HPO) and the cytochrome P450 reductase from Arabidopsis thaliana, (+)-valencene was hydroxylated to produce trans-nootkatol. Trans-nootkatol was then oxidized to (+)-nootkatone by the intrinsic activity of P. pastoris. The production of (+)-nootkatone was 17 mg/L in a shake flask and 208 mg/L in a bioreactor, respectively [19]. Interestingly, the overexpression of RAD52, which is responsible for DNA repair and recombination, improved the production of trans-nootkatol by 5-fold [79].

Dammarenediol-II is a triterpenoid with multiple pharmacological activities. On the basis of the natural triterpene biosynthesis pathway [80,81], Liu et al. introduced PgDDS from Panax ginseng, encoding a dammarenediol-II synthase that catalyzed the production of dammarenediol-II from 2,3-oxidosqualene, to successfully construct a dammarenediol-II producing P. pastoris strain (Fig. 3). By increasing the expression of ERG1 to enhance the supply of 2,3-oxidosqualene and downregulating the expression of ERG7 to decrease the production of lanosterol from 2,3-oxidosqualene, the yield of dammarenediol-II was increased from 0.03 mg/g DCW to 0.736 mg/g DCW. Finally, by extra supplementation of 0.5 g/L squalene into the culture medium, the yield of dammarenediol-II reached up to 1.073 mg/g DCW.

Similarly, Sun et al. established a menaquinoine-4 (MK-4) P. pastoris cell factory by introducing a heterologous gene encoding Homo sapiens UBIAD1 (HsUBIAD1), which can produce MK-4 from phylloquinone (VK1) or menadione (VK3). HsUBIAD1 was cloned into pGAPZA (with the constitutive promoter pGAP) and pPICZA (with the inducible promoter pAOX1) and the effect of promoters on the expression of the target gene was investigated. It was found that the vector pGAPZA (with the target gene HsUBIAD1 under the control of pGAP) resulted in higher protein expression level. Then the geranylgeranyl pyrophosphate synthase gene (GGPPS) from Sulfolobus acidocaldarius was fused with the endogenous isopentenyl diphosphate isomerase gene (IDI1), and the resultant ID1-GGPPS chimeric gene was integrated into the 28S rDNA loci in a multi-copy manner using a modified integrative vector (pGrG, based on pGAPZA). In combination with the optimization of the fermentation conditions (i.e. pH and temperature) resulted in the maximum yield of MK-4 up to 0.24 mg/g DCW [82].

Table 3

| Cas9 promoter | sgRNA promoter, promoter type | Host | Target(s) | Donor length | Efficiency | References |
|---------------|-------------------------------|------|-----------|--------------|------------|------------|
| pHITX1        | pHITX1, II                    | CBS7435 | GUT1       | 1000 bp      | 87–94%     | [70]       |
| pDN01         | pRNA-dRNA1, III               | NBRRL Y-11430 | GUT1 | 500 bp      | 95%        | [71,72]    |
| pHITX1        | pHITX1, II                    | GS115 Δku70 | 2 lociα | 1000 bp      | 57.7–70%   | [72]       |
| pHITX1        | pHITX1, II                    | GS115 Δku70 | 3 lociα | 1000 bp      | 12.5–32%   | [72]       |
| pGAP          | pHITX1, II                    | GS115 | MRK1      | –600 bp    | >80%       | [74]       |
| pGAP          | pSER, III                     | GS115 | ADE2      | 250 bp      | 80%        | [32]       |
| pGAP          | pHITX1, II                    | GS115 | Gt1       | None        | 100%       | [31]       |
| pHITX1        | pHITX1, II                    | CBS7435 Δku70 | GUT1 | 1000 bp      | 78–91%     | [75]       |
| pHITX1        | pHITX1, II                    | CBS7435 Δku70 | GUT1 | 1000 bp      | 80–95%α    | [75]       |
| pHITX1        | pHITX1, II                    | CBS7435 Δku70 | GUT1 | 1000 bp      | 100%β      | [75]       |
| pHITX1        | pHITX1, II                    | KFM1 | PDC1      | 1000 bp     | N.A        | [76]       |

α Any two loci of pAOX1, pFDL1, and pTEFI were simultaneously targeted.
β pAOX1, pFDL1, and pTEFI were simultaneously targeted. None means that no donor was added and DSB was repaired by NHEJ during CRISPR editing.
γ Integration efficiency using the canonically designed DNA donor.
αβ Integration efficiency using DNA donor with a selection marker (zeocin’).
β Integration efficiency using DNA donor with a replicon (ARS).
3.2. Polyketides

Polyketides are a class of secondary metabolites produced by bacteria, fungi, plants, and animals and the most important source of natural product-based drugs. 6-Methylsalicylic acid (6-MSA) is the first polyketide produced by \textit{P. pastoris}. The 6-MSA biosynthetic pathway consisting of the phosphopantetheinyl transferase (PPTase) gene from \textit{A. nidulans} and the 6-MSA synthase (6-MSAS) gene from \textit{A. terreus} was successfully reconstituted in \textit{P. pastoris}. After methanol induction, the production of 6-MSA was up to 2.2 g/L in 20 h in a 5 L bioreactor, which established \textit{P. pastoris} as a promising cell factory for future industrial production of fungal polyketides [83].

![Fig. 3. Biosynthetic pathway of dammarenediol-II and MK4 in \textit{P. pastoris}. This study was carried out on the basis of the natural triterpene synthesis pathway. By introducing the exogenous \textit{PgDDS} gene, encoding a dammarenediol synthase, the target compound dammarenediol-II was produced from 2,3-oxidosqualene in \textit{P. pastoris}. The green arrow indicated the down-regulated gene (\textit{ERG7}) and the red arrow indicated the overexpressed genes. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)]

![Fig. 4. Biosynthetic pathway for lovastatin and simvastatin. Heterologous genes integrated into the genome of \textit{P. pastoris} were shown in red. \textit{lovB} and \textit{lovF}: two PKS genes; \textit{lovC}: an enoyl-reductase gene; \textit{lovG}: a thioesterase gene; \textit{lovA}: a cytochrome P450 monooxygenase gene; and \textit{lovD}: an acyl-transferase gene. \textit{NpgA} is from \textit{A. nidulans}. \textit{lovB}, \textit{lovC}, \textit{lovF}, \textit{lovG}, and \textit{CPR} were amplified from the \textit{A. terreus} genome. \textit{lovA} and \textit{lovD} are synthetic and codon-optimized DNA sequences for \textit{lovA} and \textit{lovD}, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)]
Recently, *P. pastoris* was engineered for de novo biosynthesis of citrinin, a value-added compound. The structure and biosynthetic pathway of citrinin are more complicated than 6-MSA, serving as an excellent model compound for further investigations [84,85]. Besides the citrinin polyketide synthase gene PksCT (CitS) from *Monascus purpureus* and the phosphopantetheinylltransferase gene NpgA from *A. nidulans*, the citrinin gene cluster from *M. purpureus*, including a serine hydrolase gene MPL1 (CitA), an oxygenase gene MPL2 (CitB), a dehydrogenase gene MPL4 (CitD), and other two intron-removed genes MPL6 (CitE) and MPL7 (CitC), was introduced to enable citrinin biosynthesis in *P. pastoris*. After 24 h induction with methanol, the yield of citrinin reached up to 0.6 ± 0.1 mg/L [86].

Production of monacolin J and lovastatin is another classic example of the production of polyketides using *P. pastoris* (Fig. 4). Seven enzymes, including lovastatin nonaketide synthase (LovB), enoyl reductase (LovC), thioesterase (LovG), a cytochrome P450 enzyme (LovA) together with a cytochrome P450 reductase (CPR), and cytochrome LovD (LovD) from *A. terreus*, as well as phosphopantetheinyl transferase (PpTase or NpgA) from *A. nidulans*, were heterologously expressed in *P. pastoris*. The expression of all these genes were driven by PpAOX1, LovB, LovC, LovG, and NpgA expression cassettes were cloned into the vector pPICZ B. LovA and CPR were cloned into the vector pPIC3.5K. After these two vectors were linearized and integrated into the GS115 genome, the recombinant strains were screened using zeocin-containing plates and histidine-deficient plates, respectively. Under pH-controlled fermentation (i.e. carbon sources), the periodic hydrogen peroxide shock strategy was employed to further increase the production of 3′-hydroxygenisterin to 20.3 mg/L in a 5 L fermenter [91].

4. Conclusions and perspectives

Many synthetic biology tools have been developed to precisely control the expression of heterologous genes and the assembly and integration of multi-gene pathways in *P. pastoris* [75]. However, the development of *P. pastoris* cell factories for natural products is still limited to a few examples [82], particularly when compared with that of *S. cerevisiae*, indicating a need to develop novel synthetic biology tools. For example, the most commonly used promoters in *P. pastoris* are still pAOX1 and pGAP [51], while the construction of efficient cell factories generally requires to precisely control the expression levels of biosynthetic pathway genes. In this case, promoters with different strength (weak, medium, and strong promoters) should be characterized under various fermentation conditions (i.e. carbon sources). In addition, most of exogenous genes are currently integrated into the HIS4 or AOX1 locus of the *P. pastoris* genome, while natural product biosynthetic pathways generally contain multiple genes [72]. In other words, well-characterized integration sites are highly demanded to assemble natural product biosynthetic pathways. Ideally, the integration sites should enable efficient and stable expression of heterologous genes, without affecting cell fitness. Although CRISPR/Cas9 system has been established for genome

![Fig. 5.](image_url) The ortho-hydroxylation of five flavonoids. CYP57B3-CPR is a fusion protein consisting of CYP57B3 from *A. oryzae* and the cytochrome P450 reductase (CPR) from *S. cerevisiae.*
Due to the complexity of cellular metabolic network, systems level understanding is generally a prerequisite to establish efficient \textit{P. pastoris} cell factories. Based on the genome sequencing results of \textit{P. pastoris} DSMZ 70382 [93] and GS115 [94], genome-scale metabolic models (GEMs), PpaMBEL1254 [95], iPP668 [96], and ilc915 [97] have been established. Tomas-Gamisans et al. reconstructed and verified a new consensus model IMT1026 based on these three early reports of GEMs. New discoveries related to glycosylation, fatty acid metabolism, and cell energy were complemented to the new model. Growth rate, carbon dioxide production, arabbitol production, and other parameters predicted by the model IMT1026 were consistent with the experimental results, confirming that the prediction and simulation capabilities have been improved [98]. The \textit{P. pastoris} GEMs have been successfully implemented to identify targets to increase the production of recombinant proteins. Nocon et al. predicted 9 engineering targets based on GEMs, 5 of which significantly increased the production of cytosolic human superoxide dismutase (hsSOD) [99]. Cankour-Cetinkaya et al. used GEMs to analyze the metabolic burdens caused by heterologous protein synthesis and found that supplementation of tyrosine to the culture medium could increase the yield of human lysozyme and antibody fragment Fab-3H6 [100]. Currently, the production of natural products is still mainly limited by the low efficiency of the biosynthetic pathways, and there is a lack of examples on the application of GEMs in improving the production of natural products in \textit{P. pastoris}. Nevertheless, once the bottleneck of the pathway enzymes has been addressed and the titer reaches to a certain level, the supply of the precursors and cofactors will become rate-limiting and GEMs can play a more important role in guiding the design of efficient yeast cell factories.

As a non-model yeast strain, our understanding of the metabolic and regulatory networks is still rather limited for \textit{P. pastoris}. For example, Wiesznesgger et al. found that the overexpression of \textit{RAD52}, encoding a protein responsible for DNA repair and recombination, significantly improved the production of trans-nootkatal [79]. Therefore, it is highly desirable to develop genome-scale metabolic engineering strategies that can perturb all the genes at once, with an aim to identify non-intuitive engineering targets to improve the production of the desirable compounds and map genotype-phenotype relationships in a high-throughput manner [101]. Currently, CRISPR-based genome-scale engineering has been well established in \textit{E. coli} [102], \textit{S. cerevisiae} [103–105], and mammalian cells [106]. Nevertheless, several challenges should be addressed to establish genome-scale engineering for \textit{P. pastoris}, such as an efficient CRISPR system, high transformation efficiency, as well as high throughput screening.

In summary, great progress has been made in genetic manipulation (i.e. heterologous gene expression system and genome editing system) of \textit{P. pastoris} for the production of recombinant proteins and value-added compounds. Nevertheless, more efforts should be devoted to developing novel synthetic biology tools (i.e. multi-gene pathway assembly and multiplex genome engineering) to establish \textit{P. pastoris} as a robust chassis for the biosynthesis of natural products.

CRediT author statement

Juncan Gao and Lihong Jiang drafted the manuscript. Jiazhang Lian conceived the review idea and revised the manuscript. All authors approved the manuscript.

Declaration of competing interest

None.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (2018YFA0901800), the Natural Science Foundation of China (21808199), and the Natural Science Foundation of Zhejiang Province (LR20B060003).

References

[1] Rodrigues T, et al. Counting on natural products for drug design. Nat Chem 2016; 8:531–41.
[2] De Luca V, et al. Mining the biodiversity of plants: a revolution in the making. Science 2012;336:1658–61.
[3] Pena DA, et al. Metabolic engineering of \textit{Pichia pastoris}. Metab Eng 2018;50:2–15.
[4] Schwartzhans JP, et al. Towards systems metabolic engineering in \textit{Pichia pastoris}. Biotechnol Adv 2017;35:681–710.
[5] Liu J, Mishra S, Zhao H. Recent advances in metabolic engineering of \textit{Saccharomyces cerevisiae}: new tools and their applications. Metab Eng 2018;50:85–108.
[6] Jiang L, et al. Functional expression of eukaryotic cytochrome P450s in yeast. Biotechnol Bioeng 2020.
[7] Creff MCI, Shi J, Higgins DR. Recombinant protein expression in \textit{Pichia pastoris}. Mol Biotechnol 2000:23–52.
[8] Duan X, Gao J, Zhou YJ. Advances in engineering methylotrophic yeast for biosynthesis of valuable chemicals from methanol. Chin Chem Lett 2018;29:66–70.
[9] Thak EJ, et al. Yeast synthetic biology for designed cell factories producing secretory recombinant proteins. FEMS Yeast Res 2020:20.
[10] Yang Z, Zhang Z. Engineering strategies for enhanced production of protein and bio-products in \textit{Pichia pastoris}: a review. Biotechnol Adv 2018;36:182–95.
[11] Carneiro M, et al. Development of quantitative metabolomics for \textit{Pichia pastoris}. Metabolomics 2012;8:284–98.
[12] Zhou Q, et al. High-level production of a thermostable mutant of \textit{Yarrowia lipolytica} lipase 2 in \textit{Pichia pastoris}. Int J Mol Sci 2019;21.
[13] Yamada R, et al. Toward the construction of a technology platform for chemicals production from methanol: D-lactic acid production from methanol by an engineered yeast \textit{Pichia pastoris}. World J Microbiol Biotechnol 2019;35:37.
[14] Yang Z, Zhang Z. Production of (2R, 3R)-2,3-butanediol using engineered \textit{Pichia pastoris}: strain construction, characterization and fermentation. Biotechnol Biofuels 2018;11:35.
[15] Kong S, et al. De novo biosynthesis of 2-phenylethanol in engineered \textit{Pichia pastoris}. Enzym Microb Technol 2020:103:109459.
[16] Siripong W, et al. Metabolic engineering of \textit{Pichia pastoris} for production of isobutanol and isobutyl acetate. Biotechnol Biofuels 2018;11:1.
[17] Araya-Garay JM, et al. Construction of new \textit{Pichia pastoris} X-33 strains for production of lycopene and beta-carotene. Appl Microbiol Biotechnol 2012;93:2483–92.
[18] Liu Y, et al. Engineered monoculture and co-culture of methylotrophic yeast for de novo production of monocerin J and lovastatin from methanol. Metab Eng 2018:45:189–99.
[19] Wiesznesgger T, et al. Production of the sesquiterpenoid (-)-nootkatone by metabolic engineering of \textit{Pichia pastoris}. Metab Eng 2014;24:18–29.
[20] Juergen N, Grantham N. In: Yeast strains for protein production. United States patent US 2010/0279,348.
[21] Yoshimaru MA, et al. Soluble expression and purification of porcine pepginon from \textit{Pichia pastoris}. Protein Expr Purif 2002;25:229–36.
[22] Ha SH, et al. Molecular cloning and high-level expression of G2 protein of hantavirus (HTN) virus 76–118 strain in the yeast \textit{Pichia pastoris} KM71. Virus Gene 2001;22:167–73.
[23] Wang Y, et al. Expression of \textit{Aspergillus niger} glucose oxidase in \textit{Pichia pastoris} and its antimicrobial activity against \textit{Agrobacterium} and \textit{Escherichia} coli. PeerJ 2020:8.e9100.
[24] Gaffar S, et al. Combination of genetic manipulation improved \textit{Saccharomycopsis fibuligera} α-amylase secretion by \textit{Pichia pastoris}. Indonesian Journal of Chemistry 2019:19.
[25] Khan M, et al. Molecular cloning and expression of recombinant \textit{Trichoderma harzianum} chitinase in \textit{Pichia pastoris}. Adv Life Sci 2020;7:122–8.
[26] Papakonstantinou T, Harris S, Hearne MT. Expression of GFP using \textit{Pichia pastoris} vectors with zeocin or G-418 sulphate as the primary selectable marker. Yeast 2009;26:311–21.
[27] Marx H, et al. Directed gene copy number amplification in \textit{Pichia pastoris} by vector integration into the ribosomal DNA locus. FEMS Yeast Res 2009;9:1260–70.
[28] Sunga AJ, Tolstoturov I, Creff M. Postransformational vector amplification in the yeast \textit{Pichia pastoris}. FEMS Yeast Res 2008;8:870–6.
[29] Aw R, Polizi KM. Liquid PTVA: a faster and cheaper alternative for generating multiple-copy clones in \textit{Pichia pastoris}. Microb Cell Factories 2016:15:26.
[30] Li D, et al. A novel vector for construction of markerless multicopy overexpression transfectants in Pichia pastoris. Front Microbiol 2017;8:1569.

[31] Yang Y, et al. High efficiency CRISPR/Cas9 genome editing system with an eliminable episomal sgRNA plasmid in Pichia pastoris. Enzym Microb Technol 2020:138.

[32] Gu Y, et al. Construction of a series of episomal plasmids and their application in the development of a highly efficient CRISPR/Cas9 system in Pichia pastoris. World J Microbiol Biotechnol 2019;35:79.

[33] Camattari A, et al. Characterization of a panARS-based episomal vector in the methylotrophic yeast Pichia pastoris for recombinant protein production and synthetic biology applications. Microb Cell Factories 2016:15:139.

[34] Zhang T, et al. The expression of recombinant human LOX-1 and identifying its mimicking ligands by fluorescence polarization-based high throughput screening. J Microbiol 2016;42:502.

[35] Zhou HY, et al. Heterologous expression and biochemical characterization of a monooxygenase in Pichia pastoris. Enzyme Microb Technol 2020;48:13000.

[36] Li X, et al. Overexpression of MFS in Pichia pastoris. Mol Biol Rep 2009;36:1611.

[37] Troshagina DS, et al. Cloning of phytase genes from Aspergillus niger CCTCC M 2018240 in Pichia pastoris. Yeast 2003;20:1279.

[38] Chen H, et al. Engineering of an episomal plasmid suitable for high-throughput recombinant protein production in Pichia pastoris. Mol Biotechnol 2019;129:152.

[39] Liu XB, et al. Metabolic engineering of Pichia pastoris for production of 5-Methyl-1-butanol. Enzym Microb Technol 2020;138:109557.

[40] Hou C, et al. Targeted editing of transcriptional activator MXR1 on the Pichia pastoris genome using CRISPR/Cas9 technology. Yeast 2020;37:305.

[41] Ding H-Y, et al. Identification of 3 phytase genes in Pichia pastoris. Curr Pharmaceut Biotechnol 2015;16:1085.

[42] Li Y-J, et al. Screening and characterization of natural antioxidants in four Glycyrrhiza species by liquid chromatography coupled with electrospray ionization quadrupole time-of-flight tandem mass spectrometry. J Chromatogr A 2013;1291:64.

[43] Chu LL, Montecillo JAV, Bae H. Recent advances in the metabolic engineering of yeasts for ginsenoside biosynthesis. Front Biotechnol 2020;9:254.

[44] Li YB, et al. Screening and characterization of natural antioxidants in four Glycyrrhiza species by liquid chromatography coupled with electrospray ionization quadrupole time-of-flight tandem mass spectrometry. J Chromatogr A 2013;1291:64.

[45] Wriessnegger T, et al. Enhancing cytochrome P450-mediated conversions in Pichia pastoris using the attenuated C. CCTCC M 2018240 in Pichia pastoris. Microb Cell Factories 2019;18:144.

[46] Shiraia bambusicola strain. J Biosci Bioeng 2008;106:466.

[47] Hou C, et al. Targeted editing of transcriptional activator MXR1 on the Pichia pastoris genome using CRISPR/Cas9 technology. Yeast 2020;37:305.

[48] Hou C, et al. Targeted editing of transcriptional activator MXR1 on the Pichia pastoris genome using CRISPR/Cas9 technology. Yeast 2020;37:305.

[49] Hou C, et al. Targeted editing of transcriptional activator MXR1 on the Pichia pastoris genome using CRISPR/Cas9 technology. Yeast 2020;37:305.

[50] Hou C, et al. Targeted editing of transcriptional activator MXR1 on the Pichia pastoris genome using CRISPR/Cas9 technology. Yeast 2020;37:305.

[51] Hou C, et al. Targeted editing of transcriptional activator MXR1 on the Pichia pastoris genome using CRISPR/Cas9 technology. Yeast 2020;37:305.

[52] Hou C, et al. Targeted editing of transcriptional activator MXR1 on the Pichia pastoris genome using CRISPR/Cas9 technology. Yeast 2020;37:305.

[53] Hou C, et al. Targeted editing of transcriptional activator MXR1 on the Pichia pastoris genome using CRISPR/Cas9 technology. Yeast 2020;37:305.

[54] Hou C, et al. Targeted editing of transcriptional activator MXR1 on the Pichia pastoris genome using CRISPR/Cas9 technology. Yeast 2020;37:305.

[55] Hou C, et al. Targeted editing of transcriptional activator MXR1 on the Pichia pastoris genome using CRISPR/Cas9 technology. Yeast 2020;37:305.

[56] Hou C, et al. Targeted editing of transcriptional activator MXR1 on the Pichia pastoris genome using CRISPR/Cas9 technology. Yeast 2020;37:305.

[57] Hou C, et al. Targeted editing of transcriptional activator MXR1 on the Pichia pastoris genome using CRISPR/Cas9 technology. Yeast 2020;37:305.

[58] Hou C, et al. Targeted editing of transcriptional activator MXR1 on the Pichia pastoris genome using CRISPR/Cas9 technology. Yeast 2020;37:305.

[59] Hou C, et al. Targeted editing of transcriptional activator MXR1 on the Pichia pastoris genome using CRISPR/Cas9 technology. Yeast 2020;37:305.

[60] Hou C, et al. Targeted editing of transcriptional activator MXR1 on the Pichia pastoris genome using CRISPR/Cas9 technology. Yeast 2020;37:305.

[61] Hou C, et al. Targeted editing of transcriptional activator MXR1 on the Pichia pastoris genome using CRISPR/Cas9 technology. Yeast 2020;37:305.

[62] Hou C, et al. Targeted editing of transcriptional activator MXR1 on the Pichia pastoris genome using CRISPR/Cas9 technology. Yeast 2020;37:305.

[63] Hou C, et al. Targeted editing of transcriptional activator MXR1 on the Pichia pastoris genome using CRISPR/Cas9 technology. Yeast 2020;37:305.
[97] Caspeta L, et al. Genome-scale metabolic reconstructions of Pichia stipitis and Pichia pastoris and in silico evaluation of their potentials. BMC Syst Biol 2012;6: 24.

[98] Tomas-Gamisans M, Ferrer P, Albiol J. Integration and validation of the genome-scale metabolic models of Pichia pastoris: a comprehensive update of protein glycosylation pathways, lipid and energy metabolism. PloS One 2016;11. e0148031.

[99] Nocon J, et al. Model based engineering of Pichia pastoris central metabolism enhances recombinant protein production. Metab Eng 2014;24:129–38.

[100] Cankorur-Cetinkaya A, et al. Process development for the continuous production of heterologous proteins by the industrial yeast. Komagataella phaffii. Biotechnol Bioeng 2018;115:2962–73.

[101] Xia S, et al. Advances in genome evolution of Saccharomyces cerevisiae. Synthetic Biology Journal 2020;1:556–69.

[102] Garst A, et al. Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. Nat Biotechnol 2017; 35:48–55.

[103] Bao Z, et al. Genome-scale engineering of Saccharomyces cerevisiae with single-nucleotide precision. Nat Biotechnol 2018;36:505–8.

[104] Lian J, et al. Multi-functional genome-wide CRISPR system for high throughput genotype-phenotype mapping. Nat Commun 2019;10:5794.

[105] Lian J, et al. Combinatorial metabolic engineering using an orthogonal tri-functional CRISPR system. Nat Commun 2017;8:1688.

[106] Gilbert LA, et al. Genome-scale CRISPR-mediated control of gene repression and activation. Cell 2014;159:647–61.