nature synthesis

Article

https://doi.org/10.1038/s44160-022-00205-2

Biosynthesis of catharanthine in engineered *Pichia pastoris*

Received: 5 January 2022

Accepted: 4 November 2022

Published online: 9 January 2023

Check for updates

Jucan Gao[1,2,3], Yimeng Zuo[1,2,3], Feng Xiao[2], Yiling Wang[2,4], Dongfang Li[2], Junhao Xu[1,2], Cuifang Ye[1,2], Linjuan Feng[1,2], Leijie Jiang[1], Tengfei Liu[1], Di Gao[1], Bin Ma[2,4], Lei Huang[1,2], Zhinan Xu[1] & Jiazhang Lian[1,2,3]

Catharanthine can be coupled with vindoline to synthesize vinblastine and vincristine, which have been used clinically as potent anticancer drugs. However, the structural complexity and low abundance in nature hamper bulk chemical synthesis and plant extraction, leading to a limited supply and a high cost of this plant natural product. Here, we engineer the methylotrophic yeast *Pichia pastoris* for complete biosynthesis of catharanthine from simple carbon sources. Through the selection of stable integration sites, screening of biosynthetic pathway enzymes with higher activity and/or specificity, amplification of flux-limiting enzyme encoding genes, rewiring of cellular metabolism and process optimization, we achieve de novo biosynthesis of catharanthine with a titre as high as 2.57 mg l^{-1}, which also represents the most complicated molecule heterologously synthesized in a non-model microorganism. Our study establishes *P. pastoris* as a cell factory for producing plant natural products with complex biosynthetic pathways.

Monoterpene indole alkaloids (MIAs), with a secoiridoid moiety and an indole moiety, include more than 3,000 diverse structures. MIAs are plant natural products with an important medicinal value[1], and include camptothecin[2] and vinblastine[3] as anticancer drugs, ajmalicine as a vascular disorder agent[4], reserpine with antihypertension properties[4] and quinine as an antimalarial drug[5]. Among these, vinblastine from the representative plant *Catharanthus roseus* attracts the greatest attention and has been used as one of the most potent drugs for treating several types of cancer. Vinblastine is a heterodimer of two MIA moieties, catharanthine and vindoline. Considering the complicated structures, the total chemical synthesis of most MIAs is rather challenging and suffers from low yield and high cost[6–8]. Currently, vinblastine is still mainly produced via plant extraction, which suffers from extremely low content (about fresh weight 0.0002%)[9,10], environmental concerns and unreliable supply. With the elucidation of pathways leading to the biosynthesis of catharanthine and vindoline in recent years[11–13], there is a growing interest in producing this high-value compound using synthetic biology approaches, if the challenges in manipulating multigene biosynthetic pathways and expressing rate-limiting enzymes to high levels can be readily addressed.

Considering the advantages in the availability of genetic engineering tools, the presence of inner membrane systems for functionally anchoring cytochrome P450 enzymes (CYPs), and the compatibility with high cell density and large-scale fermentation, the yeast *Saccharomyces cerevisiae* has often been regarded as the top option for constructing cell factories for plant natural products[14,15]. Several plant-derived alkaloids, such as opioids[16], noscapine[17], tropine[18] and scopolamine[19], have been successfully produced in *S. cerevisiae*. Nevertheless, low expression levels and hyperglycosylation are the major concerns for expressing plant proteins in *S. cerevisiae*[20], which may limit the production of several plant natural products. In particular, the enzymatic activity of precondylocarpine acetate synthase (PAS), a key enzyme of the strictosidine to tabersonine or catharanthine conversion pathway, was not detected when heterologously expressed

[1]Key Laboratory of Biomass Chemical Engineering of Ministry of Education, College of Chemical and Biological Engineering, Zhejiang University, Hangzhou, China. [2]ZJU-Hangzhou Global Scientific and Technological Innovation Center, Zhejiang University, Hangzhou, China. [3]Zhejiang Key Laboratory of Smart Biomaterials, Zhejiang University, Hangzhou, China. [4]Institute of Soil and Water Resources and Environmental Science, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou, China. e-mail: jzlian@zju.edu.cn

in *S. cerevisiae*[11]. Thus, there has been no report of the reconstitution of the vinblastine biosynthetic pathway, although the biosynthesis of strictosidine from carbon sources[21] and the conversion of tabersonine to vindoline[12,22] have been reported in *S. cerevisiae*. In other words, an alternative chassis to *S. cerevisiae* should be established to achieve de novo biosynthesis of vinblastine.

In contrast, the methylotrophic yeast *Pichia pastoris* (or *Komagataella phaffii*) possesses the basic advantages of the yeast system and demonstrates the capability of methanol utilization and much higher expression of heterologous genes than *S. cerevisiae*. For example, recombinant proteins driven by the alcohol oxidase 1 (*AOX1*) promoter can account for more than 30% of total protein[23]. Thus, *P. pastoris* has been widely used for the industrial production of recombinant proteins and potentially can be used to drive the expression of rate-limiting pathway enzymes, such as polyketide synthases and CYPs[14], for efficient biosynthesis of plant natural products. For example, a recent study reported the production of a terpenoid natural product to be much higher in *P. pastoris* than in *S. cerevisiae*[24]. Although the production of therapeutic proteins and enzymes has been explored extensively[25], there have been a few reports of the biosynthesis of natural products in *P. pastoris*, such as carotenoids[26], lovastatin[27] and nootkatone[28], probably due to the lack of synthetic biology tools for the manipulation of multigene biosynthetic pathways. Luckily, the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)–CRISPR-associated System (CRISPR/Cas) has been established[29–32] and a panel of neutral integration sites have been characterized in *P. pastoris*[33,34]. Although multiplex genome integration has been achieved with relatively high efficiency, the stability of the heterologous multigene pathways (for example, plant natural product biosynthetic pathways) has been rarely investigated in *P. pastoris*. Moreover, with the advantages in expressing rate-limiting pathway genes, the potential of *P. pastoris* in the production of plant natural products, particularly those with complicated and long biosynthetic pathways, has not been explored yet[35].

In this study, we aim to establish *P. pastoris* as a chassis for de novo biosynthesis of a vinblastine monomer catharanthine from simple carbon sources. First, we profile stable integration sites, featured by being flanked by two essential genes, in a genome-wide manner for construction and stable maintenance of multigene biosynthetic pathways. We then integrate and optimize the complete catharanthine biosynthetic pathway (Fig. 1) in a modular manner, with the whole biosynthetic pathway divided into three functional modules: (1) the CAN module, from strictosidine to catharanthine; (2) the STR module, from nepetalactol to strictosidine; and (3) the NPT module, from carbon sources to nepetalactol. Finally, we improve the production of catharanthine via combinatorial metabolic pathway engineering, including gene amplification to de-bottleneck the metabolic fluxes towards catharanthine, selecting appropriate plant enzymes with high activity and/or specificity, signal peptide engineering for correct localization of the key pathway enzymes, protein fusion expression to facilitate substrate trafficking and removal of competing metabolic pathways. More than 30 heterologous genes can be integrated and stably maintained in *P. pastoris* for complete biosynthesis of a value-added natural product (catharanthine) from simple carbon sources. The characterized stable integration sites can help install multigene pathways for producing other MIAs and the established *P. pastoris* cell factory promises the complete biosynthesis of vinblastine.

Results

Screening and characterization of stable integration sites

The construction of *P. pastoris* cell factories capable of synthesizing plant natural products such as vinblastine generally requires the integration of dozens of heterologous genes. Although the CRISPR/Cas system has been established[29–32] and a panel of neutral integration sites has been characterized[33] in *P. pastoris*, their applications in the construction and stable maintenance of complex and long biosynthetic pathways have not been well explored yet. Previous studies in *S. cerevisiae* found that integration into the sites flanked by essential genes was beneficial for the stability of heterologous genes and pathways[36]. Thus, we propose to profile the intergenic regions flanked by essential genes as potential stable integration sites in a genome-wide manner in *P. pastoris*. Unfortunately, *P. pastoris* is much less understood than *S. cerevisiae* and the essential genes have not been fully annotated. Luckily, Zhu et al. used transposon mutations and high-throughput sequencing to classify *P. pastoris* genes as putatively essential, ambiguous and putatively non-essential[37], represented by the values of non-gene essentiality (NEP) and read frequency (read density). Therefore, in addition to the well-accepted requirements (for example, high integration efficiency, decent expression level and minimal effect on cell growth), we set the following criteria for stable integration sites (Fig. 2a): (1) the sum of NEP values of the two flanked genes is less than 0.04; (2) the sum of read density values of the two flanked genes is less than 8.0; and (3) the spacing length of the intergenic region is larger than 600 bp.

Based on the above-mentioned criteria, we profiled the *P. pastoris* genome and identified 31 intergenic regions as the potential integration sites (Supplementary Table 1). With mCherry as a reporter, 24 out of the 31 intergenic regions resulted in an integration efficiency of >80% (Fig. 2b and Supplementary Fig. 1). Consistent with previous results in *P. pastoris* and other species, the expression level of heterologous genes was found to be locus dependent and could be varied as large as an order of magnitude (Fig. 2c). After serial transfers in non-selective medium for 30 generations, all the colonies were mCherry fluorescent, indicating the heterologous gene expression cassette was stably maintained in the genome of *P. pastoris* (Extended Data Fig. 1). We also evaluated multiplex genome integration efficiency and we were able to integrate reporter genes into three independent loci simultaneously with an efficiency higher than 60% (Extended Data Fig. 2).

Biosynthesis of catharanthine from strictosidine

The catharanthine biosynthetic pathway involves more than 25 heterologous genes (Fig. 1) and such a long and complex pathway has never been reconstituted in non-model strains such as *P. pastoris*. Therefore, we divided the catharanthine biosynthetic pathway into three functional modules (CAN, STR and NPT modules), then reconstituted and optimized the whole pathway in a modular manner. As a central intermediate of a wide variety of MIAs[38], strictosidine is the best branching point for module construction and optimization. In addition, for the vinblastine biosynthetic pathway, the conversion of strictosidine to catharanthine is the only module that has not yet been reconstituted in a heterologous host. Therefore, we chose to start working with the CAN module.

Enabled by the CRISPR/Cas9 system and the characterized integration sites, we were able to reconstitute the CAN module (CAN4A) in four rounds of genome editing. Because CYPs are involved in the biosynthesis of catharanthine, we integrated *CrCPR* (*C. roseus*) together with the CAN module genes, including *CrSGD*, *CrGS*, *CrGO*, *CrRedox1*, *CrRedox2*, *CrSAT*, *CrPAS*, *CrDPAS* and *CrCS*[11] (Fig. 3a and Supplementary Table 2). When strictosidine was extracellularly provided, the production of catharanthine was detected by liquid chromatography–mass spectrometry (LC–MS; Fig. 3b and Extended Data Fig. 3). With the supplementation of 4.5 mg l^{-1} strictosidine into the fermentation medium (1.5 mg l^{-1} per day for 3 days), the CAN4A strain was only able to synthesize 125 μg l^{-1} catharanthine (Fig. 3c), indicating the necessity of pathway optimization to increase the strictosidine to catharanthine conversion yield.

*Cr*PAS was determined to be localized in the vacuole of *C. roseus*[11]. To enable functional expression in yeast, we replaced the N-terminal signal peptide of *Cr*PAS with a yeast vacuolar localization tag (PEP4sp and ProAsp[39,40] for CAN4B and CAN4C, respectively) or a maltose binding protein (MBP) fusion expression tag (MBP1-tPAS, CAN4D). MBP has been shown to be effective in enhancing the soluble and functional expression of several plant-derived enzymes in the cytosol of yeast[41].

https://doi.org/10.1038/s44160-022-00205-2

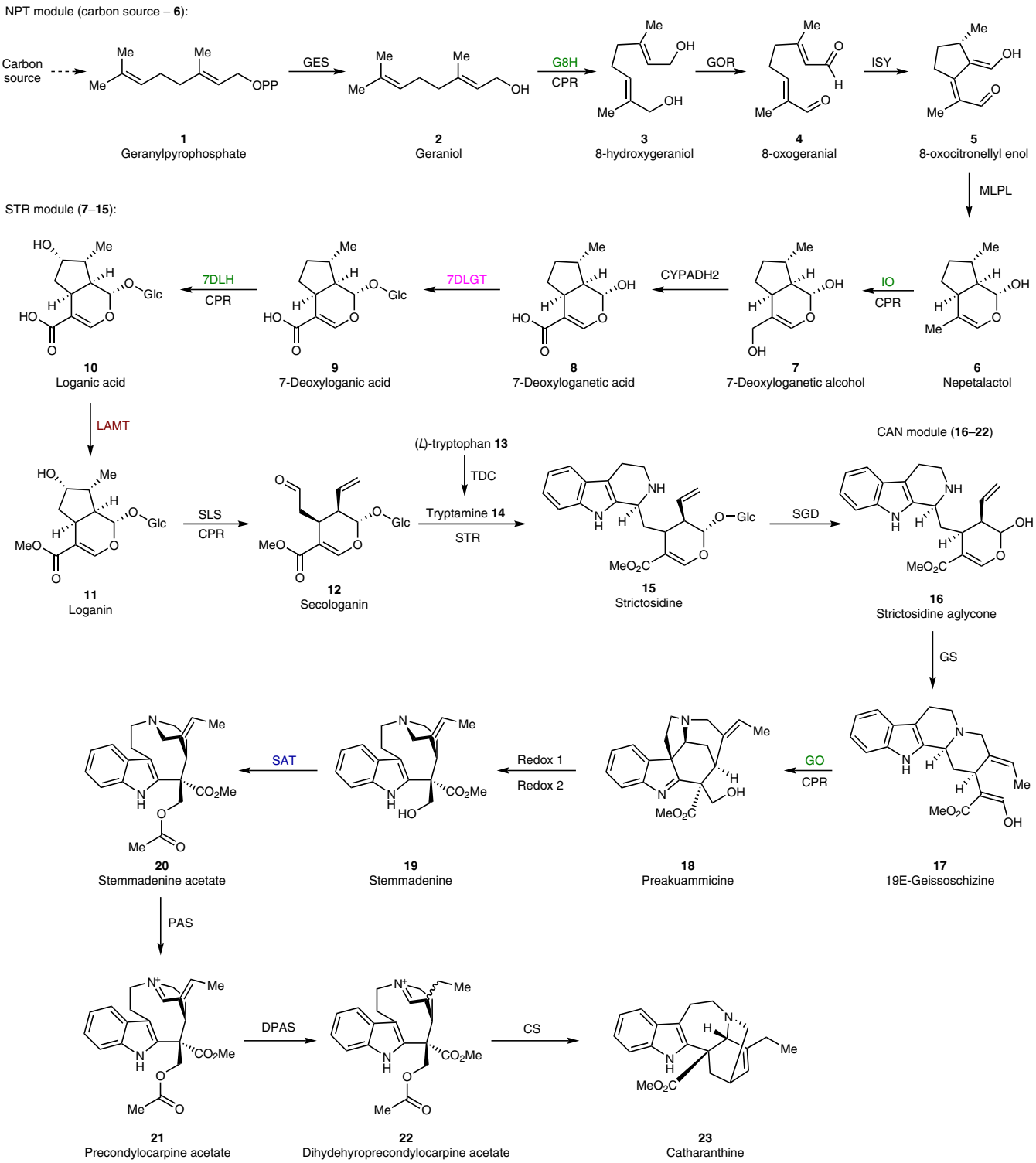


Fig. 1 | Pathway diagram of catharanthine biosynthesis. From the mevalonate pathway intermediate, a total of 25 heterologous genes including *CPR* should be integrated to achieve de novo biosynthesis of catharanthine. Thus, the whole biosynthetic pathway is divided into three functional modules: the NPT module, from carbon sources to nepetalactol **6**; the STR module, from nepetalactol **6** to strictosidine **15**; and the CAN module, from strictosidine **15** to catharanthine **23**. CYPs were shown in green, glucosyltransferases in magenta, methyltransferases in dark red and acetyltransferases in blue. It is thought that 8-oxocitronellyl enol **5** is unstable and a mixture of E/Z isomers[50]. GES, geraniol synthase; G8H, geraniol 8-hydroxylase; GOR, 8-hydroxygeraniol oxidoreductase; MLPLA, major latex protein-like gene A; IO, iridoid oxidase; CYPADH, alcohol dehydrogenase; 7DLGT, 7-deoxyloganetic acid glucosyl transferase; 7DLH, 7-deoxyloganic acid hydroxylase; LAMT, loganic acid O-methyltransferase; SLS, secologanin synthase; STR, strictosidine synthase; TDC, tryptophan decarboxylase; SGD, strictosidine β-D-glucosidase; GS, geissoschizine synthase; GO, geissoschizine oxidase; Redox1/2, cinnamyl alcohol dehydrogenase-like enzyme 1/2; SAT, stemmadenine-O-acetyltransferase; DPAS, dehydroprecondylocarpine acetate synthase; CS, catharanthine synthase; CPR, cytochrome P450 reductase.

https://doi.org/10.1038/s44160-022-00205-2

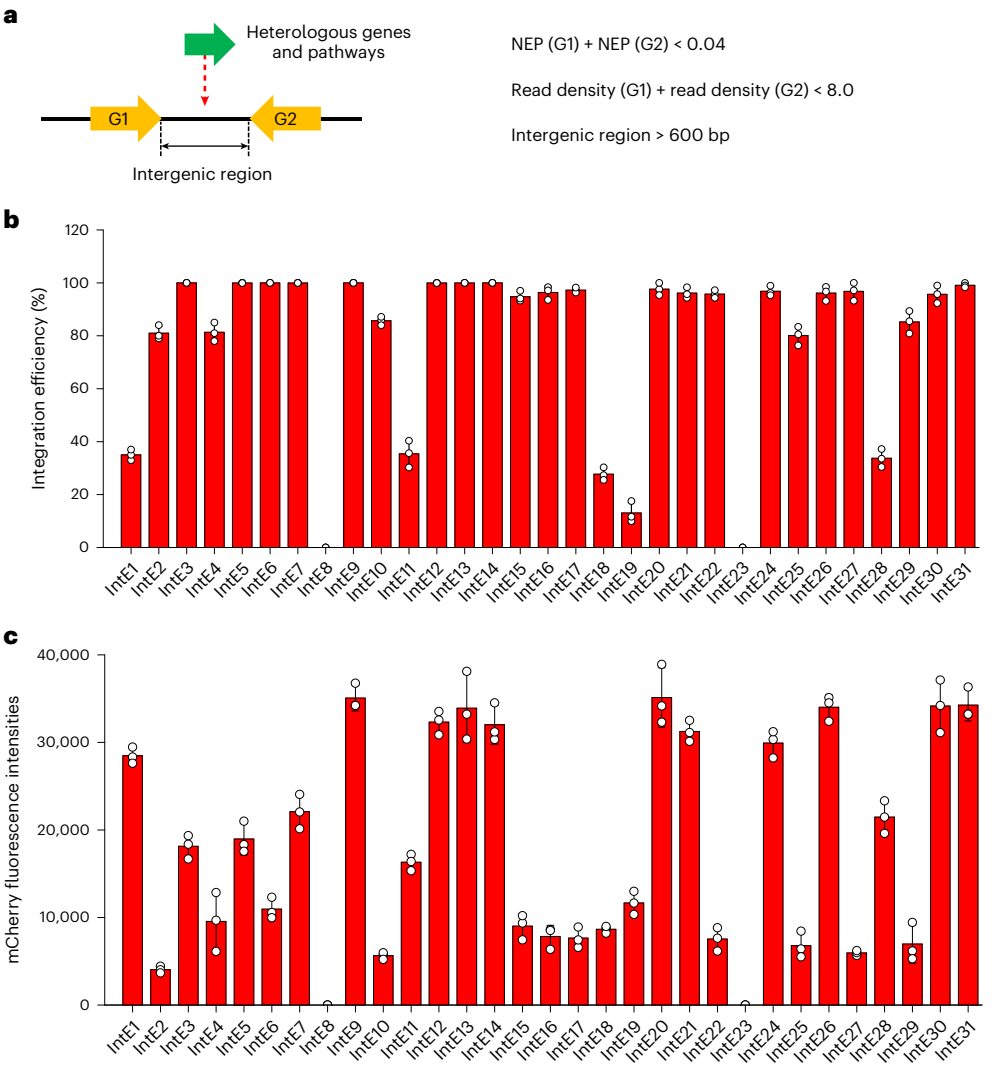


Fig. 2 | Genome-wide profiling and characterization of stable integration sites in *P. pastoris*. **a**, Criteria for the selection of stable integration sites in a genome-wide manner. The intergenic regions should be directly flanked by two essential genes, with the sum of NEP values and read density values determined by the transposon-mediated random insertion method smaller than 0.04 and 8, respectively. In addition, the intergenic regions should be longer than 600 bp, to minimize the impact on the expression of essential genes and accordingly cell fitness. G1/G2: Gene 1/Gene 2. **b**, Genome integration efficiency of the selected intergenic regions. The *mCherry* expression cassette was integrated as a reporter and the integration efficiency was calculated as the percentage of fluorescent cells. **c**, Effect of integration site on the expression level of heterologous genes. The integration efficiency and the mCherry expression level were determined by flow cytometry (Supplementary Fig. 1). The results represent the mean ± s.d. of biological triplicates ($n = 3$).

As shown in Fig. 3c, although the localization of *Cr*PAS to the yeast vacuole had no or a slightly positive effect, fusion expression with MBP resulted in the highest production of catharanthine. Confocal microscopy revealed that PEP4sp-tPAS and ProAsp-tPAS were mainly localized in the vacuole, whereas MBP1-tPAS was cytosolically expressed in *P. pastoris*. More importantly, MBP fusion expression significantly increased the expression of PAS (Extended Data Fig. 4). In other words, MBP fusion expression benefited catharanthine production through the colocalization of pathway enzymes in the cytosol and enhanced soluble expression of PAS. Because linker sequences also play an important role in the activities of fusion proteins, the GGGGS linker in CAN4D (MBP1) was replaced with a more flexible $(GGGGS)_3$ linker (MBP3-tPAS, CAN4E). The resultant CAN4E strain was able to produce approximately 270 μg l^{-1} catharanthine, which was more than twofold higher than that of CAN4A.

To further increase the strictosidine to catharanthine conversion yield, the rate-limiting enzymes of the CAN module should be identified and overexpressed with additional copies. Thus, based on the CAN4E strain, we integrated an additional copy of each gene expression cassette, leading to the construction of strains CAN5A–CAN5H (Supplementary Table 3). Among these serial strains, CAN5A, CAN5B, CAN5C, CAN5E and CAN5G increased the production of catharanthine the most. These results indicated that *CrSGD*, *CrGS*, *CrPAS*, *CrDPAS* and *CrCS* were rate-limiting enzymes of the CAN module (Fig. 3c) and should be integrated with extra copies. In terms of SGD for the formation of aglycone moiety, *Cr*SGD and *Rv*SGD (*Rauvolfia serpentina*) have been studied in detail and both were investigated in this study (CAN5A and CAN5B). In addition, the introduction of Phe373Thr mutation (F373T, *Rv*SGD*) was reported to enhance the catalytic activity of RvSGD[42], which was included into the CAN5C strain. *Cr*SGD, *Rv*SGD and *Rv*SGD* resulted in comparable improvement in the production of catharanthine, with *Rv*SGD* performing slightly better. When compared with *Rv*SGD, *Rv*SGD* demonstrated both higher K_{cat} (rate constant, 952.4 versus 9.8 s^{-1}) and K_m (Michaelis constant, 2,640 versus 90 μM)[42]. In other words, *Rv*SGD* was expected to perform better with a high concentration of strictosidine, whereas *Rv*SGD was more suitable when the

https://doi.org/10.1038/s44160-022-00205-2

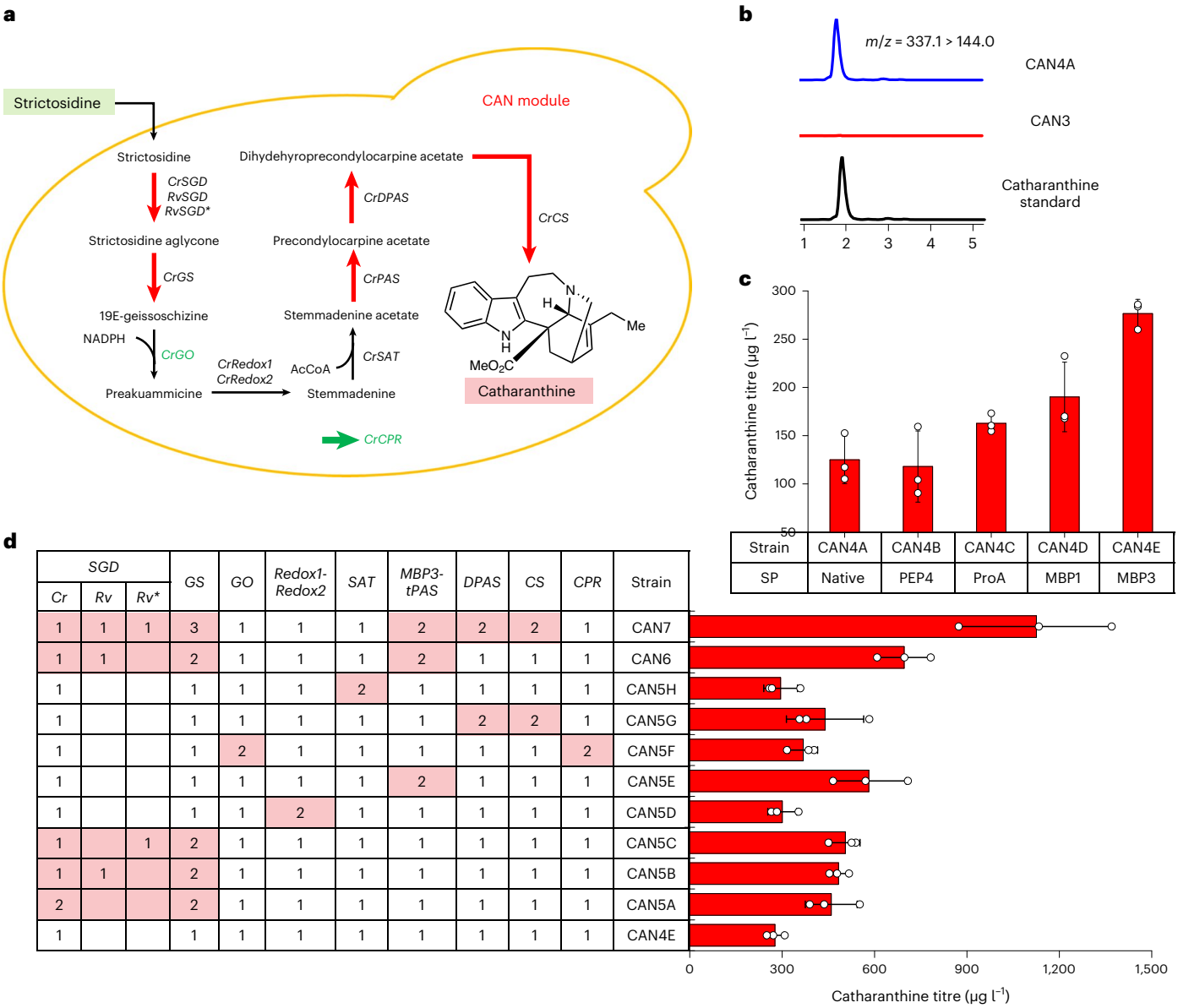


Fig. 3 | Reconstitution and optimization of CAN module. a, The biosynthetic pathway from strictosidine to catharanthine. *Cr, C. roseus*; *Rv, Rauvolfia serpentina*; NADPH, reduced form of nicotinamide adenine dinucleotide phosphate; AcCoA, acetyl-coenzyme A. **b**, LC–MS analysis of the production of catharanthine in *P. pastoris*. MRM spectra (m/z = 337.1 > 144.0) of catharanthine standard, CAN3 (control strain) sample and CAN4A (catharanthine-producing strain) sample. **c**, The effect of signal peptides (ProAsp and PEP4sp for yeast vacuolar localization) or fusion tags (MBP with a GGGGS linker and MBP with a $(GGGGS)_3$ linker for soluble expression in cytosol) on the expression and localization of *Cr*PAS and accordingly the production of catharanthine. PEP4, vacuolar peptidase; ProA, Proteinase A; MBP1, MBP with a GGGGS linker; MBP3, MBP with a $(GGGGS)_3$ linker. **d**, Metabolic pathway engineering for enhanced production of catharanthine from strictosidine. The rate-limiting enzyme encoding genes were amplified by integrating additional copies into strain CAN4E to enhance the production of catharanthine. Strictosidine was supplemented into the growth medium with a final concentration of 4.5 mg l^{-1} (1.5 mg l^{-1} per day for 3 days). The results represent the mean ± s.d. of biological triplicates (n = 3).

concentration of strictosidine was low. Therefore, we integrated both *RvSGD* and *RvSGD** and additional copies of *CrGS*, *MBP3-CrPAS*, *CrDPAS* and *CrCS* to construct the catharanthine-producing yeast strain CAN7. By supplementing a total of 4.5 mg l^{-1} strictosidine, the catharanthine titre reached as high as 1.1 mg l^{-1} (Fig. 3d), which was nearly tenfold higher than that of the original strain (CAN4A, 125 μg l^{-1}).

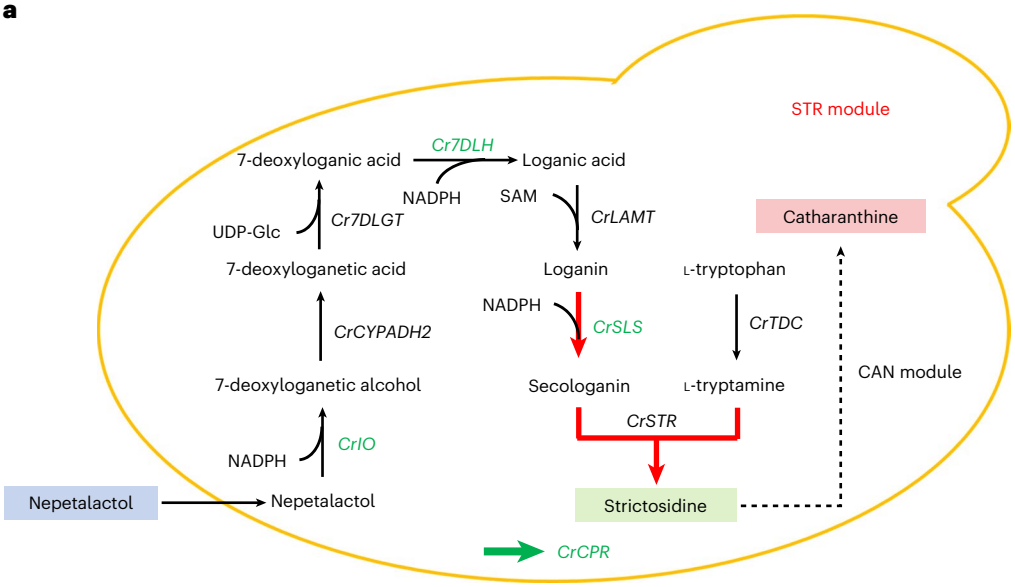
Biosynthesis of catharanthine from nepetalactol

Because the biosynthesis of nepetalactol from simple carbon sources has been reconstituted and optimized in yeast[9,43,44], we designed the STR module for converting nepetalactol to strictosidine (Fig. 4a). Based on the CAN7 strain, we integrated the STR module genes through two rounds of genome editing. The resultant CAN9 strain harboured the biosynthetic pathway from nepetalactol to catharanthine. When 150 mg l^{-1} nepetalactol (50 mg l^{-1} per day for 3 days) was supplemented into the growth medium, CAN9 was able to synthesize catharanthine with a titre of 1.2 mg l^{-1}. Although reconstituted successfully, the STR module should be optimized to improve the pathway efficiency.

Similar to the above gene amplification strategy, we integrated an additional copy of the STR module genes into strain CAN9, leading to the construction of yeast strains CAN10A–CAN10E (Supplementary Table 3). As shown in Fig. 4b, an extra copy of *CrLAMT-CrSTR* (CAN10A) and *Cr7DLH-CrSLS* (CAN10B) increased the production of catharanthine by approximately 1.5-fold, whereas additional copies of *CrIO-Cr7DLGT*

https://doi.org/10.1038/s44160-022-00205-2

a



b

IO	ADH2	7DLGT	7DLH	LAMT	SLS	STR	TDC	Strain
1	1	1	1	1	2	2	1	CAN11
1	1	1	2	2	1	1	1	CAN10E
1	2	1	1	1	1	1	2	CAN10D
2	1	2	1	1	1	1	1	CAN10C
1	1	1	2	1	2	1	1	CAN10B
1	1	1	1	2	1	2	1	CAN10A
1	1	1	1	1	1	1	1	CAN9

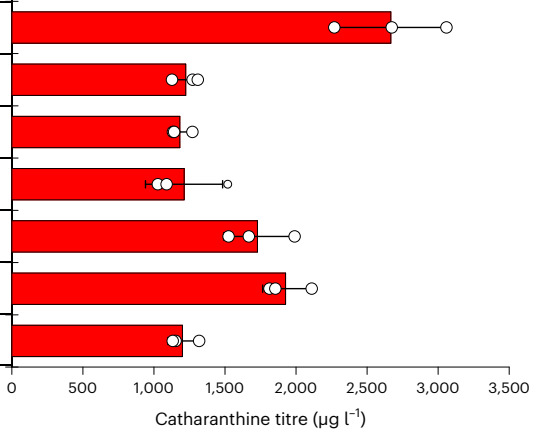


Fig. 4 | Reconstitution and optimization of the STR module. a, The biosynthetic pathway from nepetalactol to strictosidine. UDP-Glc, uridine diphosphate glucose. **b**, Metabolic pathway engineering for enhanced production of catharanthine from nepetalactol. Additional copies of the STR module genes were integrated to evaluate their contribution to catharanthine biosynthesis. Nepetalactol was supplemented into the growth medium at a final concentration of 150 mg l^{-1}. The results represent the mean ± s.d. of biological triplicates ($n = 3$).

(CAN10C), *CrCYPADH-CrTDC* (CAN10D) and *Cr7DLH-CrLAMT* (CAN10E) only increased catharanthine production marginally. These results indicated that *CrSTR* and *CrSLS* were rate-limiting enzymes of the STR module and were combined to construct the CAN11 strain, which was able to produce 2.7 mg l^{-1} catharanthine from 150 mg l^{-1} nepetalactol.

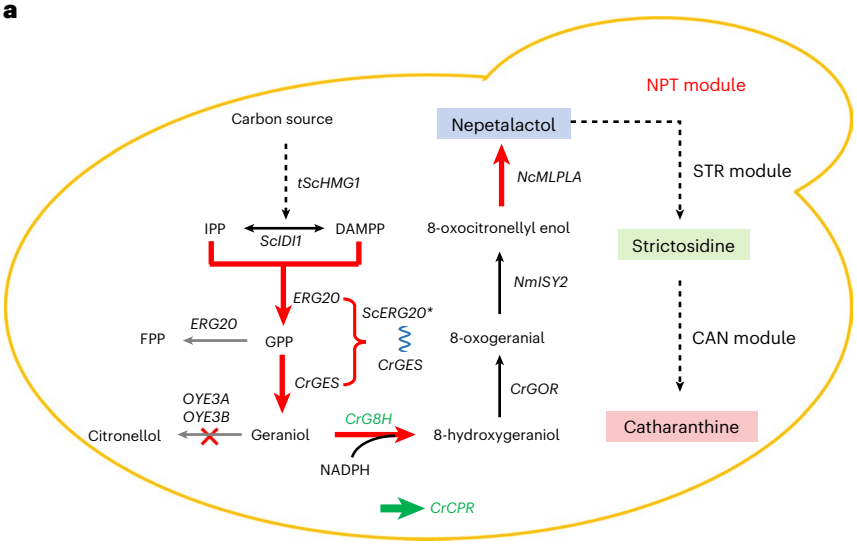
De novo biosynthesis of catharanthine

Aiming to achieve de novo biosynthesis of catharanthine, we further introduced the NPT module into strain CAN11 (Fig. 5a). Although nepetalactol biosynthesis has been attempted and engineered in yeast[9,43,44], the results were still unsatisfactory, probably due to the promiscuity of pathway enzymes and the branching of intermediate metabolites towards undesirable byproducts. Taking the iridoid synthase (ISY) for example, *Cr*ISY has been studied in detail[45] and was found to be rather promiscuous, leading to the synthesis of byproducts (for example, citronellol) other than nepetalactol[46]. In addition, ERG20 demonstrates much higher preference towards the conversion of geranyl pyrophosphate (GPP) to farnesyl pyrophosphate (FPP) and Old Yellow Enzymes (OYEs, NADPH-dependent oxidoreductases) were found to reduce geraniol to citronellol in *S. cerevisiae*[46], making the availability of geraniol and metabolic fluxes towards nepetalactol rather limited.

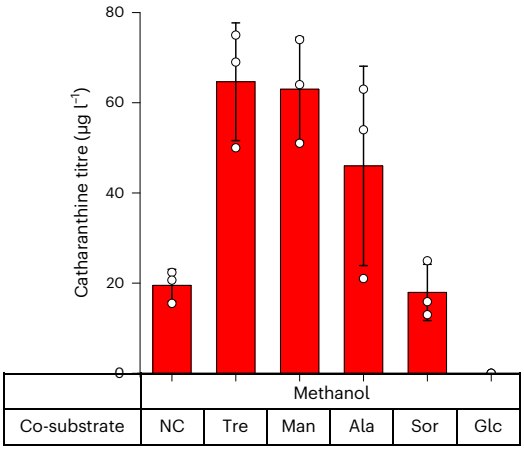
Compared with *Cr*ISY, *Nm*ISY2 (*Nepeta mussinii*) demonstrated comparable activity towards 8-oxogeranial[47], but much lower activity towards the undesirable substrates (for example geraniol)[48], which was chosen instead of *Cr*ISY for constructing the catharanthine biosynthetic pathway. The introduction of F96W and N127W mutations into ERG20 (ERG20^{WW}) or the overexpression of GPP-specific synthase (GPPS from *Abies grandis*, *AgGPPS*) redirected the metabolic fluxes towards GPP formation and the fusion expression of the ERG20 mutant with GES (ERG20^{WW}-GES) resulted in high-level production of geraniol[49]. Based on BLAST search and sequence alignment, OYE3A (PAS_chr3_0377) and OYE3B (PAS_chr3_1184) were identified as potential candidates for geraniol reduction in the *P. pastoris* genome. Therefore, in addition to the NPT module genes (*AgGPPS, CrGES, CrG8H, CrGOR* and *NmISY2*) and the upstream rate-limiting genes (*tScHMG1* and *ScIDI1*), we integrated

https://doi.org/10.1038/s44160-022-00205-2

a



c



b

tHMG1	*IDI1*	*GPPS*		*GES*		*G8H*	*GOR*	*ISY*	*MLPL A*	*Δoye 3A*	*Δoye 3B*	*MXR1*	Strain
		Ag	*Sc**	*Sc*-GES*	*GES*								
1	1	1	1	2	1	2	1	1	1	1	1	1	CAN17
1	1	1		2		2	1	1	1	1	1	1	CAN16
1	1	1		1		1	1	1		1	1	1	CAN15C
1	1	1		2		2	1	1		1	1		CAN15B
1	1	1		1		1	1	1	1	1	1		CAN15A
1	1	1		1		1	1	1		1	1		CAN14
1	1	1		1		1	1	1		1			CAN13

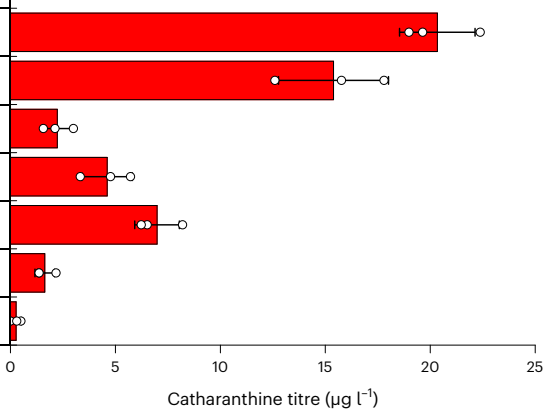


Fig. 5 | Reconstitution and optimization of NPT module for de novo biosynthesis of catharanthine. a, The biosynthetic pathway from carbon source (for example, methanol) to nepetalactol. *Sc*, *S. cerevisiae*; *Nm*, *Nepeta mussinii*; *Nc*, *Nepeta cataria*. **b**, Metabolic pathway engineering for enhanced production of catharanthine from methanol. The copy numbers of the rate-limiting enzyme encoding genes were increased; the supply of biosynthetic precursor geraniol was enhanced by fusion expression of *ScERG20*^{WW} and *CrGES* and removal of competing pathways (*OYE3A* and *OYE3B*); and the transcriptional machinery of the methanol-inducible system was enhanced by *MXR1* overexpression. *GPPS*, geranylpyrophosphate synthase; *Ag*, *Abies grandis*; *Sc**, *S. cerevisiae ERG20*^{WW}. **c**, Effect of non-repressing carbon sources on the production of catharanthine from methanol, with sole methanol fermentation and glucose-fed fermentation included as negative controls. D-trehalose (Tre), mannitol (Man), alanine (Ala), sorbitol (Sor) or glucose (Glc) was supplemented into the fermentation broth as a co-substrate with a final concentration of 1%. The results represent the mean ± s.d. of biological triplicates ($n = 3$).

the *ERG20*^{WW}*-GES* expression cassette and knocked out *OYE3A* and *OYE3B* (Fig. 5a). The resultant yeast strain CAN14 was able to synthesize catharanthine with a titre of 1.6 μg l^{-1} from methanol, achieving de novo biosynthesis of catharanthine in a heterologous host (Fig. 5b).

Although we have established a catharanthine-producing yeast strain, the titre should be further improved for practical applications. Recently, the major latex protein-like (*MLPL*) gene cluster was identified and the overexpression of *MLPLA* significantly increased the production of nepetalactol when incubated with ISY[50]. Thus, CAN15A was constructed by overexpressing *NcMLPLA* (*Nepeta cataria*) in CAN14, and the production of catharanthine was increased by 4.4-fold (7.0 μg l^{-1}). Because all the heterologous genes were driven by methanol-inducible promoters, geraniol and other pathway precursors were synthesized only during the methanol induction phase. Based on the assumption that the production of geraniol and 8-hydroxygeraniol in both glucose phase and methanol phase could enhance the supply of precursors for catharanthine biosynthesis, we integrated an additional copy of *CrG8H* and *CrGES-ScERG20*^{WW} under the control of constitutive promoters (*GAPp* and *TEF1p*) into CAN14. The resultant strain CAN15B was able to produce catharanthine with a titre approaching 4.6 μg l^{-1} (2.9-fold higher than CAN14). In addition, because more than 30 heterologous gene expression cassettes under the control of methanol-inducible promoters have been integrated into *P. pastoris*, the transcriptional machinery might be not sufficient to drive the expression of all catharanthine biosynthetic pathway genes. The overexpression of the transcriptional activator of the methanol regulation machinery (MXR1) increased the production of catharanthine by 1.4-fold (CAN15C). The combination of these three engineering strategies (CAN17) further increased the titre of catharanthine to as high as 20.3 μg l^{-1} (Fig. 5b).

Because more than 30 heterologous genes under the same sets of promoters were integrated and the catharanthine biosynthetic pathway was found to affect cell fitness to some extent (Extended Data Fig. 5),

we evaluated the stability of the engineered strain CAN17. After serial transfers in non-selective YPAM medium (1% yeast extract, 2% peptone A and 2% methanol), we randomly picked ten independent colonies for catharanthine fermentation and quantification. As shown in Extended Data Fig. 6, we observed comparable production of catharanthine in all clones, indicating that the catharanthine biosynthetic pathway genes were stably maintained.

When methanol is used as the sole carbon and energy source, more than 70% of the metabolic fluxes flow to the dissimilation pathway (energy generation and carbon dioxide formation) rather than the assimilation pathway (biomass formation and biosynthesis)[51]. To take advantage of methanol as a decent inducer for high-level expression of heterologous genes and ensure a sufficient supply of metabolic precursors (for example AcCoA A and GPP) for natural product biosynthesis, we proposed to supplement a non-repressing carbon source as a co-substrate for fermentative production of catharanthine. Although methanol-inducible promoters are completely inhibited by glucose and glycerol, a couple of non-repressing carbon sources, such as alanine, sorbitol, mannitol and D-trehalose, have been reported[52,53]. Thus, these non-repressing carbon sources were supplemented to evaluate their effect on catharanthine biosynthesis, with glucose included as a negative control. As expected, the addition of glucose completely abolished the production of catharanthine, probably due to the repression of heterologous genes under the control of methanol-inducible promoters. Among several non-repressing carbon sources, the supplementation of D-trehalose or mannitol resulted in the highest production of catharanthine, reaching up to 65 μg l^{-1} (Fig. 5c).

Metabolic and bioprocess engineering

After optimizing the three modules for de novo catharanthine biosynthesis, we performed metabolic engineering to further boost catharanthine production. Because several P450 enzymes (such as *Cr*G8H, *Cr*1O, *Cr*7DLH, *Cr*SLS and *Cr*GO) are involved in the biosynthetic pathway, we first focused on the engineering strategies to enhance the expression and/or activity of P450 enzymes. The overexpression of endogenous *RAD52* has been reported to increase the activity of a P450 enzyme and accordingly the production of trans-nootkatone by fivefold in *P. pastoris*[28], and we also observed an approximately 1.8-fold increase in the production of catharanthine in the *PpRAD52* overexpression strain (CAN18A). In addition, cytochrome b5 (CYB5) has been determined to increase the electron transfer efficiency of P450 enzymes[54], and the introduction of *CYB5* from *C. roseus* (*CrCYB5*, CAN18B) increased the production of catharanthine by approximately 1.7 fold. Another metabolic engineering strategy for natural product biosynthesis is to enhance the supply of *S*-adenosyl-L-methionine (SAM), which serves as a universal methyl donor for methyltransferases (for example LAMT of the STR module)[22]. The overexpression of *SAM2* from *S. cerevisiae* (*ScSAM2*, CAN18C) increased the production of catharanthine for more than twofold. A combination of these three metabolic engineering strategies resulted in the construction of strain CAN19, whose production of catharanthine was increased by around sixfold to as high as 383 μg l^{-1} (Fig. 6a).

Although extensive pathway engineering and metabolic engineering have been performed, the production of catharanthine was still far from industrial applications. Thus, we tried to analyse the accumulation of pathway intermediates, with an aim to identify rate-limiting steps of the catharanthine biosynthetic pathway for further optimization. Due to the lack of most standard chemicals for the pathway intermediates, we constructed a series of knockout strains (Supplementary Table 4) for the synthesis and potentially accumulation of these intermediates, with MS fragmentation information obtained from previous studies[11,21] (Supplementary Table 5). For example, we knocked out *CrSAT* in the strain containing the CAN module to detect the accumulation of stemmadenine, which should have disappeared in the *CrRedox1* knockout strain. Using such a genetic engineering strategy, we could only detect four pathway intermediates with decent MS signals, including preakuammicine, stemmadenine, stemmadenine acetate, and precondylocarpine acetate (Supplementary Fig. 2). The MS signals for other pathway intermediates were either too low or too noisy, probably due to the compounds having low stability or a high MS detection limit. The four detectable intermediates were not accumulated at high levels except for stemmadenine in the CAN19 strain, indicating CrRedox1/CrRedox2, CrPAS and CrDPAS were not rate limiting for catharanthine biosynthesis in CAN19, and CrSAT should be manipulated for further engineering. Because including extra copies of *CrSAT* failed to decrease the accumulation of stemmadenine and accordingly increase the production of catharanthine (Fig. 3), CrSAT activity was not limited by the expression level and other engineering efforts should be performed for future studies.

Finally, we performed catharanthine fermentation in a 1 l bioreactor with 0.7 l working volume. As for the fermentation with methanol as sole carbon source, 2% methanol was fed into the bioreactor every 4 h after induction, with the cell densities (OD_{600}) and catharanthine production reached as high as approximately 170 and 0.62 mg l^{-1} (Fig. 6b, top panel). For fed-batch fermentation with a mixture of carbon sources, 1% methanol and 1% mannitol was fed into the bioreactor every 4 h after induction. After 70 h of fermentation, cell densities (OD_{600}) were steadily increased to a maximum of approximately 230, and the production of catharanthine reached a titre as high as 2.57 mg l^{-1} (Fig. 6b, bottom panel).

Discussion

MIAs are a family of plant natural products with important medicinal values and complicated biosynthetic pathways[1]. With the development of synthetic biology and metabolic engineering, there is a growing interest in producing these medicinally relevant compounds in microbial cell factories, particularly in yeasts[15]. Nevertheless, grand challenges remain to be addressed for efficient biosynthesis of MIAs: (1) only a few MIA biosynthetic pathways have been elucidated and can be reconstituted in microbes; (2) the assembly and stable maintenance of multigene biosynthetic pathways (general more than 20 heterologous genes) is not an easy task; (3) the biosynthesis efficiency can be limited by the poor expression, low activity and/or low specificity of plant-derived enzymes in a heterologous host. For example, Brown et al. reconstituted the whole biosynthetic pathway and achieved de novo production of strictosidine, a general precursor for the biosynthesis of more than 3,000 MIA molecules[21]. Unfortunately, due to the low activity of G8H and other P450 enzymes[21] and the poor specificity of ISY[46], strictosidine was produced at an extremely low titre, making the further reconstitution of MIA biosynthetic pathways nearly impossible.

In this study, considering the advantages in expressing rate-limiting enzymes (such as P450 enzymes and PAS) to high levels, we aimed to establish *P. pastoris* as a cell factory for the de novo biosynthesis of catharanthine, a monomer for producing the anticancer drug vinblastine, both biologically and chemically. Through combinatorial metabolic engineering, such as increasing the copy number of rate-limiting enzyme encoding genes (such as *CrG8H, CrSLS, CrSTR* and *CrCS*), choosing the appropriate plant enzymes with high activity (such as *RvSGD* and its mutant) and specificity (such as *NmISY2*), locating key pathway enzymes to the correct compartment (such as *CrPAS*), fusing CrGES with an ERG20 mutant for enhanced metabolite trafficking and precursor supply, and blocking competing pathways (such as *OYE3A* and *OYE3B*), we integrated more than 30 heterologous gene expression cassettes and achieved the de novo biosynthesis of catharanthine in *P. pastoris*. Although a previous study has reported the production of catharanthine in cell cultures of *C. roseus*, *P. pastoris* showed advantages of high production at a low cost for large-scale applications[55]. While this manuscript was under review, Zhang et al. reported the de novo biosynthesis of catharanthine and vindoline in *S. cerevisiae*[56]. Using fed-batch fermentation, the production of catharanthine in *P.*

https://doi.org/10.1038/s44160-022-00205-2

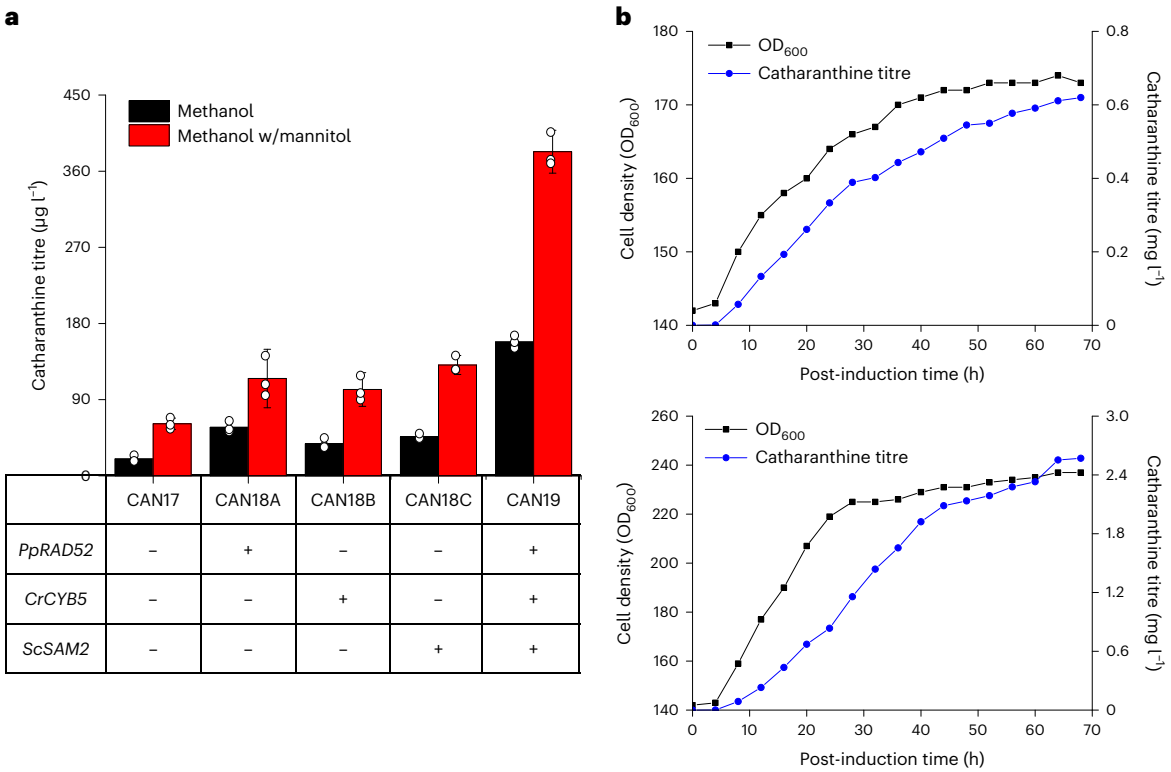


Fig. 6 | Optimization and fed-fed fermentation of catharanthine-producing strains. a, Metabolic engineering of catharanthine-producing strains. *PpRAD52*, endogenous *RAD52* encoding a recombinase involved in the repair of double-strand breaks; *CrCYB5*, cytochrome b5 encoding gene from *C. roseus*; *ScSAM2*, SAM synthetase encoding gene from *S. cerevisiae*. The results represent the mean ± s.d. of biological triplicates ($n = 3$). **b**, Bioreactor fermentation of CAN19 for enhanced biosynthesis of catharanthine. The post-induction fermentation profiles with methanol (top panel) or methanol/mannitol (bottom panel) as the substrate were provided. At each data point, methanol (2%) or methanol/mannitol (1% each) was fed into the bioreactor as carbon sources for fermentative production of catharanthine. Fermentation profiles of other parameters were provided in Extended Data Fig. 7.

pastoris was more than 20-fold higher, highlighting the advantages of *P. pastoris* in expressing plant proteins and/or biosynthesis of plant natural products. In contrast, the production of catharanthine indicated that PAS could be functionally expressed in *S. cerevisiae*, which was contradictory to what Caputi et al.[11] reported, probably due to the use of different yeast-expression systems. Nevertheless, the production is still rather low and further metabolic engineering should be used to achieve industrial production, particularly for P450 enzymes, several of which (such as *Cr*G8H, *Cr*1O, *Cr*7DLH, *Cr*SLS and *Cr*GO) are involved in the biosynthetic pathway and generally considered rate limiting for plant natural product biosynthesis[14]. Although our metabolic engineering strategies increased the production of catharanthine, our limited understanding of membrane protein expression-folding network remains the biggest challenge for further P450 engineering. For example, this and previous studies reported that the overexpression of *RAD52*, encoding a protein responsible for DNA repair and recombination, significantly improved the expression and/or activity of P450 enzymes and the production of the corresponding plant natural products[57], whereas the detailed mechanism has not yet been elucidated. In other words, we should develop genome-scale metabolic engineering strategies that can theoretically perturb all the genes to identify non-intuitive engineering targets to enhance P450 expression and natural product biosynthesis. Although CRISPR-based genome-scale engineering has been well established in model organisms[58,59], its application in *P. pastoris* is yet to be explored.

A series of integration sites with different integration efficiencies and expression levels were characterized, laying the foundation for the reconstitution of multigene pathways in *P. pastoris*. To construct engineered strains with complex biosynthetic pathways, many heterologous genes should be integrated into the genome in an iterative manner. In this case, off-target effect of the CRISPR-based genome integration becomes a concern for the construction of stable cell factories. Accordingly, we analysed strain CAN19 for potential off-target effect. First of all, the seed sequences (12 bp) of each guide RNA (gRNA) used for the construction of CAN19 were blasted against the *P. pastoris* GS115 genome and a total of 47 potential off-targeting sites sharing the highest homology were chosen for genotyping (Supplementary Table 6). After PCR amplification and sequencing of the potential off-target regions, we found no mutations and accordingly no off-target effect in strain CAN19 (Supplementary Fig. 3). In addition, we performed whole-genome sequencing and analysed point mutations, deletions, and insertions of CAN1 (GS115-Cas9, parent strain), CAN7 (with optimized CAN module), CAN11 (with optimized CAN and STR modules) and CAN19 (the finally engineered strain with optimized CAN, STR and NPT modules). During the generation of the engineered strains, although no unexpected insertions or deletions were detected, some point mutations (a total of 47 mutations, Supplementary Table 7) were indeed observed. Because these mutations were not correlated with gRNAs, they were estimated to have occurred and accumulated randomly during cell division, and not caused by the off-target effect of the CRISPR-based genome editing system.

In summary, we profiled intergenic regions flanked by essential genes as stable integration sites in a genome-wide manner, which was used for multiplex integration and stable maintenance of multigene biosynthetic pathways in *P. pastoris*. In combination with modular metabolic engineering strategies, we achieved de novo biosynthesis of catharanthine in a heterologous host. With more than 30 heterologous gene expression cassettes integrated, it represents the most

complicated biosynthetic pathway ever reconstituted in *P. pastoris*. The characterized stable integration sites and multi-tier modular metabolic pathway engineering strategy can be used to establish *P. pastoris* cell factories for the production of other MIAs (such as vinblastine) and plant natural products.

Methods

Plasmid construction

All plasmids used in this study were constructed by digestion, ligation or Gibson assembly. The flow chart for plasmid construction was described in Supplementary Fig. 4. The gRNA sequences targeting the integration sites were designed using the Benchling CRISPR online tool (https://benchling.com/crispr) and cloned into *Bsa*I digested gRNA helper plasmids (HZP-sgRNA, HHP-sgRNA and HGP-sgRNA). The integration donor helper plasmids were constructed by cloning the upstream (approximately 500 bp) homology arms, dual promoter and terminator pairs (*AOX1p-KpnI-BsaI-CYC1t-FLD1p-BamHI-XhoI-DUS4t* for inducible expression and *GAPp-XhoI-BglII-AOX1t-TEF1p-HindIII-NdeI-PRX5t* for constitutive expression) and downstream (approximately 500 bp) homology arms into pUC19 using Gibson assembly. *CrGOR*, *CrLAMT*, *CrTDC*, *Cr7DLGT*, *AgGPPS* and *CrCYPADH* were amplified from Strain4, *CrGES-ERG20^{WW}* was amplified from pJGZ200 (ref. 49), *tScHMG1* and *ScIDI1* were amplified from the genomic DNA of *S. cerevisiae* BY4741, *CrG8H*, *NmISY2*, *NcMLPLA*, *CrIO*, *Cr7DLH*, *CrSLS* (isoform 2), *CrSTR*, *CrSGD*, *RvSGD*, *CrGS*, *CrGO*, *CrRedox1*, *CrRedox2*, *CrSAT*, *CrPAS*, *CrDPAS* and *CrCS* were codon optimized for yeast expression and chemically synthesized by Genscript Biotech (Nanjing, China). The catharanthine biosynthetic pathway genes were cloned into the restriction sites of the integration donor helper plasmids. All the catharanthine biosynthetic pathway genes are listed in Supplementary Table 2 and the corresponding DNA sequences are listed in Supplementary Data 1. All the plasmids constructed in this study are listed in Supplementary Table 8 and the corresponding oligonucleotides are listed in Supplementary Table 9.

Strain construction

All genome editing and pathway integration were performed through the CRISPR/Cas9 system in *P. pastoris*. The strain construction procedures are in Supplementary Fig. 5. The dual gene expression cassettes were amplified together with the homology arm sequences by PCR (approximately 1 µg for each) and co-transformed with the corresponding gRNA plasmids (approximately 500 ng for each) for integrating into the specific loci of the genome of *P. pastoris*. The integrated cassettes were verified by diagnostic PCR and DNA sequencing. All strains constructed in this study are listed in Supplementary Table 3.

Fluorescence intensity measurement

P. pastoris colonies were randomly selected from the screening plates, inoculated into in 1 ml YPAD medium (1% yeast extract, 2% peptone A and 2% glucose), and cultured until saturation (approximately 48 h). The seed culture was then transferred to 1 ml YPAM medium with 1% inoculum. The yeast strains were cultured for 16 h to measure mCherry fluorescence intensities using the Attune NxT Flow Cytometer (Thermo Fisher Scientific) equipped with an excitation laser of 561 nm and an emission filter of 620 nm. All yeast strains were cultivated in 96-well plates in a high-speed shaker (30 °C and 800 rpm).

Confocal microscopy

Yeast cells expressing a vacuole-targeting eGFP (PEP4sp-eGFP) and different PAS-mCherry variants (including PAS-mCherry, PEP4sp-tPAS-mCherry, ProAsp-tPAS-mCherry, MBP1-tPAS-mCherry and MBP3-tPAS-mCherry) were cultured in YPM for 16 h, washed twice and then resuspended in 10 mM PBS. Around 5 µl of the resuspension was taken for confocal fluorescence detection using the OLYMPUS IX83-FV3000 Confocal Microscope (Olympus, Japan). The fluorescence parameters were set as 500/540 nm and 570/670 nm for eGFP and mCherry, respectively.

Evaluation of yeast strain stability

To test the stability of the selected integration site, the mCherry containing strains (pink) were cultured in non-selective medium (YPAD) for approximately 30 generations (via serial transfers), and then approximately 1,000 colonies were spread on YPAD agar plates. The percentage of pink colonies represented the stability of the integrated mCherry expression cassette in *P. pastoris*. Similarly, to evaluate the stability of the catharanthine-producing strain, CAN17 was cultured in YPAM for approximately 72 generations and spread into YPAD agar plates. Then ten colonies were randomly selected and the stability of the engineered strain was evaluated by comparing the titre of catharanthine.

Whole-genome sequencing and analysis

Genomic DNA of strains CAN1, CAN7, CAN11, CAN17 and CAN19 were extracted and subject to shotgun sequencing with an Illumina NovaSeq PE150 (Illumina). Paired reads were mapped to reference genome or to in silico-generated sequences using the Bowtie2 v.2.3.2 (ref. 60); --very-sensitive-local and processed with Pilon v.1.24 (ref. 61). The resulting sequences were aligned using MAFFT v.7 (ref. 62). Raw sequencing data of strains (CAN1, CAN7, CAN11, CAN17 and CAN19) were deposited at the National Center for Biotechnology Information under BioProject ID PRJNA859235.

Statistics and reproducibility

All experimental data were at least in triplicate and expressed as means ± standard errors. All data analysis was performed by Excel or OriginPro.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data involved in the research are included in the manuscript, supplementary materials and supplementary data. NGS data have been deposited into NCBI with an accession number of PRJNA859235.

References

1. Facchini, P. J. & De Luca, V. Opium poppy and *Madagascar periwinkle*: model non-model systems to investigate alkaloid biosynthesis in plants. *Plant J.* **54**, 763–784 (2008).
2. Kang, M. et al. A chromosome-level *Camptotheca acuminata* genome assembly provides insights into the evolutionary origin of camptothecin biosynthesis. *Nat. Commun.* **12**, 3531 (2021).
3. Liu, J. et al. Enhancement of vindoline and vinblastine production in suspension-cultured cells of *Catharanthus roseus* by artemisinic acid elicitation. *World J. Microbiol. Biotechnol.* **30**, 175–180 (2014).
4. Mohammed, A. E. et al. Chemical diversity and bioactivities of monoterpene indole alkaloids (MIAs) from Six *Apocynaceae Genera*. *Molecules* https://doi.org/10.3390/molecules26020488 (2021).
5. Achan, J. et al. Quinine, an old anti-malarial drug in a modern world: role in the treatment of malaria. *Malar. J.* **10**, 144 (2011).
6. Feng, L. W. et al. Reaction of donor-acceptor cyclobutanes with indoles: a general protocol for the formal total synthesis of (+/-)-strychnine and the total synthesis of (+/-)-akuammicine. *Angew. Chem. Int. Ed. Engl.* **56**, 3055–3058 (2017).
7. Smith, J. M., Moreno, J., Boal, B. W. & Garg, N. K. Cascade reactions: a driving force in akuammiline alkaloid total synthesis. *Angew. Chem. Int. Ed. Engl.* **54**, 400–412 (2015).
8. Tokuda, R. et al. Asymmetric total synthesis of kopsiyunnanine K, a monoterpenoid indole alkaloid with a rearranged skeleton. *Org. Lett.* **18**, 3490–3493 (2016).

9. Miettinen, K. et al. The seco-iridoid pathway from *Catharanthus roseus*. *Nat. Commun.* **5**, 3606 (2014).
10. Sharma, A., Amin, D., Sankaranarayanan, A., Arora, R. & Mathur, A. K. Present status of *Catharanthus roseus* monoterpenoid indole alkaloids engineering in homo- and hetero-logous systems. *Biotechnol. Lett* **42**, 11–23 (2020).
11. Caputi, L. et al. Missing enzymes in the biosynthesis of the anticancer drug vinblastine in *Madagascar periwinkle*. *Science* **360**, 1235–1239 (2018).
12. Qu, Y. et al. Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proc. Natl Acad. Sci. USA* **112**, 6224–6229 (2015).
13. Qu, Y., Safonova, O. & De Luca, V. Completion of the canonical pathway for assembly of anticancer drugs vincristine/vinblastine in *Catharanthus roseus*. *Plant J.* **97**, 257–266 (2019).
14. Jiang, L., Huang, L., Cai, J., Xu, Z. & Lian, J. Functional expression of eukaryotic cytochrome P450s in yeast. *Biotechnol. Bioeng.* **118**, 1050–1065 (2021).
15. Lian, J., Mishra, S. & Zhao, H. Recent advances in metabolic engineering of *Saccharomyces cerevisiae*: new tools and their applications. *Metab. Eng.* **50**, 85–108 (2018).
16. Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M. & Smolke, C. D. Complete biosynthesis of opioids in yeast. *Science* **349**, 1095–1100 (2015).
17. Guo, L. et al. The opium poppy genome and morphinan production. *Science* **362**, 343–347 (2018).
18. Ping, Y. et al. *De novo* production of the plant-derived tropine and pseudotropine in yeast. *ACS Synth. Bio.* **8**, 1257–1262 (2019).
19. Srinivasan, P. & Smolke, C. A.-O. Biosynthesis of medicinal tropane alkaloids in yeast. *Nature* **585**, 614–619 (2020).
20. Thak, E. J., Yoo, S. J., Moon, H. Y. & Kang, H. A. Yeast synthetic biology for designed cell factories producing secretory recombinant proteins. *FEMS Yeast Res.* https://doi.org/10.1093/femsyr/foaa009 (2020).
21. Brown, S., Clastre, M., Courdavault, V. & O'Connor, S. E. *De novo* production of the plant-derived alkaloid strictosidine in yeast. *Proc. Natl Acad. Sci. USA* **112**, 3205–3210 (2015).
22. Liu, T. et al. Efficient production of vindoline from tabersonine by metabolically engineered *Saccharomyces cerevisiae*. *Commun. Biol.* **4**, 1089 (2021).
23. Duan, X., Gao, J. & Zhou, Y. J. Advances in engineering methylotrophic yeast for biosynthesis of valuable chemicals from methanol. *Chin. Chem. Lett.* **29**, 681–686 (2018).
24. Zuo, Y. et al. Establishing komagataella phaffii as a cell factory for efficient production of sesquiterpenoid alpha-santalene. *J. Agric. Food Chem.* **70**, 8024–8031 (2022).
25. Zhou, Q. et al. High-level production of a thermostable mutant of *Yarrowia lipolytica* lipase 2 in *Pichia pastoris*. *Int. J. Mol. Sci.* https://doi.org/10.3390/ijms21010279 (2019).
26. Araya-Garay, J. M., Feijoo-Siota, L., Rosa-dos-Santos, F., Veiga-Crespo, P. & Villa, T. G. Construction of new *Pichia pastoris* X-33 strains for production of lycopene and beta-carotene. *Appl. Microbiol. Biotechnol.* **93**, 2483–2492 (2012).
27. Liu, Y. et al. Engineered monoculture and co-culture of methylotrophic yeast for de novo production of monacolin J and lovastatin from methanol. *Metab. Eng.* **45**, 189–199 (2018).
28. Wriessnegger, T. et al. Production of the sesquiterpenoid (+)-nootkatone by metabolic engineering of *Pichia pastoris*. *Metab. Eng.* **24**, 18–29 (2014).
29. Weninger, A., Hatzl, A. M., Schmid, C., Vogl, T. & Glieder, A. Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*. *J. Biotechnol.* **235**, 139–149 (2016).
30. Gu, Y. et al. Construction of a series of episomal plasmids and their application in the development of an efficient CRISPR/Cas9 system in *Pichia pastoris*. *World J. Microbiol. Biotechnol.* **35**, 79 (2019).
31. Dalvie, N. C. et al. Host-informed expression of CRISPR guide RNA for genomic engineering in *Komagataella phaffii*. *ACS Synth. Biol.* https://doi.org/10.1021/acssynbio.9b00372 (2019).
32. Liu, Q. et al. CRISPR/Cas9-mediated genomic multiloci integration in *Pichia pastoris*. *Microb. Cell Fact.* **18**, 144 (2019).
33. Cai, P. et al. Recombination machinery engineering facilitates metabolic engineering of the industrial yeast *Pichia pastoris*. *Nucleic Acids Res.* **49**, 7791–7805 (2021).
34. Gao, J. et al. A synthetic biology toolkit for marker-less integration of multi-gene pathways into *Pichia pastoris* via CRISPR/Cas9. *ACS Synth. Biol.* **11**, 623–633 (2022).
35. Gao, J., Jiang, L. & Lian, J. Development of synthetic biology tools to engineer *Pichia pastoris* as a chassis for the production of natural products. *Synth. Syst. Biotechnol.* **6**, 110–119 (2021).
36. Mikkelsen, M. D. et al. Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. *Metab. Eng.* **14**, 104–111 (2012).
37. Zhu, J. et al. Genome-wide determination of gene essentiality by transposon insertion sequencing in yeast *Pichia pastoris*. *Sci. Rep.* **8**, 10223 (2018).
38. Brown, S., Clastre, M., Courdavault, V. & O'Connor, S. E. De novo production of the plant-derived alkaloid strictosidine in yeast. *Proc. Natl Acad. Sci. USA* **112**, 3205–3210 (2015).
39. Luo, X. et al. Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* **567**, 123–126 (2019).
40. Zirpel, B., Stehle, F. & Kayser, O. Production of Delta9-tetrahydrocannabinolic acid from cannabigerolic acid by whole cells of *Pichia (Komagataella) pastoris* expressing Delta9-tetrahydrocannabinolic acid synthase from *Cannabis sativa L*. *Biotechnol. Lett* **37**, 1869–1875 (2015).
41. Reider Apel, A. et al. A Cas9-based toolkit to program gene expression in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **45**, 496–508 (2017).
42. Barleben, L., Panjikar, S., Ruppert, M., Koepke, J. & Stöckigt, J. Molecular architecture of strictosidine glucosidase: the gateway to the biosynthesis of the monoterpenoid indole alkaloid family. *Plant Cell* **19**, 2886–2897 (2007).
43. Duan, Y., Liu, J., Du, Y., Pei, X. & Li, M. *Aspergillus oryzae* biosynthetic platform for de novo iridoid production. *J. Agric. Food Chem.* **69**, 2501–2511 (2021).
44. Billingsley, J. M. et al. Engineering the biocatalytic selectivity of iridoid production in *Saccharomyces cerevisiae*. *Metab. Eng.* **44**, 117–125 (2017).
45. Kries, H. et al. Structural determinants of reductive terpene cyclization in iridoid biosynthesis. *Nat. Chem. Biol.* **12**, 6–8 (2016).
46. Campbell, A. et al. Engineering of a nepetalactol-producing platform strain of *Saccharomyces cerevisiae* for the production of plant seco-iridoids. *ACS Synth. Biol.* **5**, 405–414 (2016).
47. Sherden, N. H. et al. Identification of iridoid synthases from *Nepeta* species: iridoid cyclization does not determine nepetalactone stereochemistry. *Phytochemistry* **145**, 48–56 (2018).
48. Jiang, G., Yao, M., Wang, Y., Xiao, W. & Yuan, Y. A "push-pull-restrain" strategy to improve citronellol production in *Saccharomyces cerevisiae*. *Metab. Eng.* **66**, 51–59 (2021).
49. Jiang, G. Z. et al. Manipulation of GES and ERG20 for geraniol overproduction in *Saccharomyces cerevisiae*. *Metab. Eng.* **41**, 57–66 (2017).
50. Lichman, B. R. et al. The evolutionary origins of the cat attractant nepetalactone in catnip. *Sci. Adv.* **6**, eaba0721 (2020).

51. Vanz, A. L. et al. Physiological response of *Pichia pastoris* GS115 to methanol-induced high level production of the Hepatitis B surface antigen: catabolic adaptation, stress responses, and autophagic processes. *Microb. Cell Fact.* **11**, 103 (2012).
52. Inan, M. & Meagher, M. M. Non-repressing carbon sources for alcohol oxidase (*AOX1*) promoter of *Pichia pastoris*. *J. Biosci. Bioeng.* **92**, 585–589 (2001).
53. Nakagawa, T., Wakayama, K. & Hayakawa, T. Selection of suitably non-repressing carbon sources for expression of alcohol oxidase isozyme promoters in the methylotrophic yeast *Pichia methanolica*. *J. Biosci. Bioeng.* **120**, 41–44 (2015).
54. Paddon, C. J. et al. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496**, 528–532 (2013).
55. Zhou, P. et al. Effects of beta-cyclodextrin and methyl jasmonate on the production of vindoline, catharanthine, and ajmalicine in *Catharanthus roseus* cambial meristematic cell cultures. *Appl. Microbiol. Biotechnol.* **99**, 7035–7045 (2015).
56. Zhang, J. et al. A microbial supply chain for production of the anti-cancer drug vinblastine. *Nature* **609**, 341–347 (2022).
57. Wriessnegger, T. et al. Enhancing cytochrome P450-mediated conversions in *P. pastoris* through *RAD52* over-expression and optimizing the cultivation conditions. *Fungal Genet. Biol.* **89**, 114–125 (2016).
58. Lian, J., Schultz, C., Cao, M., HamediRad, M. & Zhao, H. Multi-functional genome-wide CRISPR system for high throughput genotype-phenotype mapping. *Nat. Commun.* **10**, 5794 (2019).
59. Garst, A. D. et al. Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat. Biotechnol.* **35**, 48–55 (2017).
60. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
61. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).
62. Rozewicki, J., Li, S., Amada, K. M., Standley, D. M. & Katoh, K. MAFFT-DASH: integrated protein sequence and structural alignment. *Nucleic Acids Res.* **47**, W5–W10 (2019).

Acknowledgements

This work was supported by National Key Research and Development Program of China (grant nos. 2018YFA0901800 and 2021YFC2103200 to J.L.), Natural Science Foundation of Zhejiang Province (grant no. LR20B060003 to J.L.), Natural Science Foundation of China (grant no. 22278361 to J.L.) and Fundamental Research Funds for the Central Universities (grant nos. 226-2022-00214 and 226-2022-00055 to J.L.). We thank S. E. O'Connor from Max Planck Institute of Chemical Ecology for kindly sharing Strain4 and we thank Y. Yuan and W. Xiao from Tianjin University for generously providing pJGZ200. We also would like to thank iBioFoundry and Core Facility at the Institute for Intelligent Bio/Chem Manufacturing, ZJU-Hangzhou Global Scientific and Technological Innovation Center for analytical support.

Author contributions

J.L. conceptualized and supervised the study. J.G. performed experiments, analysed data and drafted the manuscript. J.G., Y.Z., J.X., C.Y., L.F. and L.J. selected and characterized the integration sites. J.G., T.L. and D.G. designed and constructed the catharanthine biosynthetic pathway. J.G. and F.X. performed the bioreactor fermentation. J.G. and D.L. performed the LC–MS analysis. J.G. and Y.W. performed the whole-genome sequencing and analysis. J.L., B.M., L.H. and Z.X. revised the manuscript. All authors approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

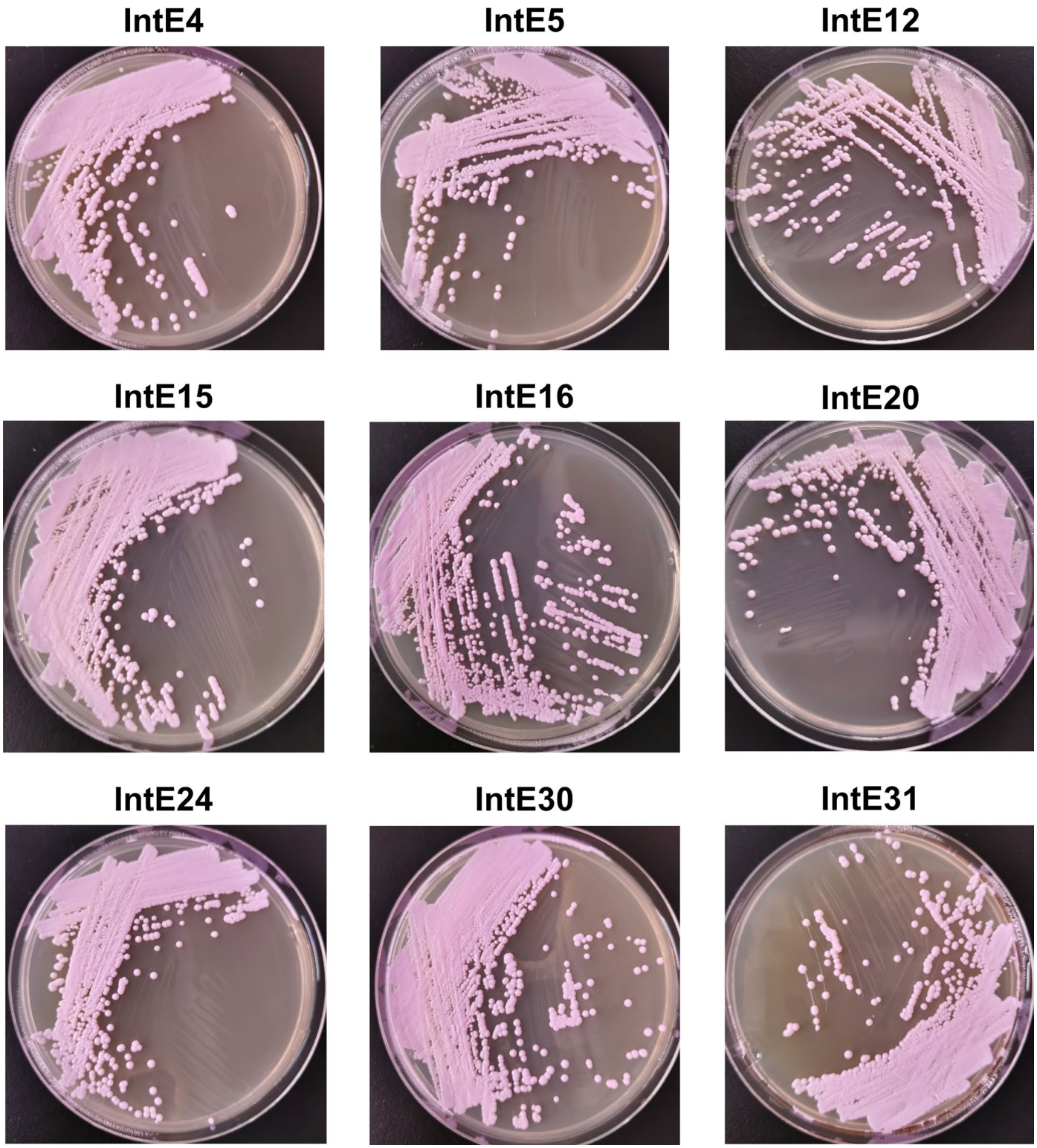
Extended data is available for this paper at https://doi.org/10.1038/s44160-022-00205-2.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44160-022-00205-2.

Correspondence and requests for materials should be addressed to Jiazhang Lian.

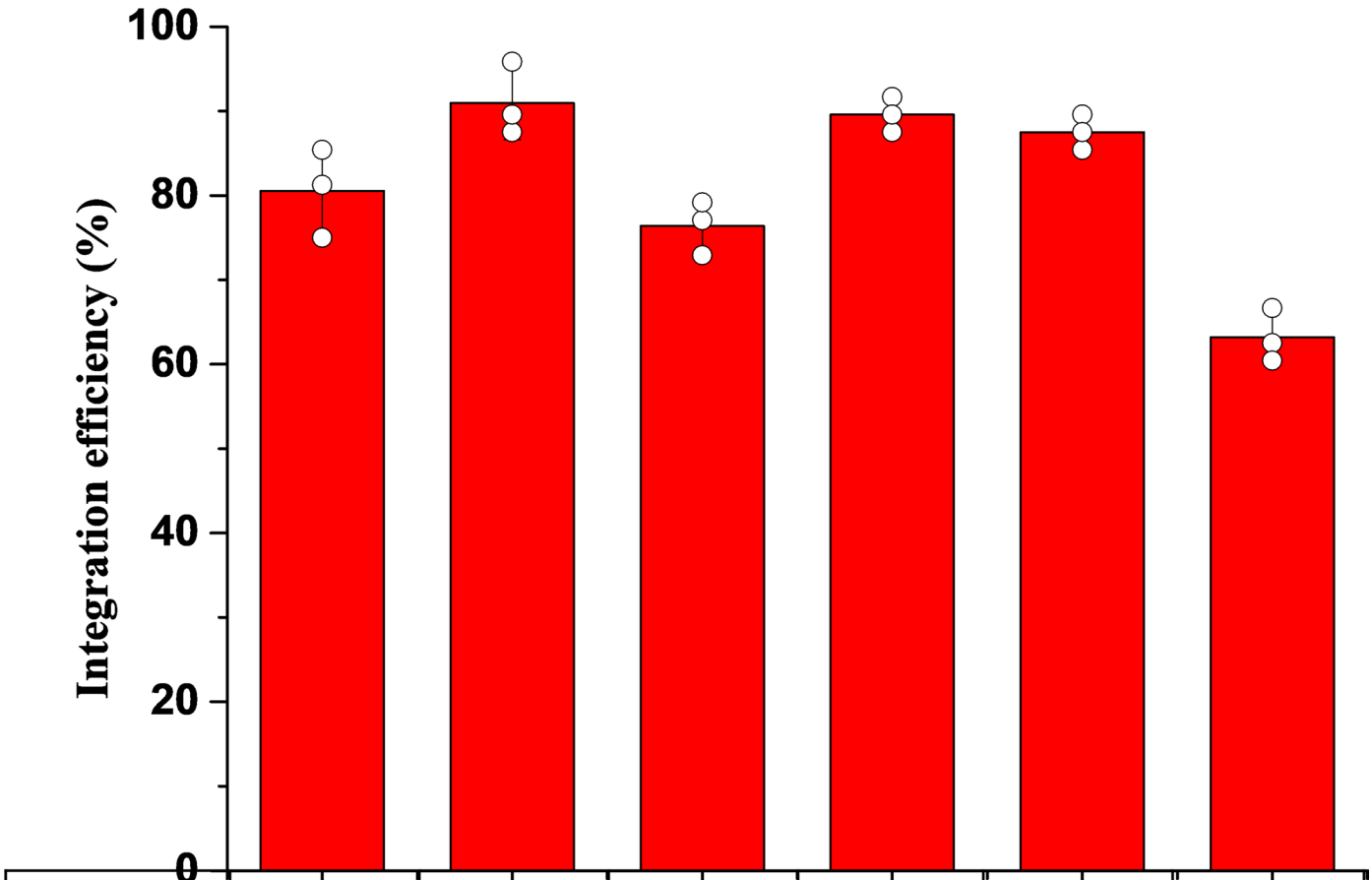
Peer review information *Nature Synthesis* thanks Cameron Kim and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Thomas West, in collaboration with the *Nature Synthesis* team.

https://doi.org/10.1038/s44160-022-00205-2



Extended Data Fig. 1 | Evaluation of the stability of heterologous genes integrated into the selected intergenic regions. Evaluation of the stability of heterologous genes integrated into the selected intergenic regions. A single yeast colony containing the *mCherry* expression cassette was cultured in YPD medium for ~30 generations, and then streak out into YPD agar plates. The stability of evaluated by the percentage of mCherry fluorescent (pink) colonies. IntE4, IntE5, IntE12, IntE15, IntE16, IntE20, IntE24, IntE30, and IntE31 were shown as representative examples.

https://doi.org/10.1038/s44160-022-00205-2

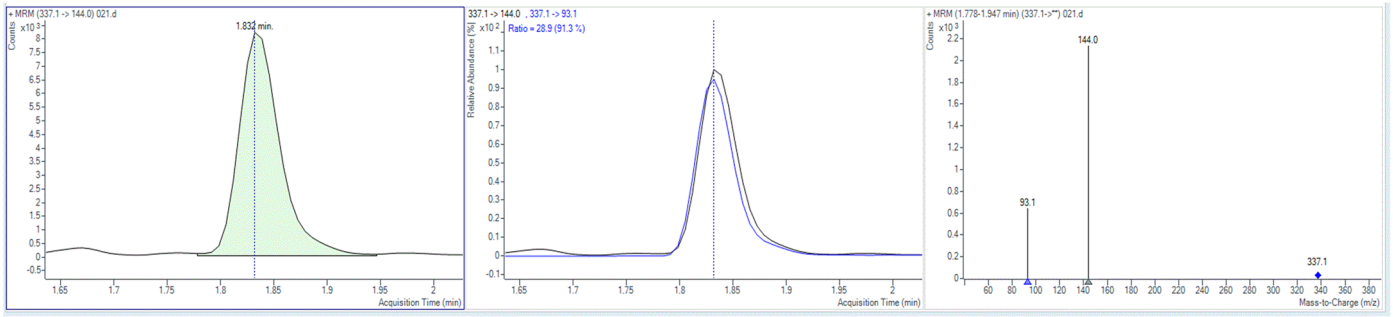


IntE12	+					
IntE17		+				
IntE20		+				
IntE21			+			+
IntE22				+		
IntE24	+					
IntE25				+		
IntE26					+	+
IntE31			+		+	+

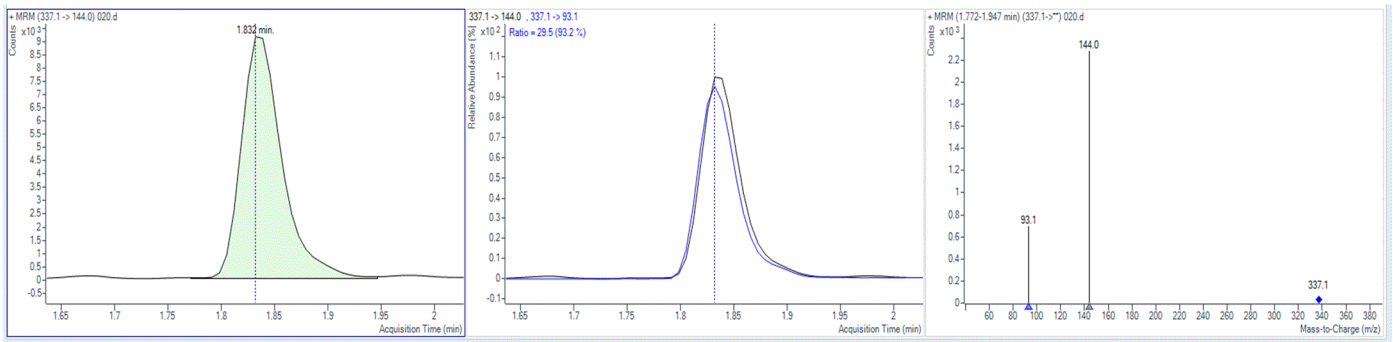
Extended Data Fig. 2 | Multiplex genome integration efficiency. Multiplex genome integration efficiency. *mVenus*, *mCherry*, and *HygB* were chosen as the reporter genes to test the efficiency of editing several loci simultaneously. For two-loci integration, *mCherry* (*TEF1p-mCherry-PRX5t*) and *mVenus* (*GAPp-mVenus-AOX1t*) were used as reporters and the genome integration efficiency was calculated as the percentage of cells showing both mCherry and mVenus fluorescence. For three-loci integration, *mCherry* (*TEF1p-mCherry-PRX5t*), *mVenus* (*GAPp-mVenus-AOX1t*), and *HygB* (*TEF1p-HygB-PRX5t*) were used as reporters and the genome integration efficiency was calculated as the percentage of colonies showing mCherry and mVenus fluorescence as well as hygromycin resistance. The results represent the mean ± s.d. of biological triplicates ($n = 3$).

https://doi.org/10.1038/s44160-022-00205-2

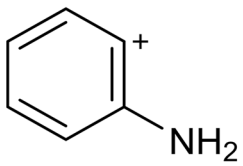
A) CAN4A Sample



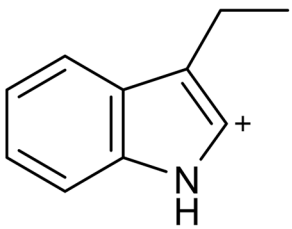
B) Catharanthine Standard



C) Proposed ion structures



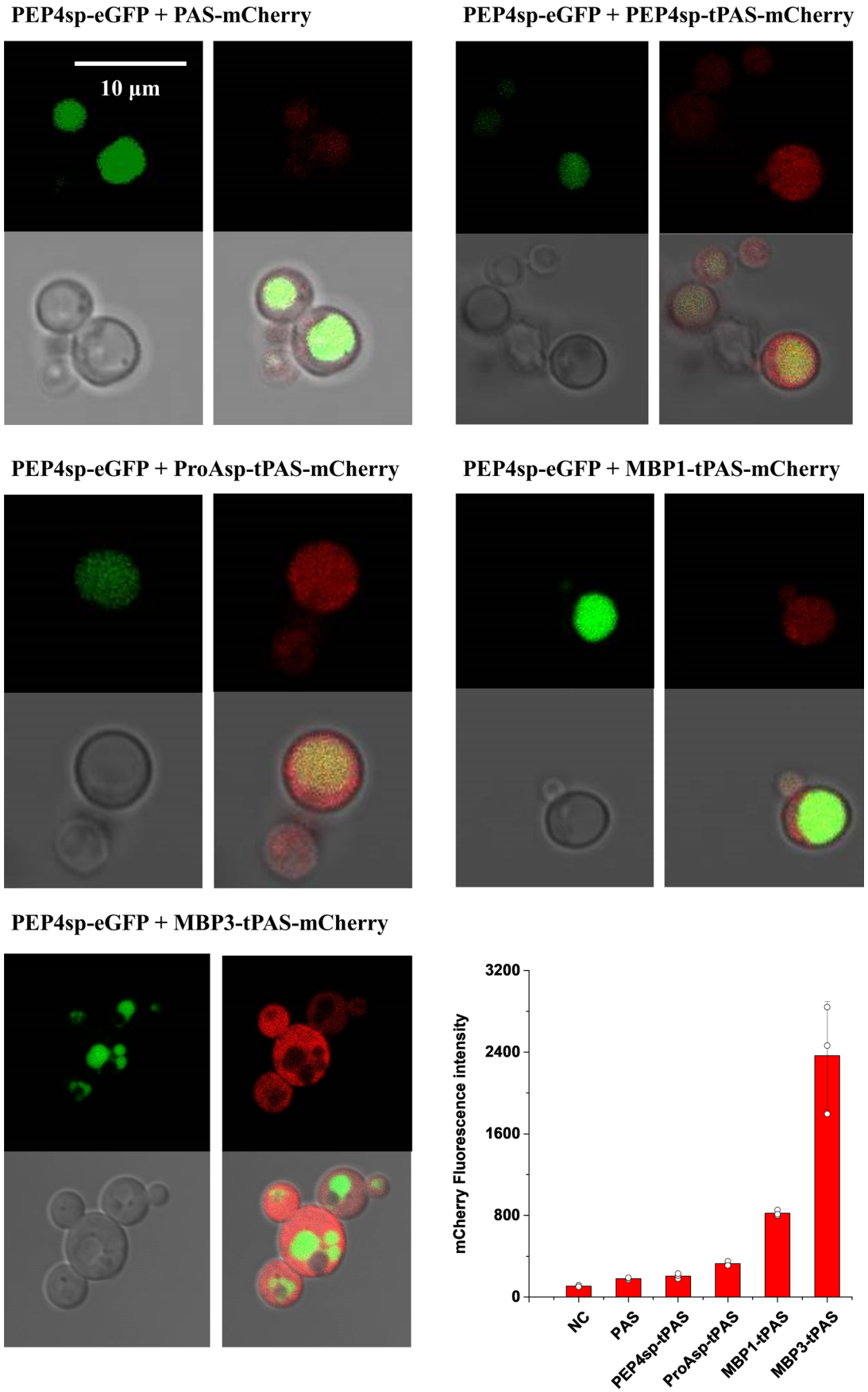
24
m/z=93.1



25
m/z=144.0

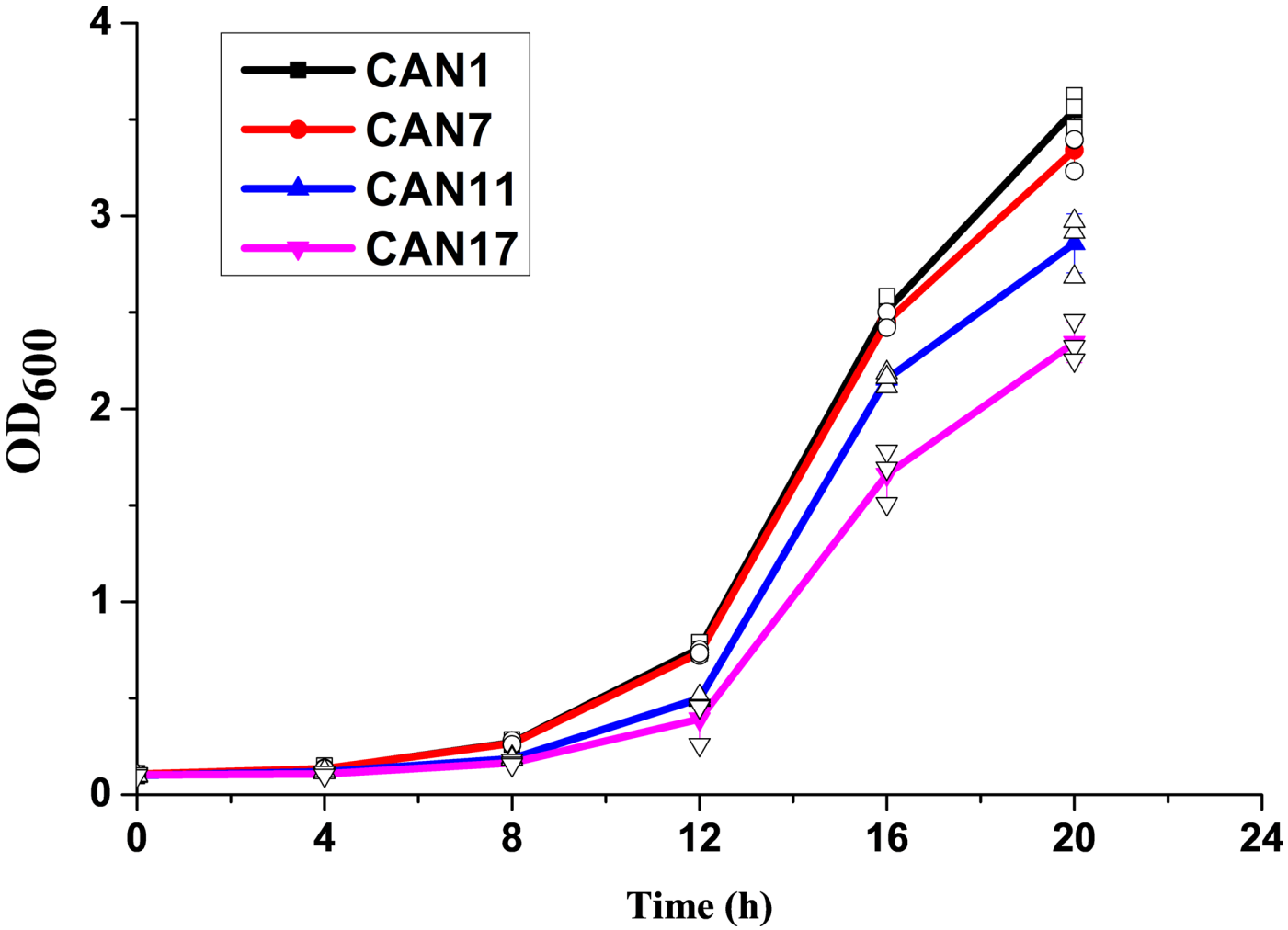
Extended Data Fig. 3 | LC-MS analysis of the production of catharanthine. LC-MS analysis of the production of catharanthine. MRM chromatograms and spectra (*m/z* = 337.1 > 144.0 and *m/z* = 337.1 > 93.1) for CAN4A sample (**a**) and catharanthine standard (**b**). (**c**) Proposed ion structures of the two MS/MS fragments of catharanthine with an *m/z* of 144.0 and 93.1, respectively.

https://doi.org/10.1038/s44160-022-00205-2



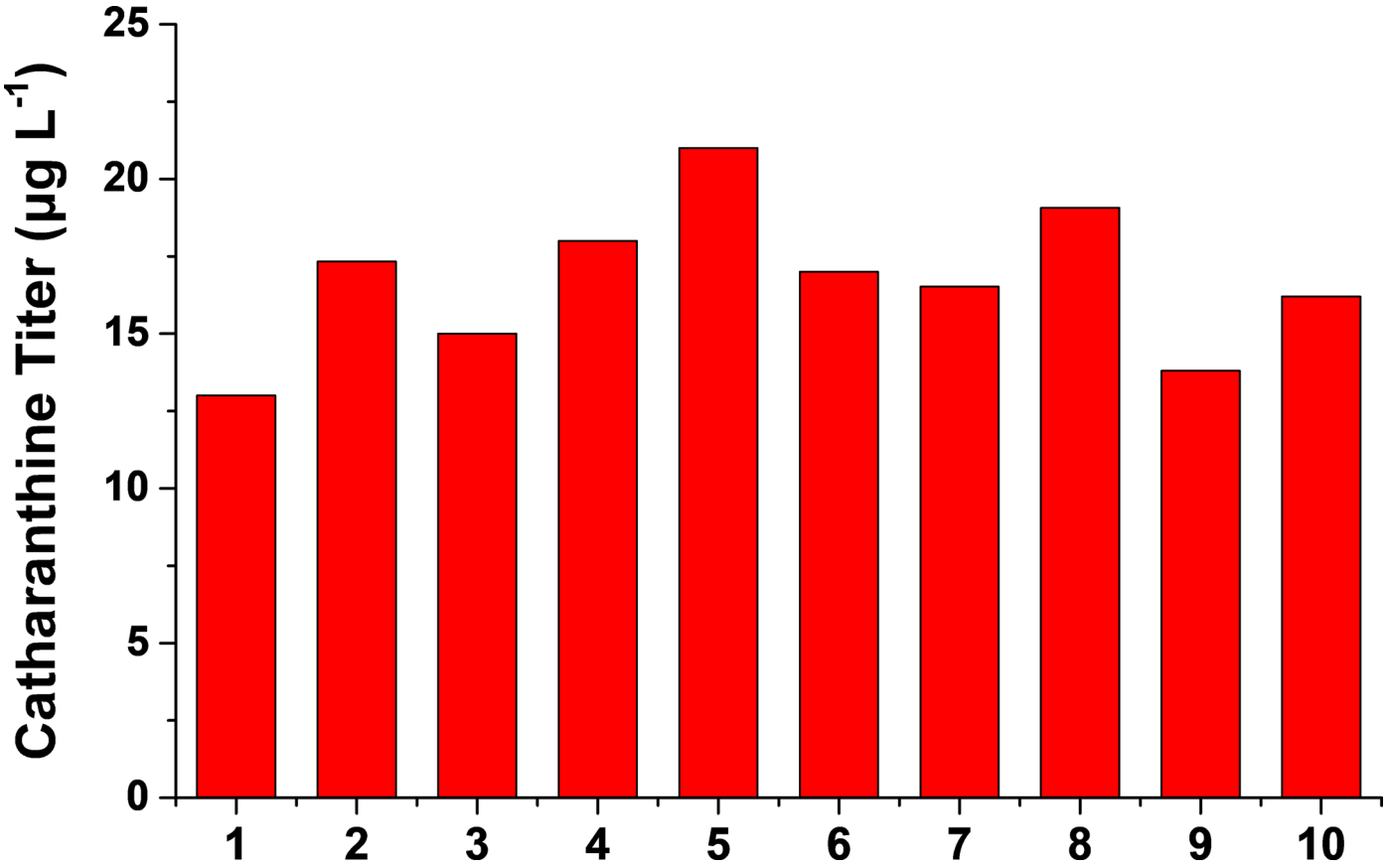
Extended Data Fig. 4 | Confocal microscopy analysis of the localization of PAS variants expressed in *P. pastoris*. Confocal microscopy analysis of the localization of PAS variants expressed in *P. pastoris*. The vacuole was targeted by eGFP (PEP4sp-eGFP), while PAS variants were fused with mCherry and their localization was analyzed by comparing the green (500/540 nm) and red (570/670 nm) fluorescence. Yeast cells were cultured in YPM medium until mid-log phase and subject to confocal microscopy analysis (OLYMPUS IX83-FV3000). The results represent the mean ± s.d. of biological triplicates ($n = 3$).

https://doi.org/10.1038/s44160-022-00205-2



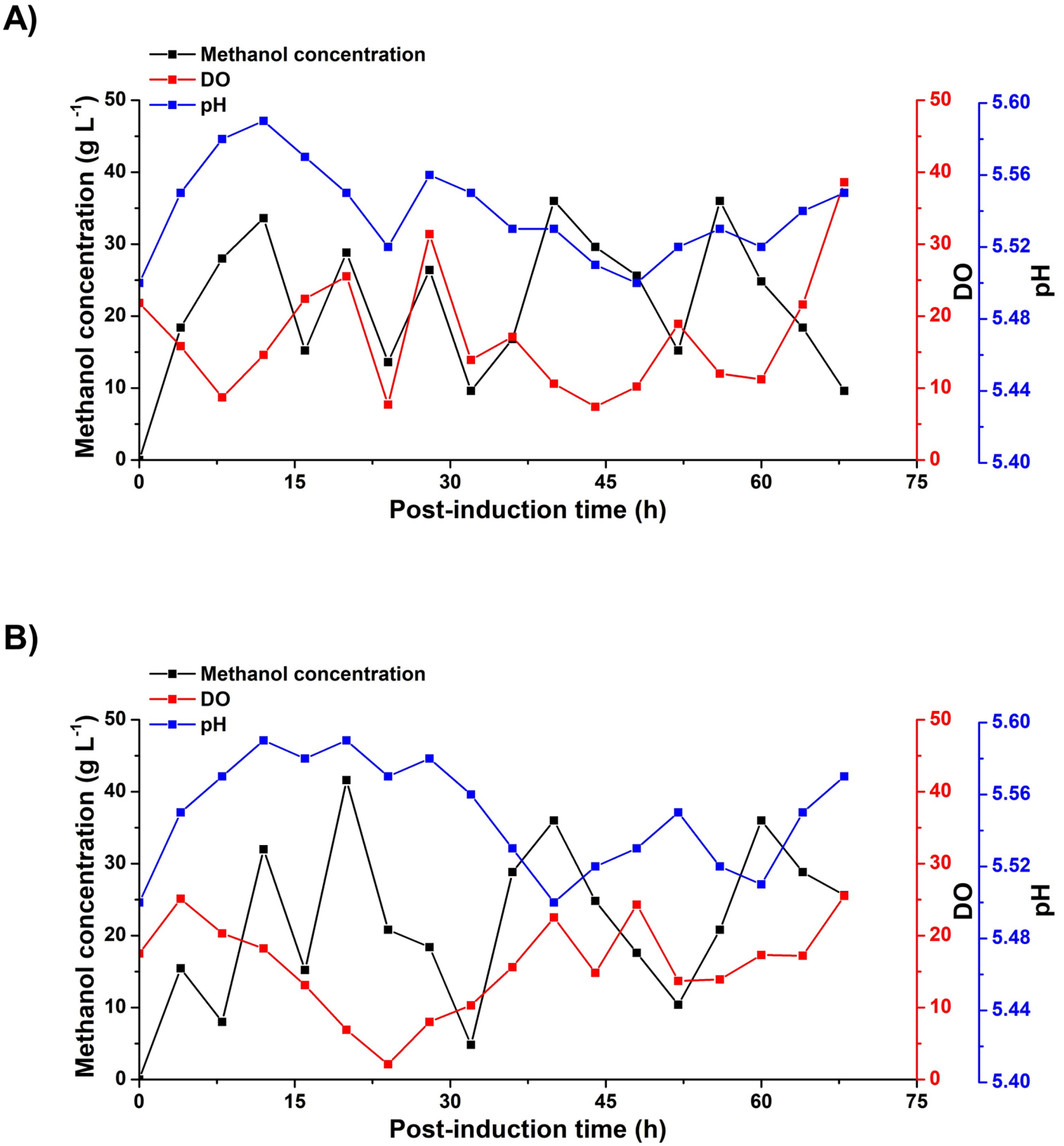
Extended Data Fig. 5 | Growth curves of *P. pastoris* strains, wild-type strain (CAN1) and catharanthine-producing strains (CAN7, CAN11, and CAN17). Growth curves of *P. pastoris* strains, wild-type strain (CAN1) and catharanthine-producing strains (CAN7, CAN11, and CAN17). A single colony of *P. pastoris* was pre-cultured in YPAD medium until saturation and then inoculated into YPAM medium with 2% inoculum. OD_{600} was measured every 4 h. The results represent the mean ± s.d. of biological triplicates ($n = 3$).

https://doi.org/10.1038/s44160-022-00205-2



Extended Data Fig. 6 | Stability evaluation of *P. pastoris* CAN17. Stability evaluation of *P. pastoris* CAN17. Strain CAN17 was cultured in non-selective YPAM medium for ~72 generations, and then spread into YPAD plates. 10 colonies were randomly picked and evaluated for fermentative production of catharanthine (with methanol as the sole carbon source). The production of catharanthine in all clones as well as their consistency in catharanthine titer indicated the stability of the catharanthine-producing *P. pastoris* strain.

https://doi.org/10.1038/s44160-022-00205-2



Extended Data Fig. 7 | Fed-batch fermentation profiles of methanol concentration, dissolved oxygen (DO), and pH after induction. Fed-batch fermentation profiles of methanol concentration, dissolved oxygen (DO), and pH after induction. Methanol (**a**) or methanol/mannitol (**b**) was fed into the bioreactor as carbon sources for fermentative production of catharanthine. After induction for 30 h, 100 mL 5-fold concentrated YPA medium was supplemented to the bioreactors as additional nitrogen source.

nature portfolio

Corresponding author(s): Jiazhang Lian

Last updated by author(s): Oct 6, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	MassHunter (Agilent) was used for LC-MS data collection; AttuneTM NXT Software v4.2 was used to collect flow cytometry data
Data analysis	MassHunter (Agilent) was used for LC-MS data analysis; FlowJo version 10.8.1 was used to analyze flow cytometry data; Excel and OriginPro were used for data analysis; Benchling CRISPR online tool (https://benchling.com/crispr) was used for gRNA design.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data involved in the research are included in the manuscript, supplementary materials, and supplementary data. All relevant data are available upon reasonable request from the corresponding author.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size with at least biological triplicates was chosen by following previous publications in the synthetic biology field.
Data exclusions	No data were excluded from analysis.
Replication	All experiments were performed with at least biological triplicates.
Randomization	Yeast colonies were randomly picked from the agar plates.
Blinding	Not applicable.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Yeast cells were inoculated into 1 mL YPD/zeocine medium in 96-well plates, cultured until mid-log phase (approximately 16 h), collected by centrifugation, washed twice by 10 mM PBS, filtered through a 70 µm cell sieve, and then subject to flow cytometry analysis. The number of injected cells was set at 10,000 and the flow rate was 12.5 µL/min. Excitation laser: 561 nm, emission filter: 620, and bandwidth: 15.
Instrument	AttuneTM NXT, Flow Cytometer (Thermo Fisher Scientific)
Software	AttuneTM NXT Software v4.2 was used to collect flow cytometry data, and FlowJo version 10.8.1 was used to analyze flow cytometrydata.

Cell population abundance	P. pastoris cells were gated based on the cell's size and complexity during the flow cytometry assay. 10,000 of gated cells were collected for analysis.
Gating strategy	Live P. pastoris cells were gated by FSC and SSC. mCherry fluorescence was detected by YL2 channel (with 561 nm excitation and 620 nm emission). No boundaries between 'positive' and 'negative' cell population were defined.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.