

Leveraging the Hermes Transposon to Accelerate the Development of Nonconventional Yeast-based Microbial Cell Factories

Yuxin Zhao, Zhanyi Yao, Deon Ploessl, Saptarshi Ghosh, Marco Monti, Daniel Schindler, Meirong Gao, Yizhi Cai, Mingqiang Qiao, Chao Yang,* Mingfeng Cao,* and Zengyi Shao*



Cite This: *ACS Synth. Biol.* 2020, 9, 1736–1752



Read Online

ACCESS |



Metrics & More



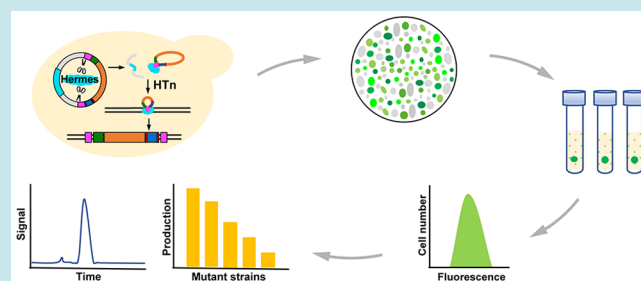
Article Recommendations



Supporting Information

ABSTRACT: We broadened the usage of DNA transposon technology by demonstrating its capacity for the rapid creation of expression libraries for long biochemical pathways, which is beyond the classical application of building genome-scale knockout libraries in yeasts. This strategy efficiently leverages the readily available fine-tuning impact provided by the diverse transcriptional environment surrounding each random integration locus. We benchmark the transposon-mediated integration against the nonhomologous end joining-mediated strategy. The latter strategy was demonstrated for achieving pathway random integration in other yeasts but is associated with a high false-positive rate in the absence of a high-throughput screening method. Our key innovation of a nonreplicable circular DNA platform increased the possibility of identifying top-producing variants to 97%. Compared to the classical DNA transposition protocol, the design of a nonreplicable circular DNA skipped the step of counter-selection for plasmid removal and thus not only reduced the time required for the step of library creation from 10 to 5 d but also efficiently removed the “transposition escapers”, which undesirably represented almost 80% of the entire population as false positives. Using one endogenous product (*i.e.*, shikimate) and one heterologous product (*i.e.*, (S)-norcoclaurine) as examples, we presented a streamlined procedure to rapidly identify high-producing variants with titers significantly higher than the reported data in the literature. We selected *Scheffersomyces stipitis*, a representative nonconventional yeast, as a demo, but the strategy can be generalized to other nonconventional yeasts. This new exploration of transposon technology, therefore, adds a highly versatile tool to accelerate the development of novel species as microbial cell factories for producing value-added chemicals.

KEYWORDS: genome integration, nonconventional microbes, Hermes transposon, nonhomologous end joining, shikimate, (S)-norcoclaurine



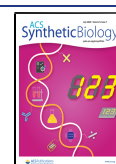
The advancement of metabolic engineering and synthetic biology in the past two decades has enabled manipulation of microorganisms as cell factories for the production of various value-added specialty chemicals from inexpensive feedstock, many of which have applications in the pharmaceutical, nutraceutical, cosmetic, and food industries.^{1,2} The construction of an efficient production strain requires many optimization steps but, nevertheless, starts from assembling the pathway from both native and heterologous biocatalysts to establish initial production. If an amenable model microorganism (*e.g.*, *Escherichia coli* or *Saccharomyces cerevisiae*) is selected as a production host, a pathway can be built in an episomal plasmid or integrated into the genome; if a non-model host (also known as a nonconventional host)³ is selected, for which stable episomal vectors are not readily available, the strategy of integrated expression remains prevalent. Recently, interest has shifted toward tailoring host selection to target products, desired feedstock, and/or preferred fermentation conditions. Many previously underutil-

ized species, such as *Rhodospiridium toruloides*,^{4–6} *Scheffersomyces stipitis*,^{7,8} and *Kluyveromyces marxianus*,^{9,10} have received unprecedented attention owing to their unique biochemical and physiological features including lipogenesis, C5 sugar assimilation, and thermotolerance. However, the toolkit for effectively manipulating these seemingly advantageous hosts is incomplete, and our knowledge of phenotype-genotype correlations in these species is insufficient. These barriers prevent the development of non-model hosts from meeting commercialization criteria.

Genome integration approaches have shifted from leveraging endogenous chromosomal double-strand breaks (DSBs)

Received: February 28, 2020

Published: May 12, 2020



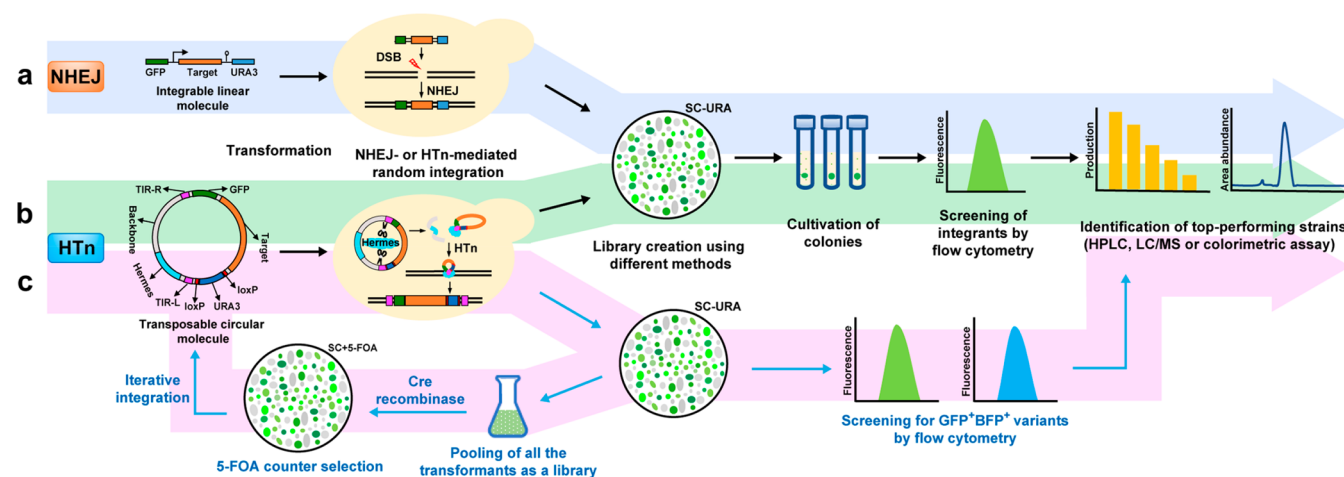


Figure 1. Schematic illustration of the NHEJ and HTn-mediated random integration methods. (a) For the NHEJ-mediated integration, a linear integrable fragment containing the target gene or pathway expression cassettes flanked by the *GFP* and *URA3* cassettes is transformed into the host strain. Only a portion of the transformants growing on the SC-URA agar medium are *GFP*⁺, since the rest only have *URA3* integrated. The *GFP*⁺ colonies are inoculated into the liquid SC-URA medium, and their *GFP* expression profiles are evaluated by flow cytometry. This step is used to remove the false positives that express *GFP* in a nonintegrated manner. The transformants confirmed to harbor the integrated *GFP* are subsequently tested for the production of the target product by high-performance liquid chromatography (HPLC), LC/MS, or colorimetric analysis. (b) For the HTn-mediated integration, a nonreplicable circular DNA molecule consisting of Hermes transposase and TIR-flanked *GFP*-target pathway-*loxP*-*URA3*-*loxP* fragment is transformed into the host, and the rest of the steps share the same protocol with the NHEJ-mediated method if only one round of transposition is performed. (c) If iterative integration is desired, all the clones appearing on the SC-URA plate are pooled as a library. The integrated *URA3* marker is then removed by *Cre-loxP* recombination followed by 5-FOA counter-selection. The resulting library is ready for next-round transpositional integration. To facilitate the screening, a new fluorescence protein (e.g., BFP) could be optionally incorporated in the transposable element to replace *GFP* in the second round of integration. Flow cytometry is used to identify *GFP*⁺*BFP*⁺ integrants. The usage of multiple fluorescence markers is not a design requirement due to the high efficiency of the HTn-mediated integration.

repaired by either homologous recombination^{11,12} or exogenously introduced recombinase^{13–17} to the widespread usage of the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas (CRISPR-associated) systems to enhance the probability of a DSB occurring at a target locus.^{18–20} However, it is time-consuming to identify a favorable locus for balanced pathway expression and optimal production outcome. Transcription is dependent on “chromosomal position effect”,²¹ which is generally a result of integration into either the accessible euchromatin regions or the relatively packed heterochromatin regions. An expression cassette located in different genomic regions undergoes different levels of transcriptional regulation based on the accessibility to the transcription factors and the dramatically different levels of local transcriptional competition. Studies in *E. coli*,^{22,23} *S. cerevisiae*,^{24,25} and other organisms^{26–28} have demonstrated that expression levels of a translocated native gene (i.e., α -galactosidase) and foreign genes (i.e., fluorescent reporter proteins) inserted at different loci could vary by ~2- to 300-fold.

Many nonconventional microorganisms, especially eukaryotic microbes, rely on nonhomologous end joining (NHEJ) to repair chromosomal DSBs, which poses severe problems for achieving precise genome modifications. This gap-connecting mechanism has been exploited to rapidly create expression libraries carrying both the pathway and the selection marker integrated at multiple loci (Figure 1a), and this random integration significantly influences the productivity of the strain.^{27,29,30} However, this method has a high false-positive rate, with many transformants carrying only the selection marker and not the intended pathway (see Results section). In the absence of a high-throughput screening method, identification of functional variants is still laborious.

Transposases recognize terminal inverted repeats (TIRs) and facilitate translocation of the enclosed DNA from a donor site to a receiver site. This feature has been commonly used for creating insertion-mutagenesis and transgenesis libraries.^{31,32} The Hermes transposon (HTn), a *hAT* (i.e., *hobo* from *Drosophila*, *Ac* from maize, and *Tam3* from snapdragon) family member, isolated from the housefly *Musca domestica*,³³ has been introduced into the genomes of *S. cerevisiae*,³⁴ *Yarrowia lipolytica*,³⁵ and *Schizosaccharomyces pombe*³⁶ for functional genomics studies. Although the insertion sites of HTn are not totally random, favoring 5'-nTnnnnAn-3' sites in nucleosome-free regions,^{34,36–38} the massive abundance of such sites allows for sufficient quasi-random coverage. For instance, on average, insertions occurred every 24 bp in intergenic and every 98 bp in intragenic regions in *Y. lipolytica*, as well as every 29 bp in nonrepetitive regions in the genome of *S. pombe*. It will be highly valuable to evaluate the usages of HTn and NHEJ to create integrated expression libraries and characterize the false-positive rates.

S. stipitis is considered an important yeast in the field of biorenewables due to its high native capacity for xylose utilization, which is essential for the economic lignocellulose bioconversion.³⁹ Many studies have focused on isolating xylose-assimilating enzymes and transporters from this species and expressing them in model hosts to improve their conversions of xylose, the second most-abundant sugar present in biomass.^{40–42} With a ¹³C-metabolic flux analysis (¹³C-MFA), a previous study has indicated that erythrose 4-phosphate (E4P) is the rate-limiting precursor in the biosynthesis of shikimate in *S. cerevisiae*.⁴³ Considering that xylose is assimilated through the pentose phosphate pathway (PPP), the more active PPP in *S. stipitis* can elevate the E4P level and thereby ultimately increase the production of

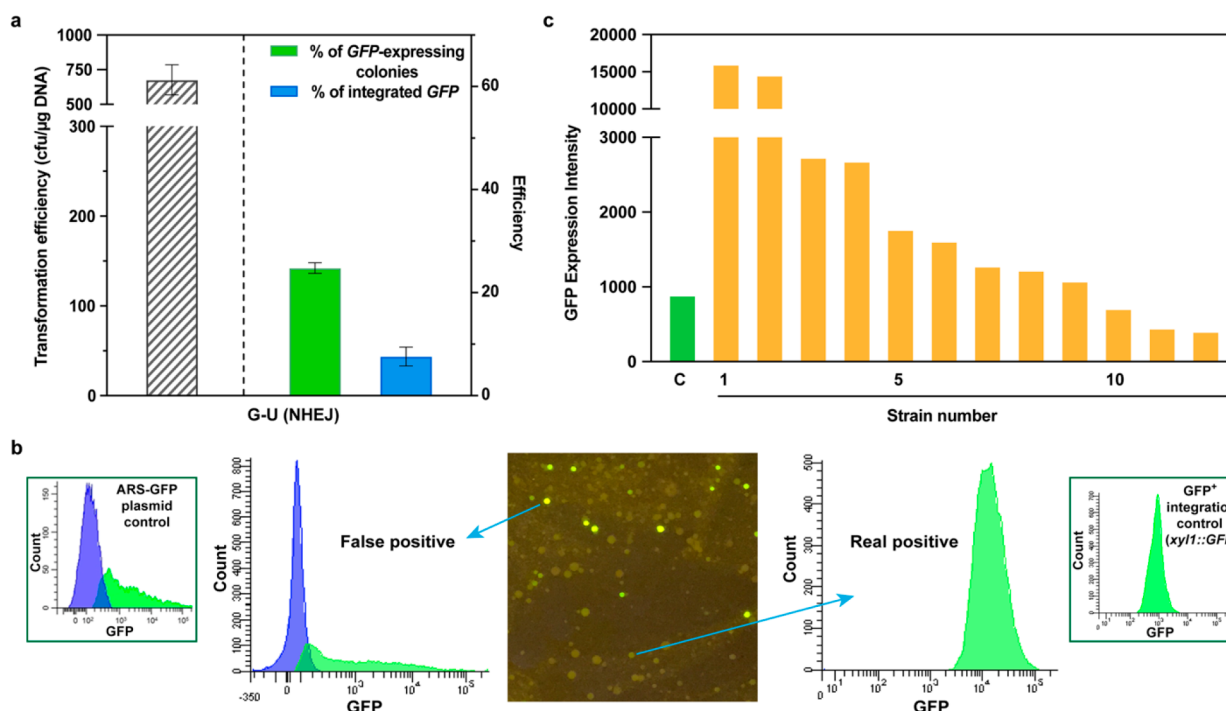


Figure 2. A linear fragment containing the expression cassettes of *GFP* and *URA3* was transformed into the strain *SS ura⁻* to achieve NHEJ-mediated random integration. (a) The transformation and integration efficiency. G-U fragment (1.2 μ g) was electroporated into strain *SS ura⁻* using a voltage of 1.5 kV/mm, and 812 colonies were obtained on average per transformation. (b) The distinct GFP expression profiles revealed by flow cytometry correspond to two different groups of GFP⁺ clones growing on the SC-URA plate. The image was taken under a darker reader transilluminator. The blue peak represents the population without GFP expression, and the green peak represents the population with GFP expression. The width of the green peak indicates the variation of GFP expression among the cells. The strain *SS ura⁻* carrying a single-copy *GFP* gene integrated into the *XYL1* locus was used as the GFP⁺ integration control. The strain *SS ura⁻* expressing *GFP* in an episomal ARS plasmid was used as the negative control. (c) The GFP expression intensity of the *SS ura⁻* variants with randomly integrated G-U fragment was evaluated by flow cytometry after 24 h of cultivation. "C" represents the GFP⁺ integration control strain containing a single-copy *GFP* gene integrated into the *XYL1* locus.

shikimate,⁸ suggesting that it is a high-potential platform host to produce shikimate pathway derivatives, including many kinds of flavonoids and alkaloids with important pharmaceutical and nutraceutical properties.^{44–47} Here, selecting *S. stipitis* as an exemplary nonconventional yeast with an entirely dominant NHEJ mechanism, we demonstrated a streamlined HTn-mediated methodology for rapid identification of variants producing the endogenous product shikimate and the heterologous product (S)-noroclaurine at high levels.

RESULTS

NHEJ-mediated Integration Is Easily Accomplished but Has a High False-positive Rate. We chose *S. stipitis* as the target host to represent the group of nonconventional yeasts that utilize NHEJ as the dominant mechanism to repair chromosomal DSBs. It has been shown that the precise genome editing efficiency of *S. stipitis* is less than 1% with the inherent homologous recombination (HR) mechanism and as low as 16% with the assistance of CRISPR-Cas9, unless the genes responsible for NHEJ were disrupted.^{48,49} However, NHEJ, a potent but error-prone mechanism for DSB repair, can be leveraged for random genome integration and expression library construction. To investigate the efficiency of NHEJ, we designed a linear G-U fragment (4.2 kb) containing the GFP reporter and the *URA3* selection marker expression cassettes and transformed it into a *S. stipitis* FLP-UC7 strain that contains an auxotrophic mutation in the *URA3* gene (hereafter the strain is labeled as *SS ura⁻*). The

transformation efficiency reached 677 cfu/ μ g DNA, but only 25% of the plated colonies exhibited green fluorescence (Figure 2a). Forty green colonies were selected for quantification of GFP expression levels by flow cytometry. Of the 40, only 12 colonies displayed a sharp, uniform peak, similar to that of the control strain (which carried a single copy of the *GFP* gene integrated into the *XYL1* locus) (Figure 2b real positive). The other 28 colonies had broad GFP expression peaks, which were presumably caused by the *in vivo* self-assembly of the G-U fragment with a genome-originating autonomously replicating sequence (ARS);⁵⁰ a similar expression profile has been previously observed in a strain transformed with an unstable plasmid that contains an ARS and misses a plasmid-stabilizing centromeric element (Figure 2b false positive).⁴⁹ The efficiency of G-U genomic integration was therefore calculated to be 7.5%. The GFP intensity of the 12 integration strains showed a broad range from 0.4- to 18-fold of that expressed by the control strain integrated with a single-copy *GFP* gene (Figure 2c).

Despite the high false-positive rate, the NHEJ-mediated random integration is a readily usable method for pathway integration, resulting in a wide range of transcriptional levels, which could be used to screen for high-level production of valuable chemicals. It has been previously shown that the more active pentose phosphate pathway associated with the native xylose assimilation capacity of *S. stipitis* provides a higher carbon flux toward erythrose 4-phosphate (E4P),⁵¹ the rate-limiting precursor of the shikimate pathway.⁴³ Increased E4P

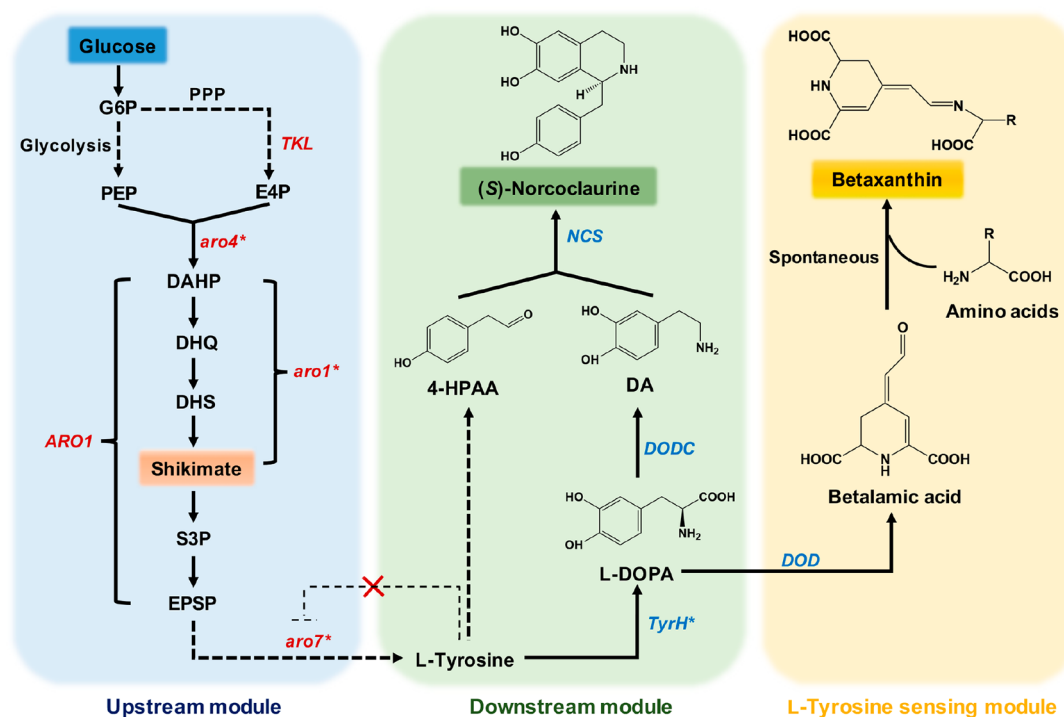


Figure 3. Schematic representation of the shikimate pathway (upstream native module), the (S)-norcoclaurine biosynthetic pathway (downstream heterologous module), and the L-tyrosine sensing module. The minimal set of the shikimate pathway consists of three enzymes, including Tkl, Aro4_{K220L}, and Aro1_{D900A}. The *de novo* biosynthesis of (S)-norcoclaurine requires the overexpression of four enzymes in the upstream module and heterologous expression of two plant-sourced enzyme variants (TyrH_{W13L, F309L} and NCS) and one bacterium-originating enzyme (DODC). The dashed line indicates the feedback inhibition of Aro7 posed by L-tyrosine. The L-tyrosine-sensing module contains betaxanthin pathway, which harbors TyrH_{W13L, F309L} and DOD. Abbreviations of metabolites: G6P, glucose-6-phosphate; PEP, phosphoenolpyruvate; E4P, erythrose 4-phosphate; DAHP, 3-deoxy-D-arabino-heptulosonate 7-phosphate; DHQ, 3-dehydroquinic acid; DHS, 3-dehydroshikimate; S3P, shikimate-3-phosphate; EPSP, 5-enolpyruvylshikimate-3-phosphate; L-DOPA, L-3,4-dihydroxyphenylalanine; DA, dopamine; 4-HPAA, 4-hydroxyphenylacetaldehyde. Enzymes: PPP, pentose phosphate pathway; Tkl, transketolase; Aro4*, feedback inhibition-insensitive mutant (Aro4_{K220L}) of DAHP synthase; Aro1, pentafunctional aromatic enzyme; Aro1*, Aro1 mutant (Aro1_{D900A}) with the shikimate kinase domain inactivated; Aro7*, feedback-resistant chorismate mutase variant (Aro7_{G139S}); TyrH*, tyrosine hydroxylase mutant (TyrH_{W13L, F309L}) from *Berberis vulgaris* with a low L-DOPA oxidase activity; DOD, L-DOPA dioxygenase from *Mirabilis jalapa*; DODC, L-DOPA decarboxylase from *Pseudomonas putida*; NCS, norcoclaurine synthase from *Papaver somniferum*.

availability enables *S. stipitis* to produce substantially more shikimate than *S. cerevisiae*.⁸ The minimal set of enzymes accounting for shikimate accumulation consists of transketolase (Tkl), the feedback inhibition-insensitive mutant of 3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase (Aro4_{K220L}), and the pentafunctional aromatic enzyme mutant with an inactivated shikimate kinase domain (Aro1_{D900A}) that enables shikimate accumulation (Figure 3).^{52–54} A linear G-SA-U fragment (14.9 kb, enclosing the shikimate pathway between the GFP and URA3 expression cassettes) was transformed into strain SS ura[−] with a transformation efficiency of 248 cfu/μg DNA (Figure 4a). Of the plated colonies, 13% exhibited green fluorescence; of those, 60% carried the integrated GFP gene as verified by flow cytometry. Further investigation of these URA⁺GFP⁺ integrants showed that the majority of them (92%) could produce shikimate. The efficiency of random integration of the G-SA-U pathway was therefore calculated to be ~7.2%, consistent with the efficiency of integrating the G-U fragment. No transformants were obtained when the linear G-SA-U fragment was transformed into the SS ura[−] *ku70Δku80Δ* strain, in which the NHEJ mechanism was disrupted,⁴⁸ indicating that random integration was indeed mediated by NHEJ in *S. stipitis*.

Fifteen strains confirmed to carry the integrated G-SA-U were randomly picked to evaluate shikimate production after a

5 d fermentation (Figure 4b). The titers varied by 7.7-fold, ranging from 0.44 ± 0.01 to 3.4 ± 0.1 g/L. The highest producer, strain SA1, was selected to test the stability of pathway expression. Twenty colonies were randomly selected from the restreaked SA1 and compared with a control strain that carries the same shikimate fragment by a previously reported episomal plasmid.⁸ Shikimate production via the integrated pathway was substantially more stable than the production by the episomal expression plasmid, which varied dramatically among the clones, likely as a result of plasmid instability (Figure 4c). Furthermore, GFP expression did not correlate with shikimate production, indicating that GFP could only be used to facilitate the screening of potential integrants (Figure 4b). Additionally, this confirms the notion that selecting a predetermined locus for heterologous pathway expression is tedious and ineffective, as a higher level of transcription does not necessarily equate to an increased pathway flux presumably due to the imbalanced gene expression in a multistep pathway. Integration by NHEJ occurs randomly at different loci, each presenting a distinct transcriptional environment. This strategy therefore automatically creates a fine-tuning mechanism to screen for expression balancing in multigene pathways and thus high production levels.

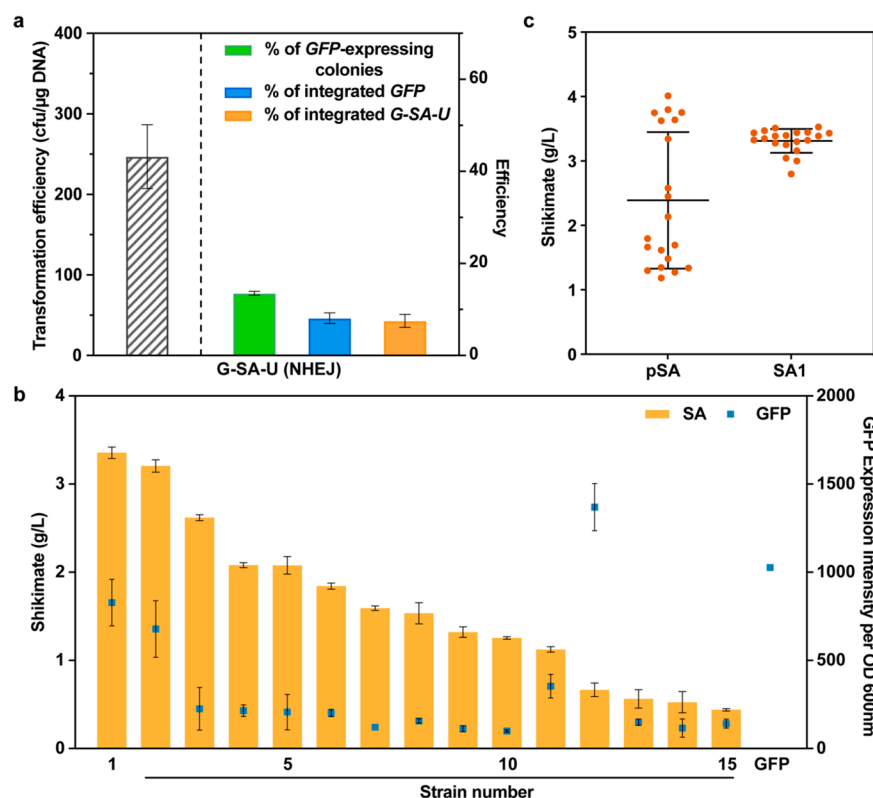


Figure 4. A linear DNA fragment containing the shikimate pathway flanked by the expression cassettes of *GFP* and *URA3* was transformed into strain *SS ura⁻* to achieve random pathway integration via NHEJ. (a) The transformation and integration efficiency. G-SA-U DNA fragment (1.2 μg) was electroporated into strain *SS ura⁻* using a voltage of 1.5 kV/mm, and 297 colonies were obtained on average per transformation. The percentage of integrated *GFP* was calculated based on multiplication of the percentage of *GFP*-expressing colonies in total and the percentage of the confirmed genome integration based on flow cytometry of the *GFP* expression profile. The number shown inside each bar indicates the average total number of observed colonies belonging to each group per transformation of 1.2 μg of DNA. (b) Shikimate production and normalized *GFP* expression intensity of the *SS ura⁻* variants carrying the randomly integrated G-SA-U fragment after 120 h of cultivation. Error bars indicate mean $\pm$ standard deviation (s.d.) of three biological replicates. (c) Comparison of the shikimate production between integrated expression and plasmid expression in the strain *SS ura⁻* after 120 h of cultivation. SA1: the highest producer was restreaked, and 20 colonies were randomly selected; pSA: 20 colonies were picked from the streaked strain *SS ura⁻* carrying the shikimate pathway in a CEN/ARS plasmid, whose stability was already improved significantly over the previous ARS plasmid.^{49,55} Error bars indicate the mean $\pm$ s.d. of 20 biological replicates.

HTn-mediated Random Integration of the Shikimate Pathway using a Nonreplicable Circular Molecule. The NHEJ mechanism results in a low pathway integration efficiency; the majority of the transformants appearing on SC-URA plates are only guaranteed to carry a functional *URA3* gene. Even in the small subset of the transformants that are both *URA3⁺* and *GFP⁺*, many clones are non-integrants. To address these issues, we sought to implement a DNA transposon-mediated integration strategy and compare it with the NHEJ-mediated integration. To date, integration of a large biosynthetic pathway using the former method has been demonstrated in bacteria,^{56–59} but, surprisingly, an analogous concept has not been adapted for yeasts.

Our design was based on the HTn system that has been used to study gene essentiality in yeasts.^{34–36} In these studies, the Hermes transposase was expressed via a plasmid carrying a selection marker (e.g., *URA3*) on the plasmid backbone, with another selection marker (e.g., *LEU2*, *NatMX*, or *KanMX6*) flanked by TIRs (Figure 5a). The transformants were selected on the corresponding selection plates, and a secondary 5-fluoroorotic acid (5-FOA) counter-selection was applied to remove clones containing the plasmid in the subsequent step. However, many “transposition escapers” surviving from both selections dominated as false positives. In our study in *S.*

stipitis, the percentage of transposition escapers reached 79%, presumably because the SC-LEU condition cannot differentiate integrated expression versus episomal expression of *LEU2*.

To overcome this barrier, we designed a nonreplicable circular DNA molecule for *S. stipitis*. The Hermes transposase expression cassette was assembled with the TIR-flanked G-SA-U fragment and other helper elements using a DNA assembler method reported previously.^{60–62} The circular molecule was built in *S. cerevisiae* as a plasmid via homologous recombination and transformed into *E. coli* for plasmid enrichment. As there is no replicating element (i.e., ARS) specific to *S. stipitis* in the transformed molecule, the transposition escapers were automatically removed. Therefore, only one selection marker was needed, and the cells that underwent transposition were identified in a single round of selection (Figure 5b). Moreover, using a circular molecule as a carrier potentially reduced the interference from the NHEJ-mediated integration, which has been reported to act preferentially on linear DNA.⁶³

To evaluate the integration efficiency, the nonreplicable circular molecule cHTn-GSAU (G-SA-U fragment flanked by TIRs) and two control circular molecules (either lacking the transposase or one side of the TIRs) were transformed into the *SS ura⁻* strain in parallel (Figure 1b). The percentage of *GFP*-

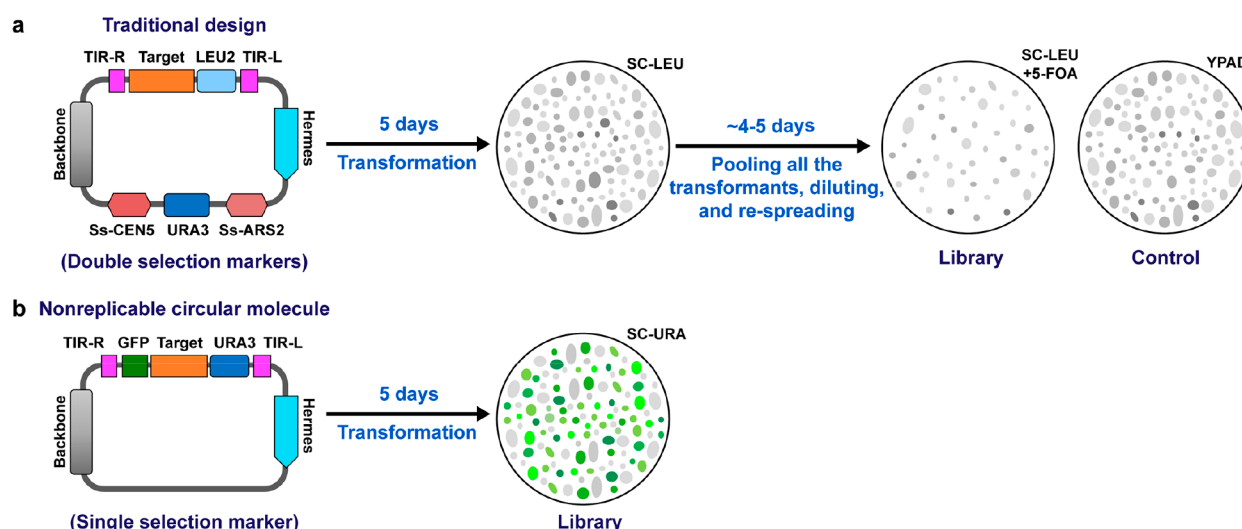


Figure 5. Comparison of the traditional transposition protocol and the new design based on a nonreplicable circular DNA molecule. (a) The traditional design reported in the previous studies.^{34–36} A plasmid consists of a Hermes transposase expression cassette, TIR-flanked transposable target region containing a selection marker (e.g., *LEU2*), ARS, and CEN (the elements required for plasmid replication and stable segregation in the host), and a second selection marker *URA3* for subsequent plasmid removal. The transposition efficiency was calculated based on the ratio of the colony numbers on SC-LEU+5-FOA and YPAD plates. The usage of a replicable plasmid causes the appearance of many transposition escapers, because the SC-LEU condition cannot differentiate integrated expression vs episomal expression of *LEU2*. (b) The improved design relies on a nonreplicable circular molecule, which does not contain ARS and CEN specific to *S. stipitis*, and only uses one selection marker. Only the cells that undergo transposition can propagate on the SC-URA medium, thus reducing the number of transposition escapers and shortening the experiment protocol.

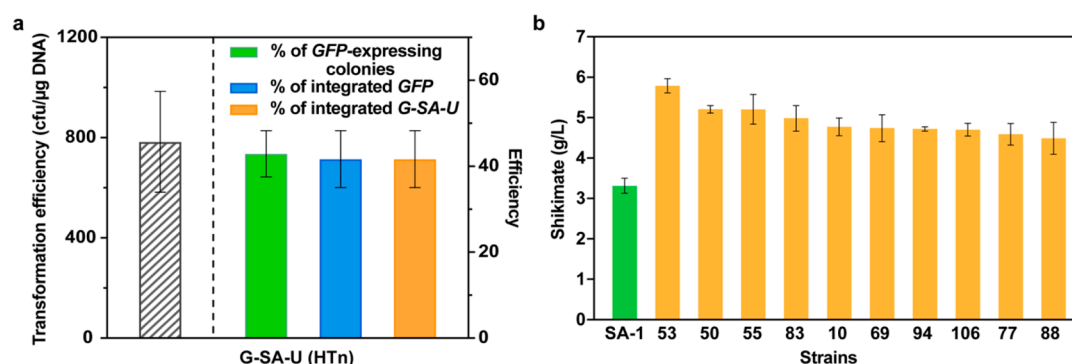


Figure 6. The transposition efficiency of cHTn-GSAU in the SS *ura*[−] strain and the shikimate production by the variants carrying the randomly integrated shikimate pathway. (a) The transformation and transposition efficiency. cHTn-GSAU (1.2 μg) was electroporated into strain SS *ura*[−] using a voltage of 1.8 kV/mm, and 940 colonies were obtained on average per transformation. (b) Shikimate production by the SS *ura*[−] variants that underwent transpositional insertion of the G-SA-U fragment. The top 10 shikimate producers were selected from 40 variants (summarized in Figure S2), and their shikimate production levels were retested after 120 h of fermentation with three biological replicates. SA1, the best producer created based on NHEJ-mediated integration, was used as a benchmark.

expressing colonies observed in the cHTn-GSAU group was 43%, in sharp contrast to the low percentage (~1%) of GFP-expressing colonies observed in the controls (Figure S1), indicating that GFP expression was likely obtained via transpositional integration. To further confirm whether they were true integrants, GFP⁺URA⁺ colonies were randomly selected for the measurement of shikimate production and GFP expression. As verified by flow cytometry, 97% of the GFP-expressing clones turned out to carry the integrated *GFP* gene, and all of these integrants produced shikimate (Figure 6a). The cHTn-GSAU library generated a 20-fold range of shikimate titers (Figure S2). Fermentation of the top 10 high-production variants confirmed that all produced shikimate at titers higher than that of the best producer SA1 created by the NHEJ-mediated integration, among which the maximum titer reached 5.8 ± 0.2 g/L (Figure 6b). Consistently, GFP intensity

failed to correlate with the production level, emphasizing that GFP can only be used for the initial round of screening. On the basis of these results, the circular HTn-based platform had an integration efficiency of 42% (i.e., $43\% \times 97\% \times 100\%$), which is 2-fold and 5.5-fold of that obtained with the plasmid-based HTn platform (Figure 5a) and that obtained with the NHEJ-mediated random integration, respectively (Figure 4a). Furthermore, while 40% of GFP⁺URA⁺ transformants obtained by transforming the linear G-SA-U fragment were not true integrants, the colonies isolated from the circular HTn-based integration platform had a false positive rate of only 3%. Therefore, the nonreplicable circular DNA platform streamlined the single-step library creation and the significantly increased integration efficiency indicated the feasibility of using HTn for the integration of more complex biosynthetic pathways. With the same amount of screening effort, the low

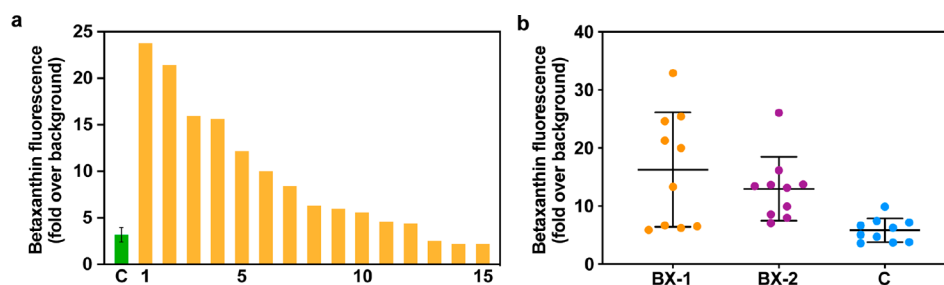


Figure 7. Development of an L-tyrosine sensor and screening for the mutants with a higher flux in the upstream module. (a) The knockout variants from the transpositional insertion library produced betaxanthin at different levels when transformed with the sensor plasmid. Fluorescence values were normalized to the autofluorescence of a negative control strain (strain SS *ura*⁻ transformed with a plasmid expressing just DOD). “C” represents the parental SS *ura*⁻ strain transformed with the sensor plasmid. (b) A secondary measurement to confirm a higher flux in the upstream module in the top two variants BX-1 and BX-2. Ten transformants were randomly selected for each variant. Error bars indicate mean $\pm$ s.d. of 10 biological replicates.

false-positive rate increased the probability of identifying high-producing variants.

Identification of Knockout Variants Suitable for Producing Benzylisoquinoline Alkaloids. Shikimate, a committed precursor in the aromatic amino acids biosynthesis, serves as a reporter for assessing flux toward aromatic amino acids. Considering that *S. stipitis* enables a higher flux toward the shikimate pathway, we proceeded to explore its potential to produce higher-value products downstream from the aromatic amino acids. As proof of concept, (S)-norcoclaurine, the entrance molecule to the synthesis of the diverse family of compounds named benzylisoquinoline alkaloids (BIAs), was selected as the target. BIAs are L-tyrosine-derived, specialized plant secondary metabolites with many potent pharmacological properties. Currently, exemplary compounds, such as the anticancer agent sanguinarine,⁶⁴ the cough suppressant noscapine,⁶⁵ the analgesics morphine and codeine,⁴⁴ and the anti-atherosclerosis agent magnoflorine,⁶⁶ have all been derived from (S)-norcoclaurine. Its biosynthesis starts with the decarboxylation and hydroxylation of L-tyrosine to 4-hydroxyphenylacetaldehyde (4-HPAA) and L-3,4-dihydroxyphenylalanine (L-DOPA), respectively, under the catalyzation of tyrosine decarboxylase and hydroxylase. Subsequently, (S)-norcoclaurine is formed through the condensation of 4-HPAA and dopamine (converted from L-DOPA by L-DOPA decarboxylase) under the catalyzation of norcoclaurine synthase^{67–69} (Figure 3). A *de novo* (S)-norcoclaurine biosynthetic pathway consists of an upstream module that includes four endogenous proteins and a downstream heterologous module that includes three heterologous proteins. Overexpression of the four yeast proteins leads to an increased accumulation of L-tyrosine: DAHP synthase (Aro4_{K220L}) and chorismate mutase (Aro7_{G139S}) are two mutants insensitive to aromatic amino acid feedback inhibition;^{54,70} transketolase (Tkt1) and the pentafunctional aromatic enzyme (Aro1) are usually overexpressed to increase the availability of the precursor of E4P and thus the flux toward the shikimate pathway.^{8,71} The three heterologous proteins are the tyrosine hydroxylase variant (TyrH_{W13L, F309L}) from *Berberis vulgaris* (reported to have a low L-DOPA oxidase activity to avoid the draining of L-DOPA⁶⁹), L-DOPA decarboxylase (DODC) from *Pseudomonas putida*,⁷² and norcoclaurine synthase (NCS) from *Papaver somniferum*.⁶⁹

Despite the capabilities of nonconventional hosts to outperform model hosts in certain aspects, our knowledge of phenotype-genotype relationships in nonconventional mi-

crobes is limited. Prior to introducing the two modules of the (S)-norcoclaurine pathway, we used the HTn technology to create a knockout library for screening of host strains carrying unknown knockouts beneficial to (S)-norcoclaurine biosynthesis. The screening relied on the conversion of L-DOPA to the yellow pigment betaxanthin via L-DOPA dioxygenase (DOD), which was developed for *S. cerevisiae* (Figure 3).^{69,73} A *S. stipitis* version of the sensor plasmid (named pUra-TyrH*-DOD) was constructed and transformed into the SS *ura*⁻ strain. Uniformly yellow colonies were directly observed on the SC-URA plate.

A plasmid pHTn-Leu expressing Hermes transposase and the TIR-flanked *LEU2* was transformed into the uracil and leucine double auxotrophic strain *S. stipitis* FLP-UC7 *leu2* Δ (hereafter the strain is labeled as SS *ura*⁻ *leu*⁻), which was created by introducing 7 bp *indel* mutations to *LEU2* of the strain SS *ura*⁻ using CRISPR technology.^{48,49} Transformants on SC-LEU agar plates were collected as a pool and diluted onto SC-LEU agar plates supplemented with 5-FOA. Colonies grown on the SC-LEU+5-FOA plates were pooled again, rediluted, and spread onto new SC-LEU+5-FOA agar plates. The 5-FOA-resistant colonies were collected to generate a transposon-mediated mutagenesis library G0. Subsequently, the sensor plasmid pUra-TyrH*-DOD was transformed into the library G0. Colonies with higher yellow intensity were inoculated into SC-LEU-URA medium, and betaxanthin fluorescence was measured by a microplate reader after 48 h of cultivation (Figure 7a). A 10-fold variation in fluorescence was observed among the selected colonies, and the majority of them produced more betaxanthin than the SS *ura*⁻ strain transformed with the sensor plasmid as a control. The top two mutants, namely, BX-1 and BX-2, were selected for further analysis. To avoid the selection of false positives (e.g., mutations occurring in TyrH* or DOD that happened to lead to higher betaxanthin fluorescence), the sensor plasmid was removed by applying 5-FOA counter-selection, and the resulting cells were retransformed with the freshly prepared sensor plasmid. Ten colonies were transferred from the same transformation plate to liquid medium, and betaxanthin fluorescence was measured again (Figure 7b). Although variation among the ten colonies was relatively large, the average production levels of BX-1 and BX-2 both surpassed that of the control strain. We speculated that the large variations were a result of using the plasmid-born sensor considering that the *S. stipitis* plasmid is not as stable as the ones used in *S. cerevisiae*. We elected not to proceed with site-

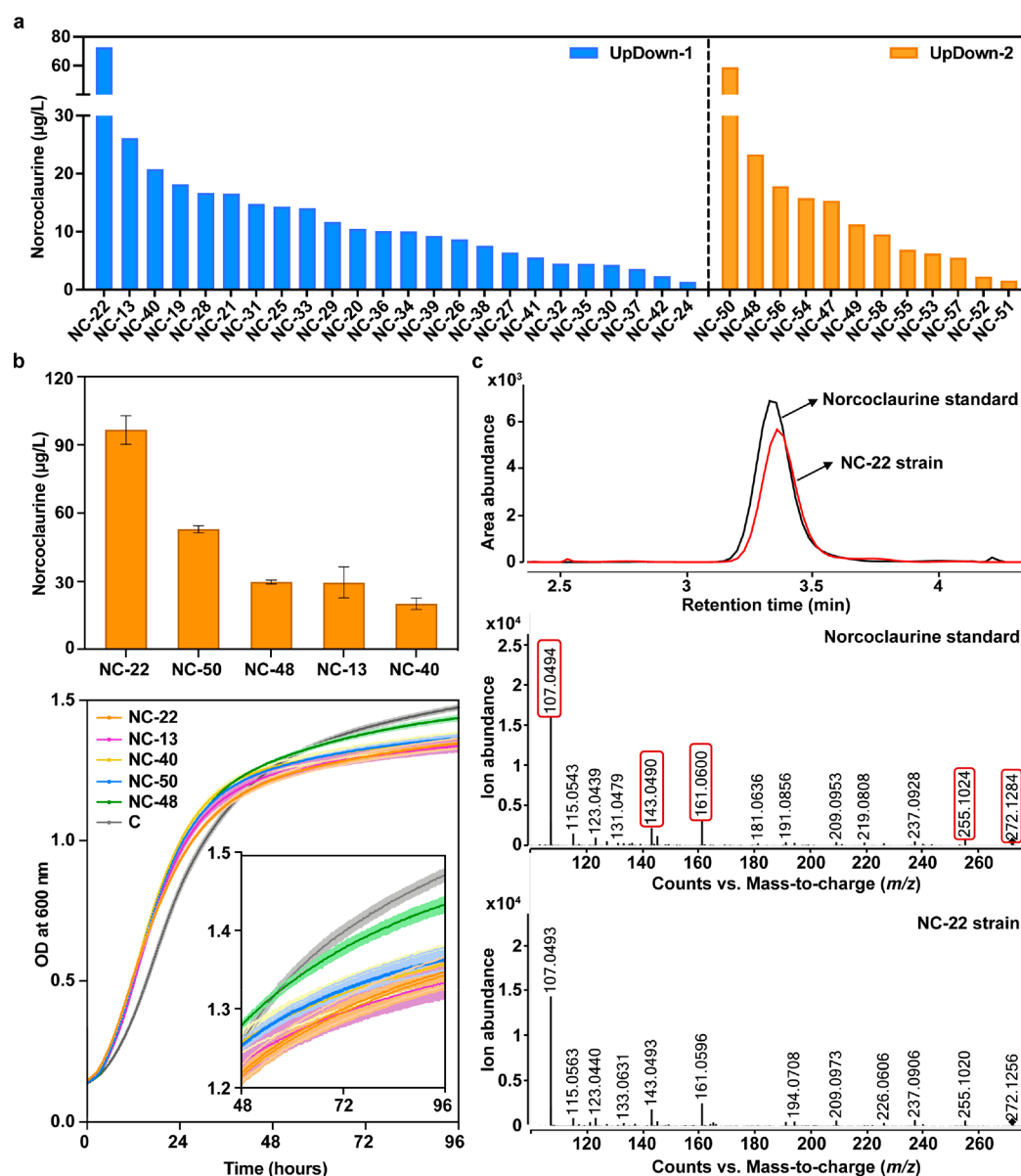


Figure 8. Iterative integration of the (S)-norcoclaurine biosynthetic pathway using a combination of HTn and Cre-loxP recombination. (a) The production of (S)-norcoclaurine by the variants created by transpositional integration of the upstream and downstream modules into the strains Tyr-1 and Tyr-2. (b) The (S)-norcoclaurine production levels and the growth curves of the top five high-titer variants screened from the UpDown-1 and UpDown-2 libraries. “C” represents the negative control, which is the SS *leu⁻ ura⁻* strain containing an empty plasmid pSSLeuUra. Error bars indicate mean $\pm$ s.d. of three biological replicates. (c) LC-MS/MS analysis of (S)-norcoclaurine in the supernatant of strain NC-22 after 120 h. Fermentation was performed in 2X selective synthetic medium supplemented with 6% glucose and 0.5% L-ascorbic acid. Red boxes indicate the five signature $[M+H]^+$ ions of the (S)-norcoclaurine standard ($[m/z]$ 722.1, 255.1, 161.0, 143.0, and 107.0).

specific integration of TyrH* and DOD into the genomes of BX-1 and BX-2, because the easy removal of the plasmid-born sensor made the strains ready for execution of the HTn-mediated (S)-norcoclaurine pathway integration.

Combining Hermes Transposon and Cre-loxP for Iterative Integration of the (S)-Norcoclaurine De Novo Biosynthetic Pathway. With the removal of the sensor plasmid, the BX-1 and BX-2 mutants were renamed strains Tyr-1 and Tyr-2 (considering that the product was not betaxanthin anymore). To randomly integrate the (S)-norcoclaurine pathway, we designed two systems: Design I combined all seven genes into one 22.5-kb piece flanked by the *GFP* and *URA3* expression cassettes and the TIRs; Design II split the 16.6 kb upstream module (*GFP*, *TKL*, *aro4_{K220L}*, *ARO1*, *aro7_{G139S}*, and *URA3*)

from the 8.8 kb downstream module (*BFP*, *TyrH_{W13L, F309L}*, *DODC*, *NCS*, and *URA3*). We conjectured that Design I would probably result in lower transposition efficiency, but the integration could be done in a single step. Design II might not only generate a larger expression library to be screened; in addition, placing the two modules into different transcriptional environments would also increase our ability to fine-tune the entire pathway. Unsurprisingly, the nonreplicable circular module cHTn-GNCU (30.9 kb including the backbone) containing the entire (S)-norcoclaurine pathway only led to 2–3 cfu/ μ g DNA, which was possibly caused by the reduced transposition and transformation efficiencies of DNA larger than 20 kb.

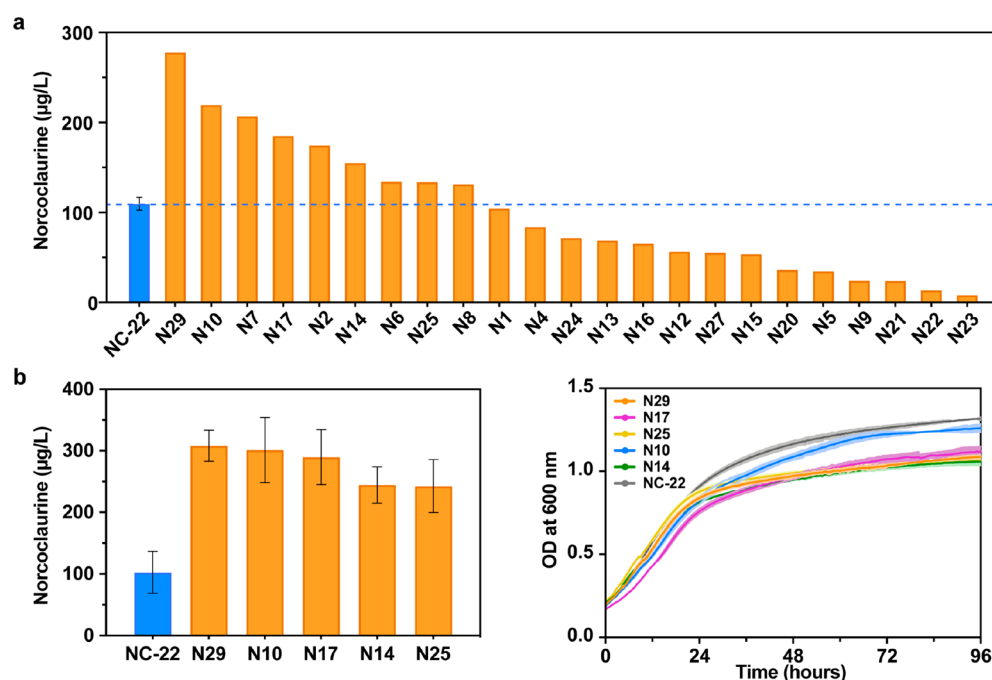


Figure 9. The third round of HTn-mediated integration targeting the rate-limiting norcoclaurine synthase NCS led to the identification of significantly improved variants. (a) The (*S*)-norcoclaurine production by the variants created by transpositional integration of cBNU-loxP in the NC-22 strain. (b) Verification of the (*S*)-norcoclaurine production levels and the growths of the top five high-titer variants screened from the NC-22-NCS library. Error bars indicate mean $\pm$ s.d. of three biological replicates.

As the strains Tyr-1 and Tyr-2 only contained one auxotrophic selection marker *URA3* to be used, to implement Design II, we used a Cre-loxP plasmid, which has been previously utilized for marker rescue in *S. stipitis*⁷⁴ (this marker rescue step can be skipped in future studies if the host is a multi-auxotrophic strain). Two circular molecules, namely, cHTn-GSA4U-loxP and cHTn-BTDNU-loxP, were constructed, each with the *URA3* selection marker flanked by two *loxP* sites. The “G” and “B” in cHTn-GSA4U-loxP and cHTn-BTDNU-loxP refer to GFP and BFP, respectively. The molecule cHTn-GSA4U-loxP was transformed into the strains Tyr-1 and Tyr-2 in parallel, and then the colonies surviving on the SC-LEU-URA plates were pooled (Figure 1c). The integrated *URA3* gene was then removed by transforming the plasmid pJMLS45 expressing Cre recombinase, and the resulting libraries were named Up-1 and Up-2. Subsequently, cHTn-BTDNU-loxP containing the downstream module was transformed into these two libraries to generate libraries UpDown-1 and UpDown-2, and then the cells underwent selection on SC-LEU-URA plates. BFP and GFP were used to screen for the strains in which both modules were integrated. This strategy yielded 73% (24/33) and 46% (12/26) GFP⁺BFP⁺ colonies in the two libraries, and all of these clones were confirmed to include both modules and produce the product (Figure S3). The production of (*S*)-norcoclaurine was found to vary from 1.4 to 73 μg/L in the UpDown-1 library and from 1.6 to 59 μg/L in the UpDown-2 library in the initial screening (Figure 8a). The five highest-producing strains were selected to retest their titers and growth rates (Figure 8b). The results indicated that the growth of the selected strains was not significantly disturbed by the integration of the pathway; the best producing strain, NC-22, produced 97 ± 6 μg/L (*S*)-norcoclaurine using glucose as the substrate after 120 h of fermentation (Figure 8c).

A significant accumulation of dopamine was observed during the product detection, which was consistent with the previous study that reported significant decrease in titers from dopamine (~25 mg/L) to (*S*)-norcoclaurine (~100 μg/L) in *S. cerevisiae*.⁶⁹ There were also studies that selected NCS homologues^{12,69,75} and increased NCS copy number¹² to improve the conversion efficiency of the NCS-catalyzed reaction. In these studies, the best producer could produce 50–60 μg/L of (*S*)-norcoclaurine *de novo* with a single-copy integration of the NCS gene from *Sanguinaria canadensis* or *Nandina demestica* using 4% sucrose as a substrate; expressing additional three copies of NCS further increased the titer of (*S*)-norcoclaurine to ~130 μg/L in *S. cerevisiae*. To confirm that NCS is the bottleneck in *S. stipitis*, we proceeded with another round of transpositional integration of NCS based on the top producer NC-22 strain after removing *URA3* using Cre recombinase. The production of (*S*)-norcoclaurine by the resulting NC-22-NCS variants varied from 8.4 to 278 μg/L in the initial round of screening (Figure 9a), and 9 of 29 transformants yielded higher titers compared to the starting strain NC-22. The top-eight producers were then subjected to the second round of verification with three biological replicates. Five of them robustly produced (*S*)-norcoclaurine at titers at least twofold of that produced by NC-22 (Figure 9b). The highest production reached 308 ± 25 μg/L after 120 h of fermentation, which is threefold of the titer previously achieved in *S. cerevisiae* (105 μg/L using glucose as the substrate).⁶⁹

We further analyzed the copy number of integration achieved by the HTn strategy using quantitative polymerase chain reaction (qPCR). The *ARO1* and *NCS* genes were selected for the quantification of the upstream and downstream modules, respectively. The results showed that all five strains identified from the UpDown libraries contained a single-copy *ARO1* and one- to two-copy *NCS*, with the top two high-

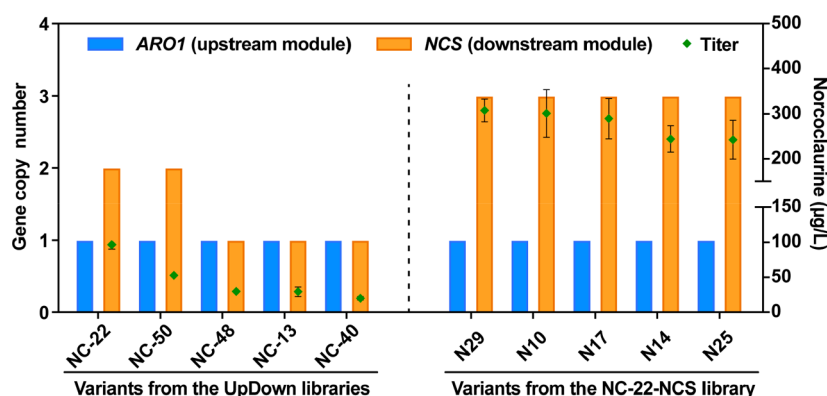


Figure 10. Copy-number analysis of *ARO1* (upstream module) and *NCS* (downstream module) in 10 variants that produced (S)-norcoclaurine at high titers. The five variants on the left were identified from the UpDown-1 and UpDown-2 libraries, and the other five variants on the right were identified from the NC-22-NCS library.

Table 1. Summary of Genomic Integration Methods for Expressing Multi-gene Pathways in Microbial Hosts^a

principles	strategies	description of genomic integration	references
recombinase-mediated	RAGE mediated by Cre recombinase	59-kb gene cluster containing alginate utilization and ethanol production pathways in <i>E. coli</i>	13
	CRAGE by transposon and Cre recombinase	48-kb putative biosynthetic gene cluster in <i>E. coli</i>	14
	λ -red recombinase-based system	50-kb fragment amplified from <i>Bacillus subtilis</i> genomic DNA, divided into seven fragments (each 6–7 kb), and sequentially integrated into <i>E. coli</i> genome in seven rounds	15
	MSGE mediated by single integrase and multiple <i>attB</i> sites	5-copy 67-kb pristinamycin II biosynthetic gene cluster in <i>Streptomyces pristinaespiralis</i> (two rounds)	16
	aMSGE mediated by five orthogonal site-specific recombination systems	4-copy 72-kb 5-oxomilbemycin biosynthetic gene cluster in <i>Streptomyces hygroscopicus</i> (two rounds)	17
CRISPR/Cas9	Di-CRISPR	18-copy 24 -kb pathway containing xylose utilization and 2,3-butanediol production pathways in <i>S. cerevisiae</i>	19
	CRISPR-Cas	24 -kb muconic acid pathway, divided into six fragments, and integrated into three loci in <i>S. cerevisiae</i>	18
	CRISPR-assisted	10-kb isobutanol production pathway in <i>E. coli</i>	20
	HR	14.5-kb fragment in <i>KU70</i> -deficient <i>Yarrowia lipolytica</i>	76
inherent genetic mechanism	NHEJ	12.5-kb fragment containing pyruvic carboxylase and fumaric reductase expression cassettes in <i>Y. lipolytica</i>	27
	MycoMar transposon	58-kb epothilone gene cluster in <i>Myxococcus xanthus</i>	56
transposon-mediated	MycoMar transposon	56-kb epothilone gene cluster in <i>Burkholderiales</i> strain	58
	MycoMar transposon	21-kb coronafacic acid gene cluster in <i>Pseudomonas putida</i>	59
	TREX and yTREX mediated by Tn5 transposon	22-kb prodigiosin pathway in <i>P. putida</i>	57,77
	Hermes transposon	14.9-kb shikimate pathway and 22.5-kb (S)-norcoclaurine pathway in <i>S. stipitis</i>	this study

^aNotes: RAGE, recombinase-assisted genome engineering; CRAGE, chassis-independent recombinase-assisted genome engineering; MSGE, multiplexed site-specific genome engineering; aMSGE, advanced multiplex site-specific genome engineering; Di-CRISPR, delta integration CRISPR-Cas; HR, homologous recombination; NHEJ, nonhomologous end joining; TREX, transfer and expression of biosynthetic pathways; yTREX, yeast recombinational cloning-enabled pathway transfer and expression tool.

production variants (*i.e.*, NC-22 and NC-50) containing two-copy NCS (Figure 10). It is interesting to point out that, even with the same copy number, the (S)-norcoclaurine production levels in NC-22 and NC-50 differed by 45%, supporting the strong impact of a transcriptional environment on the pathway expression. The other five variants derived from the strain NC-22 and yielding higher titers were verified to contain three-copy NCS, indicating that NCS is indeed the rate-limiting step for the production of (S)-norcoclaurine. Our results demonstrated the simplicity and the efficiency of the HTn-mediated genome integration strategy for identifying high-performance strains.

DISCUSSION

Despite the rapidly increasing interest in expanding the current collection of microbial production hosts, significant technology hurdles still lay ahead. Tracing the path of *E. coli* and *S. cerevisiae* research development, researchers recognize clearly that stable expression platforms and precise genome-editing tools are needed as the foundational technologies for any new species. *S. stipitis* was previously demonstrated as a potentially better-suited host than *S. cerevisiae* for producing derivatives of the shikimate pathway, which was attributed to the highly active pentose phosphate pathway in *S. stipitis*.^{8,45} We previously developed a centromere/autonomous replicative sequence (CEN/ARS)-based low-copy plasmid^{49,55} and

CRISPR editing tool⁴⁸ for this species. However, when a complex pathway was expressed using this CEN/ARS episomal plasmid, the productivity from a single genotype background was not satisfactorily stable (varying from 1.2 to 4.0 g/L, as illustrated by the shikimate example, Figure 4c). Note that the stability of this CEN/ARS episomal plasmid had already been significantly improved over the previous ARS plasmid that was missing a CEN sequence. If the genome is the preferred pathway carrier, when we engineer a nonconventional species, the genome integration locus of the pathway is an important consideration, especially considering the paucity of the knowledge regarding integration hotspots and genotype-phenotype correlations in a new species.

To tackle this issue, we evaluated two distinct mechanisms to achieve random pathway integration and expression library creation in yeasts. Table 1 summarizes the commonly used strategies for large pathway integration in both bacteria and yeasts. CRISPR technology, heterologous recombinases, and HR are the three main mechanisms used to achieve site-specific integration. In addition, to make full use of different transcriptional environments present in a given genome, the NHEJ strategy was implemented to integrate the β -carotene pathway into the genome of *Y. lipolytica*, with high-production strains identified based on the orange intensity of the product.²⁷ Despite the ease of exploiting this inherent mechanism to create an expression library, in the absence of a directly visible phenotype, the high false-positive rate significantly influences the efficiency of identifying interesting variants. In our study, integration of the G-U fragment via NHEJ resulted in merely 25% of the plated colonies that presented a GFP⁺ phenotype. On the one hand, this is because this dominant homology-independent mechanism does not guarantee the incorporation of anything except a functional *URA3* expression cassette into a chromosomal DSB. On the other hand, if we view the receiver DNA (*i.e.*, the genome) and the donor DNA (*e.g.*, the G-U fragment) with reversed roles, a URA⁺ phenotype could also be rendered by the G-U fragment (now as a receiver) that grabs an ARS-like element from the genome (now as a donor) and circularizes with it. This secondary group of false-positive clones represented as much as 70% of all the GFP⁺URA⁺ clones. The hypothesis of circularization was made on the basis of GFP expression profiles, which largely mimicked the expression profile of the strain containing an unstable ARS episomal plasmid (Figure 2b). Our previous study demonstrated that a plasmid containing a replicating signal but missing a segregation signal (given by CEN) could not be stably segregated to an offspring cell during cell division, which causes a heterogeneous plasmid distribution and a large variation in GFP expression.⁴⁹

In addition to the NHEJ-mediated integration in *Y. lipolytica*, researchers have also developed a CRISPR-derived strategy (named Di-CRISPR) to target the redundant delta sequences scattered in the *S. cerevisiae* genome to integrate a 24-kb combined xylose utilization pathway and a 2,3-butanediol pathway with multiple copies.¹⁹ Unfortunately, such a strategy could not be easily applied to the vast majority of the nonconventional yeasts, as these species do not have sufficient HR machinery to simultaneously seal the numerous chromosomal incisions introduced by Cas9.

For reasons unknown, random integration of pathways in yeasts has not benefited from the success of transpositional delivery of large natural product gene clusters to heterologous bacteria, such as *Myxococcus xanthus*, *Burkholderiales* sp., and

Pseudomonas putida.^{56–59} The concepts from these studies, taken with the functional genomics studies in yeasts,^{34–36} made the creation of a transposon-based pathway integration method straightforward, requiring only the modification of enclosing the target pathway and selection marker by TIRs (Figure 5). Unexpectedly, this attempt resulted in a high false-positive rate in the initial screen as well. After plasmid curing and second-step verification, we determined that 79% of the transformants appearing on the first selection plate had escaped transposition. To streamline the method and reduce false positives, we modified the initial design to create a nonreplicable circular DNA molecule that used only one selection marker, skipped the plasmid elimination step, and significantly shortened the library creation step to just 5 d. When GFP⁺URA⁺ was used as the screening filter, 97% of the GFP⁺URA⁺ clones were found to carry the enclosed shikimate pathway; the equivalent percentage was only ~55% (*i.e.*, 60% $\times$ 92%) when NHEJ-mediated integration was used. We also noticed that the number of colonies dramatically decreased as the size of the transposable pathway increased. On the one hand, we had to split the 22.5 kb (*S*)-norocloaurine pathway into two modules and apply a two-step transposition protocol to generate enough production variants. On the other hand, the splitting allowed for independent fine-tuning of each module; the iterative integration strategy was also useful for increasing the copy number of the gene encoding the rate-limiting step to further enhance the productivity.

CONCLUSION

We broadened the usage of DNA transposition beyond the classical application of building genome-scale knockout libraries in yeasts by demonstrating its capacity for rapid creation of pathway expression libraries. This strategy efficiently leverages the readily available fine-tuning impact provided by the diverse transcriptional environment surrounding each random integration locus. Using one endogenous product and one heterologous product as examples, we demonstrated that the nonreplicable circular platform not only offered the highest integration efficiency but also the most streamlined procedure. This new exploration of HTn therefore adds a highly versatile tool to accelerate the development of novel species as microbial cell factories for producing value-added chemicals.

MATERIALS AND METHODS

Strains and Media. The *S. stipitis* FLP-UC7 strain (*ura3*–3, NRRL Y-21448)⁷⁸ was gifted from Dr. Thomas W. Jeffries, professor emeritus from the University of Wisconsin-Madison. *S. stipitis* FLP-UC7 *leu2* Δ was derived from *S. stipitis* FLP-UC7 by introducing *indel* mutations (7 bp deletion) to *LEU2* using a previously reported genome editing protocol.⁴⁹ *S. cerevisiae* YSG50 (*MAT α* , *ade2*–1, *ade3* Δ 22, *ura3*–1, *his3*–11, 15, *trp1*–1, *leu2*–3, 112, *can1*–100) was used as the host for plasmid assembly in this study. Yeast strains were cultured at 30 °C in YPAD, YPAX, YPAD plus zeocin, or synthetic complete (SC) dropout medium. YPAD or YPAX medium contains 1% yeast extract, 2% peptone, 0.01% adenine hemisulfate, and 2% dextrose or xylose. YPAD plus zeocin medium was prepared based on the YPAD medium supplemented with 100 μ g/mL zeocin, and the pH was adjusted to 7.5 using 1 M NaOH. SC dropout medium lacking uracil, leucine, or tryptophan and consisting of 0.17% yeast nitrogen base without amino acids

and ammonium sulfate, 0.5% ammonium sulfate, 0.083% synthetic complete drop out mix, and 2% dextrose was used to select transformants containing the corresponding auxotrophic gene. SC plus 5-FOA medium was prepared based on the SC medium recipe and supplemented with 1 mg/mL 5-FOA,⁷⁹ which was used for the screening of *ura*[−] cells. The *E. coli* strain BW25141 (*lacI*^q *rrnB*_{T14} Δ *lacZ*_{WJ16} Δ *phoBR580* *hsdR514* Δ *araBAD*_{AH33} Δ *rhaBAD*_{LD78} *galU95* *endA*_{BT333} *uidA*(Δ *MluI*)::*pir*⁺ *recA1*) was used for plasmid enrichment. The *E. coli* strain DH5 α (New England BioLabs) was used as the host for Gibson Assembly. The *E. coli* strains were grown at 37 °C in Luria broth (LB) medium with 100 μ g/mL ampicillin added for plasmid selection. Strains used in this study are summarized in Table 2.

Table 2. Strains Used in This Study

strains	genotype or relevant features	refs
<i>E. coli</i> BW25141	<i>lacI</i> ^q <i>rrnB</i> _{T14} Δ <i>lacZ</i> _{WJ16} Δ <i>phoBR580</i> <i>hsdR514</i> Δ <i>araBAD</i> _{AH33} Δ <i>rhaBAD</i> _{LD78} <i>galU95</i> <i>endA</i> _{BT333} <i>uidA</i> (Δ <i>MluI</i>):: <i>pir</i> ⁺ <i>recA1</i>	43
<i>E. coli</i> DH5 α	<i>fhuA2</i> :: <i>IS2</i> Δ (<i>mmuP-mhpD</i>)169 Δ <i>phoA8</i> <i>gln</i> X44 Φ 80d[Δ <i>lacZ</i> 58(<i>M15</i>)] <i>rfdD1</i> <i>gyrA96</i> <i>luxS11</i> <i>recA1</i> <i>endA1</i> <i>rph</i> ^{WT} <i>thiE1</i> <i>hsdR17</i>	NEB strain
<i>S. cerevisiae</i> YSG50	<i>MATα</i> , <i>ade2</i> −1, <i>ade3</i> Δ 22, <i>ura3</i> −1, <i>his3</i> −11,15, <i>trp1</i> −1, <i>leu2</i> −3,112, <i>can1</i> −100	43
SS <i>ura</i> [−]	<i>S. stipitis</i> FLP-UC7 <i>ura3</i> −3, NRRL Y-21448	78
SS <i>ura</i> [−] <i>leu</i> [−]	<i>S. stipitis</i> FLP-UC7, <i>leu2</i> Δ	this study
SS <i>ura</i> [−] <i>ku70</i> Δ <i>ku80</i> Δ	<i>S. stipitis</i> FLP-UC7, <i>ku70</i> Δ <i>ku80</i> Δ	48
SS <i>ura</i> [−] -GFP	SS <i>ura</i> [−] , <i>xyl1</i> :: <i>egfp</i>	this study
pSA	SS <i>ura</i> [−] harboring pSA7.3 plasmid	8
SA1	SS <i>ura</i> [−] carrying NHEJ-mediated random integration of the G-SA-U fragment	this study
SS <i>ura</i> [−] -pDOD	SS <i>ura</i> [−] harboring pUra-DOD plasmid	this study
SS <i>ura</i> [−] -pTD	SS <i>ura</i> [−] harboring pUra-TyrH*-DOD plasmid	this study
SS <i>ura</i> [−] -pLU	SS <i>ura</i> [−] harboring pSsLeuUra plasmid	this study
SS <i>ura</i> [−] <i>leu</i> [−] -pLU	SS <i>ura</i> [−] <i>leu</i> [−] harboring pSsLeuUra plasmid	this study
BX-1	Library harboring pUra-TyrH*-DOD plasmid and HTn-mediated random insertion of <i>LEU2</i>	this study
BX-2	Library harboring pUra-TyrH*-DOD plasmid and HTn-mediated random insertion of <i>LEU2</i>	this study
Tyr-1	Library harboring HTn-mediated random insertion of <i>LEU2</i>	this study
Tyr-2	Library harboring HTn-mediated random insertion of <i>LEU2</i>	this study
NC-22	<i>S. stipitis</i> variant carrying the integrated one-copy GSA4 and two copy BTDNu-loxP	this study
N-29, N10, N17, N14, and N25	<i>S. stipitis</i> variants carrying the integrated one-copy GSA4, two-copy BTDNu, and another copy of BNU-loxP	this study

Plasmid Construction and Transformation. Recombinant plasmids were constructed using the DNA assembler approach^{60–62} or Gibson Assembly (New England BioLabs). The genomic DNA of *S. stipitis* FLP-UC7 was extracted using the Wizard genomic DNA purification kit (Promega) and used as polymerase chain reaction (PCR) templates to amplify the DNA fragments of *URA3*, *LEU2*, *ARO7*, promoters, and terminators. The linear *GFP-URA3* fragment was obtained via amplification from the plasmid pRS-GU. pGSAU was constructed via assembling pRS414 linear backbone, the *GFP*

and *URA3* expression cassettes, and the fragment carrying the shikimate pathway obtained from the pSA7.3 plasmid previously reported⁸ using the DNA assembler approach. The linear G-SA-U cassette was obtained by digesting the pGSAU plasmid using *NotI* and *XhoI* (Thermo Fisher Scientific). Hermes transposase was codon-optimized by GenScript according to the codon preference of *S. stipitis* and cloned under the control of the *ENO1* promoter. Hermes expression cassette and the transposed *LEU2* expression cassette flanked by the Hermes TIRs³⁵ were assembled into a *URA3*-marked ARS/CEN5–750 bp plasmid backbone to generate the donor plasmid pHTn-Leu for transposon-mediated mutagenesis. Two transposition-negative control plasmids were derived from the pHTn-Leu plasmid, including pHTn-NT containing TIR-flanked *LEU2* but lacking transposase and pHTn-NR containing the transposase but lacking the TIR on the right side of the *LEU2* cassette. cHTn-GSAU was constructed by assembling the Hermes transposase cassette, the G-SA-U fragment flanked by the TIRs, and the pRS414 linear backbone using the DNA assembler approach. Another two negative controls, namely, cHTn-SA-NT (no transposase) and cHTn-SA-NR (no TIR on the right side), were yielded based on cHTn-GSAU. “c” stands for “circular”. Because of the lack of replication element specific to *S. stipitis*, these circular DNA molecules cannot be propagated in *S. stipitis*. However, they contain the replication/segregation elements and the selection markers for both *S. cerevisiae* and *E. coli* in the backbones and, therefore, can be propagated as plasmids in these two hosts.

To construct the (S)-norcoclaurine biosynthetic pathway, *TKT1*, *aro4*_{K220L}, *ARO1*, and *aro7*_{G139S} were amplified from *S. stipitis* FLP-UC7 genomic DNA, and the genes *TyrH*_{W13L}, *F309L*, *DODC*, and *NCS* were codon-optimized and synthesized (GenScript). Subsequently, all seven expression cassettes were assembled into the digested cHTn-GSAU, replacing the enclosed shikimate pathway by the (S)-norcoclaurine pathway to form the cHTn-GNCU plasmid. In parallel, the molecules cHTn-GSA4U-loxP and cHTn-BTDNU-loxP were constructed by replacing the GSAU transposed region of cHTn-GSAU with GSA4U-loxP and BTDNu-loxP, respectively. Yeast transformation was performed using an electroporation method,⁶² and a voltage of 0.75 kV/mm was used for *S. cerevisiae*, whereas a higher voltage of 1.5, 1.8, or 2.5 kV/mm was used for *S. stipitis*. All the plasmids are summarized in Table 3, and the maps for the main constructs are summarized in Figure S4.

Screening for the Strains with Integrated GFP or BFP.

Colonies transformed with a linear fragment or a nonreplicable circular DNA molecule containing the *GFP* gene were checked using a DR46B transilluminator (Clare Chemical Research). The visibly green colonies were randomly selected and cultured in 3 mL of SC-URA or SC-LEU-URA medium at 30 °C with a constant shaking at 250 rpm for 36–48 h. Saturated cultures were diluted into 3 mL of fresh medium with an initial OD₆₀₀ of 0.2–0.3 and then were continuously grown with shaking for 48 h. Cells sampled at 24 or 40 h were diluted by 10 times into 10 mM phosphate-buffered solution (PBS) (pH 7.4) and analyzed for GFP or BFP fluorescence intensity at 488 or 405 nm using a FACSCanto flow cytometer (BD Biosciences). The fluorescence intensity distribution of GFP- or BFP-expressing cells was calculated using BD FACSCanto Clinical Software.

Table 3. Plasmids Used in This Study

plasmids	features	references
pRS-GU	GFP-URA3 cassettes in pRS414 backbone	this study
pSA7.3	TKT1- <i>aro4</i> _{K220L} - <i>aro1</i> _{D900A} cassettes, CEN/ARS, and URA3 marker for <i>S. stipitis</i> in pRS414 backbone	8
pGSAU	GFP-SA-URA3 cassettes in pRS414 backbone	this study
pJY3919	Hermes transposon plasmid designed for <i>Yarrowia lipolytica</i>	35
pHTn-Leu	ENO1p-Hermes-TEF1t cassette, TIRs flanking LEU2, CEN/ARS, and URA3 marker for <i>S. stipitis</i> in pRS414 backbone	this study
pHTn-NT	TIRs flanking LEU2, CEN/ARS, and URA3 marker for <i>S. stipitis</i> in pRS414 backbone	this study
pHTn-NR	ENO1p-Hermes-TEF1t cassette, one-sided TIR flanking LEU2, CEN/ARS, and URA3 marker for <i>S. stipitis</i> in pRS414 backbone	this study
chTn-GSAU	ENO1p-Hermes-TEF1t cassette and TIRs flanking GFP-SA-URA3 cassette for <i>S. stipitis</i> in pRS414 backbone	this study
chTn-SA-NT	TIRs flanking GFP-SA-URA3 cassette for <i>S. stipitis</i> in pRS414 backbone	this study
chTn-SA-NR	ENO1p-Hermes-TEF1t cassette and one-sided TIR flanking GFP-SA-URA3 cassette for <i>S. stipitis</i> in pRS414 backbone	this study
chTn-GNCU	ENO1p-Hermes-TEF1t cassette and TIRs flanking GFP-NC-URA3 cassettes in pRS414 backbone (NC contains the expression cassettes for TKT1, <i>aro4</i> _{K220L} , ARO1, <i>aro7</i> _{G139S} , <i>TyrH</i> _{W13L, F309L} , DODC, and NCS)	this study
pUra-TyrH*-DOD	OLE1p-TyrH _{W13L, F309L} -GLN1t, sGLN1p-DOD-sTDH2t, CEN/ARS, and URA3 marker for <i>S. stipitis</i> in pRS414 backbone	this study
pUra-DOD	sGLN1p-DOD-sTDH2t cassette, CEN/ARS, and URA3 marker for <i>S. stipitis</i> in pRS414 backbone	this study
chTn-GSA4U-loxP	ENO1p-Hermes-TEF1t cassette and TIRs flanking GFP-SA4-URA3-loxP cassettes for <i>S. stipitis</i> in pRS414 backbone (SA4 contains the expression cassettes for TKT1, <i>aro4</i> _{K220L} , ARO1, and <i>aro7</i> _{G139S} ; URA3 is flanked by loxP sequences)	this study
chTn-BTDNU-loxP	ENO1p-Hermes-TEF1t cassette and TIRs flanking BFP-TDN-URA3-loxP cassettes for <i>S. stipitis</i> in pRS414 backbone (TDN contains the expression cassettes for <i>TyrH</i> _{W13L, F309L} , DODC, and NCS; URA3 is flanked by loxP sequences)	this study
chTn-BNU-loxP	ENO1p-Hermes-TEF1t cassette and TIRs flanking BFP-NCS-URA3-loxP cassettes for <i>S. stipitis</i> in pRS414 backbone	this study
pJML545	XYL1p-Cre-XYL1t, TEF1p-ble-XYL2t, and ARS for <i>S. stipitis</i>	74
pSsLeuUra	LEU2, URA3, and CEN/ARS for <i>S. stipitis</i> in pRS414 backbone	this study

To analyze the correlation between shikimate production and GFP expression intensity among the *S. stipitis* variants with randomly integrated G-SA-U fragment, the total GFP fluorescence intensity was analyzed using a Synergy HTX multimode reader (BioTek) at an excitation of 485 nm and an emission of 516 nm with a gain of 85. The sample (100 μ L) was collected after 120 h of cultivation and placed into a 96-well black polystyrene plate (Fisher Scientific) for measuring fluorescence intensity, which was then normalized to the cell OD₆₀₀.

Shikimate Production in *S. stipitis*. The *S. stipitis* variants confirmed to carry the integrated shikimate pathway were inoculated into 3 mL of SC-URA medium and grown with shaking for 36–40 h. Cultures were diluted into 55 mL rimless culture tubes carrying 6 mL of fresh SC-URA medium supplemented with 4% glucose and grown at 30 °C with shaking at 250 rpm for 120 h. Culture (200 μ L) was pelleted, and the supernatant was diluted 100-fold into water. Shikimate production was quantified using a previously reported spectrophotometric assay.⁸⁰ Briefly, 100 μ L of diluted supernatant was oxidized with 50 μ L of a solution consisting of 0.5% periodic acid and 0.5% sodium *m*-periodate, and this mixture was incubated at 37 °C for 45 min. Subsequently, the reaction was quenched with 100 μ L of a solution consisting of 1 M NaOH and 56 mM Na₂SO₃ (3:2 ratio, v/v), and the absorbance was measured at 382 nm using a Synergy HTX multimode reader. Shikimate was quantified on the basis of a calibration curve using authentic shikimate (Sigma-Aldrich) as a standard at concentrations ranging from 5 to 40 mg/L.

Hermes Transposition and Knockout Library Construction in *S. stipitis*. The uracil and leucine double auxotrophic strain *S. stipitis* FLP-UC7 *leu2* Δ (also containing *ura3*–3) was used for Hermes transposition. Plasmid pHTn-

Leu or the negative control plasmids pHTn-NT and pHTn-NR were transformed into this strain. After 5 d of incubation on the SC-LEU agar plates, transformants were collected as a pool, diluted 10⁴-fold, spread on the SC-LEU agar plates supplemented with 5-FOA, and incubated at 30 °C for another 4–5 d. Colonies grown on the SC-LEU+5-FOA plates were pooled, separately rediluted with the same order of magnitude, and spread onto the YPAD and SC-LEU+5-FOA agar plates to calculate the transposition efficiency. All the cells survived on YPAD, whereas only the ones that contained the integrated LEU2-marked transposon and lost the URA3-marked pHTn-Leu plasmid survived on SC-LEU+5-FOA medium. The transposition efficiency was calculated as the ratio of the colony number on SC-LEU+5-FOA to the colony number on YPAD (Figure S5). The 5-FOA-resistant colonies were collected to generate a random transposon-mediated mutagenesis library (G0), which was built on *S. stipitis* FLP-UC7-*leu2* Δ (the strain was named SS *ura*[–] *leu*[–] in the Results section).

Measurement of Betaxanthin Production. Single colonies were grown for 36–40 h in the SC-URA medium at 30 °C with shaking at 250 rpm. Saturated cultures were subsequently diluted into fresh SC-URA medium with an initial OD₆₀₀ of 0.2–0.3. After the cultures grew with shaking at 30 °C for 24 h, 200 μ L of culture was spun down, and the collected cells were resuspended in 10 mM PBS (pH 7.4). The cells were transferred to a 96-well black polystyrene plate and measured for fluorescence intensity using Synergy HTX multimode reader (excitation: 485/20 nm, emission: 516/20 nm, gain: 50).⁶⁹ The fold change over the background fluorescence was calculated by normalizing the signal to the average fluorescence of the control strain SS *ura*[–] transformed

with the plasmid pDOD (the sensor plasmid with *TyrH*^{*} removed) cultured in SC-URA medium.

Screening for High-performance Variants with a High Flux toward L-Tyr. L-Tyrosine sensor plasmid pUra-TyrH^{*}-DOD was transformed into the G0 library described above using a voltage of 1.5 kV/mm. Approximately 846 transformants appeared on the SC-LEU-URA plates after 5–7 d of cultivation. The plated colonies with various yellow intensities were visually compared. The ones with higher intensities were restreaked onto fresh SC-LEU-URA plates and grown at 30 °C for another 3 d. The most intensely yellow colonies, compared to the control strain SS *ura*[−] carrying pUra-TyrH^{*}-DOD streaked on the same plate, were picked and inoculated into 2 mL of SC-LEU-URA medium supplemented with 2% glucose. After they were shaken for 36–48 h, the saturated cultures were diluted into fresh medium with an initial OD₆₀₀ of 0.2–0.3, and the cultivation continued for another 24 or 48 h. The *de novo* production of betaxanthin was quantified by a microplate reader as described above.

To confirm the stability of high-producing variants BX-1 and BX-2 screened from the library, the pUra-TyrH^{*}-DOD plasmid containing the *URA3* marker gene was removed by culturing the cells in SC-LEU+5-FOA medium. After 40–48 h of growing, cultures were transferred to the fresh SC-LEU+5-FOA medium, and the pUra-TyrH^{*}-DOD plasmid was retransformed. Ten yellow colonies were randomly selected for both strains and plated onto fresh SC-LEU-URA plates. The betaxanthin production was analyzed as described above.

Construction of the (S)-Norcoclaurine Production Library by Iterative Genome Integration and Selection Marker Curation. After the removal of the sensor plasmid, the resulting two strains Tyr-1 and Tyr-2 were used for (S)-norcoclaurine biosynthetic pathway integration. cHTn-GSA4U-loxP containing the upstream module was first transformed to Tyr-1 and Tyr-2 in parallel using a voltage of 1.8 kV/mm, and the cells were plated on the SC-LEU-URA medium. After 5 d of cultivation at 30 °C, all the transformants were collected as a pool and diluted to an OD₆₀₀ of 0.2 in 250 mL baffled shake flasks containing 50 mL of SC-LEU-URA medium. pJML545 plasmid⁷⁴ containing a zeocin selection marker (encoded by *Sh ble*) and Cre recombinase under the control of the *XYL1* promoter was electroporated into the libraries. Cells were plated on YPAD plus zeocin medium and cultured for 3 d. All the surviving transformants were subsequently pooled and diluted into 5 mL of YPAX medium to an initial OD₆₀₀ of 0.2 to induce the expression of Cre recombinase. Cells were grown for 24 h to simultaneously remove the *URA3* marker and the pJML545 plasmid without the supplementation of zeocin. Lastly, cultures were diluted onto the SC-LEU+5-FOA plates, and the *ura*[−] colonies were obtained after 5–6 d of cultivation, resulting in the library Up-1 and Up-2.

To ensure that the *URA3* marker was completely removed from both libraries, cells were cultured in SC-LEU+5-FOA medium prior to the second transformation of the downstream module carried by cHTn-BTDNU-loxP. Voltages of 1.8 and 2.5 kV/mm were used for library creation to improve the transformation efficiency. After 5–6 d of cultivation, two libraries UpDown-1 and UpDown-2 were obtained for (S)-norcoclaurine biosynthesis. 2.5 kV/mm yielded more colonies on the plates than 1.8 kV/mm.

(S)-Norcoclaurine Production and Analysis. Colonies were inoculated into 3 mL of SC-URA or SC-LEU-URA

medium with 2% glucose and grown at 30 °C with shaking at 250 rpm for 24–36 h. Saturated cultures were diluted into 55 mL rimless culture tube containing 6 mL of fresh 2X selective SC medium supplemented with 6% glucose and 0.5% L-ascorbic acid (Sigma-Aldrich) to prevent (S)-norcoclaurine oxidation. The cells were grown at 30 °C with shaking at 250 rpm for 120 h. One milliliter of culture was pelleted, and the supernatant was filtered and stored at −80 °C for subsequent analysis by liquid chromatography-mass spectrometry (LC-MS/MS).

Ten microliters of each culture supernatant was analyzed on an Agilent 6540 UHD Accurate-Mass Q-TOF LC/MS with a ZORBAX Eclipse Plus C18 reversed-phase column (2.1 × 100 mm, 1.8 μm, Agilent Technologies) at 40 °C using a 0.4 mL/min flow rate for separation and detection. Solvent A was water added with 0.2% (v/v) formic acid, and solvent B was acetonitrile added with 0.2% (v/v) formic acid. The samples were eluted with the following protocol: 0% B at 0–3 min, to 100% B at 10 min, maintaining 100% B at 11 min, and dropping to 0% B at 13 min, followed by 3 min of post-acquisition equilibration. The system was run in a positive electrospray (ESI⁺) mode with a fragmentor voltage of 150 V. The (S)-norcoclaurine parent ion [*M* + *H*]⁺ was observed and isolated with an *m/z* value of 272.1 at a retention time of 3.3 min. Quantification was performed via LC-MS/MS by monitoring the *m/z* signal of 107.049 fragmented from the parental ion with the *m/z* of 272.1 at a collision energy of 23 V. LC-MS/MS data were quantified using Agilent MassHunter Quantitative Analysis (version 10.0), and the integrated LC-MS/MS peak areas were quantified using a seven-point (S)-norcoclaurine (Sigma-Aldrich) calibration curve with a concentration ranging from 12.5 to 400 μg/L.

Measurement of the *S. stipitis* Growth. The high-performance variants for the production of (S)-norcoclaurine were cultured in SC-LEU-URA medium at 30 °C and 250 rpm for 24–36 h. Saturated cultures were diluted to an OD₆₀₀ of 0.2 and cultured in sterile 96-well plates containing 200 μL of 2X selective synthetic medium SC-LEU-URA supplemented with 6% glucose and 0.5% L-ascorbic acid at 30 °C with continuous shaking at 205 rpm. OD₆₀₀ were measured every 15 min using a Synergy Eon Microplate Spectrophotometer (BioTek) for a period of 96 h. The growth of the SS *leu*[−] *ura*[−] strain transformed with plasmid pSsLeuUra (only carrying the *LEU2* and *URA3* selection markers) was measured simultaneously as a control to evaluate the impact of (S)-norcoclaurine on cell growth.

Gene Copy Number Assay. The copy numbers of the *ARO1* and *NCS* genes in the (S)-norcoclaurine high producers were determined using quantitative PCR. Cells were cultured in SC-LEU-URA medium at 30 °C and 250 rpm for 40–48 h. Three milliliters of cells were collected, and the genomic DNA was extracted using a Wizard genomic DNA purification kit (Promega). Primers were designed for three genes: *ARO1* was used to quantify the upstream module, *NCS* was used to quantify the downstream module, and *XYL1* was used as an internal reference. Primer sequences are listed in Table S1. Twenty microliters of reaction was conducted in a StepOnePlus Real-Time PCR system (Thermo Fisher Scientific) using Maxima SYBR Green qPCR Master Mix (Thermo Fisher Scientific) according to the manufacturer's instructions, and 1 μL of genomic DNA (at a concentration range of 1–3 ng/μL) was used for each sample. The copy number of *ARO1* and *NCS* was calculated using the

corresponding standard curve. The assay was performed in triplicate.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acssynbio.0c00123>.

Evaluation of the transpositional integration efficiency for G-SA-U fragment enclosed by TIRs. The shikimate production levels and GFP expression intensities of the SS *ura⁻* variants carrying the transpositional insertion of the G-SA-U fragment. The transposition efficiency for iterative integration of the upstream module and the downstream module in the two libraries. The vector maps of the main constructs in this study. Verification of the functioning of Hermes transposon in *S. stipitis*. (PDF)

Primers used for qPCR assay. Genbank files of main constructs in this study (ZIP).

■ AUTHOR INFORMATION

Corresponding Authors

Zengyi Shao – Department of Chemical and Biological Engineering, NSF Engineering Research Center for Biorenewable Chemicals, Bioeconomy Institute, and Interdepartmental Microbiology Program, Iowa State University, Ames, Iowa, United States; The Ames Laboratory, Ames, Iowa, United States; orcid.org/0000-0001-6817-8006; Phone: 515-294-1132; Email: zyshao@iastate.edu

Mingfeng Cao – Department of Chemical and Biological Engineering, Iowa State University, Ames, Iowa, United States; orcid.org/0000-0002-6750-3871; Email: mfc@illinois.edu

Chao Yang – The Key Laboratory of Molecular Microbiology and Technology, Ministry of Education, College of Life Sciences, Nankai University, Tianjin, China; Email: yangc20119@nankai.edu.cn

Authors

Yuxin Zhao – The Key Laboratory of Molecular Microbiology and Technology, Ministry of Education, College of Life Sciences, Nankai University, Tianjin, China; Department of Chemical and Biological Engineering, Iowa State University, Ames, Iowa, United States; orcid.org/0000-0002-9267-2036

Zhanyi Yao – Department of Chemical and Biological Engineering, Iowa State University, Ames, Iowa, United States

Deon Ploessl – Department of Chemical and Biological Engineering, Iowa State University, Ames, Iowa, United States

Saptarshi Ghosh – Department of Chemical and Biological Engineering, Iowa State University, Ames, Iowa, United States

Marco Monti – Manchester Institute of Biotechnology and School of Chemistry, University of Manchester, Manchester, U.K.

Daniel Schindler – Manchester Institute of Biotechnology and School of Chemistry, University of Manchester, Manchester, U.K.; orcid.org/0000-0001-7423-6540

Meirong Gao – Department of Chemical and Biological Engineering, Iowa State University, Ames, Iowa, United States

Yizhi Cai – Manchester Institute of Biotechnology and School of Chemistry, University of Manchester, Manchester, U.K.

Mingqiang Qiao – The Key Laboratory of Molecular Microbiology and Technology, Ministry of Education, College of

Life Sciences, Nankai University, Tianjin, China; orcid.org/0000-0002-5435-1514

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acssynbio.0c00123>

Author Contributions

Y.Z., M.C., and Z.S. conceptualized the initial idea, designed the experiments, and performed troubleshooting; Z.Y. contributed to the confirmation of the rate-limiting step in the (S)-norcoclaurine biosynthesis; D.P. contributed to the copy number determination; M.M., D.S., and Y.C. contributed to the genome sequence analysis; S.G. and M.G. contributed to the sensor plasmid development; M.Q. and C.Y. were involved in the initial training of Y.Z.; Y.Z., M.C., and Z.S. wrote and polished the manuscript.

Funding

Z.Y., S.G., M.G., M.C., Z.S., and the consumables in this project were supported by the National Science Foundation (Grant Nos. 1716837 and 1749782) and Iowa State University Biobased Chemical/Bioprocess Seed Grant (PG110304-DD10742). We acknowledge the scholarship provided by Nankai University to support Y.Z. for her stipend as a visiting student and the NSF Graduate Research Fellowships Program to support D.P., M.M., D.S., and Y.C. acknowledge the support by the BBSRC Grant (No. R121730) and the Volkswagen Foundation “Life? Initiative” Grant (No. 94 771).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. T. W. Jeffries for sharing the strain *S. stipitis* FLP-UC7 (*ura3-3*, NRRL Y21448) and the plasmid pJML545. We thank Prof. J. Dueber for sharing the *S. cerevisiae* plasmid containing *TyrH** and *DOD*, and Prof. S. B. Sandmeyer for sharing pJY3919 containing the Hermes transposon system for *Y. lipolytica*. We also thank Dr. S. M. Rigby for flow cytometry and Dr. L. Showman and Dr. K. Narayanaswamy for the detection of (S)-norcoclaurine.

■ REFERENCES

- (1) Zhang, M. M., Wang, Y., Ang, E. L., and Zhao, H. (2016) Engineering microbial hosts for production of bacterial natural products. *Nat. Prod. Rep.* 33, 963–987.
- (2) Cravens, A., Payne, J., and Smolke, C. D. (2019) Synthetic biology strategies for microbial biosynthesis of plant natural products. *Nat. Commun.* 10, 2142.
- (3) Markham, K. A., and Alper, H. S. (2018) Synthetic biology expands the industrial potential of *Yarrowia lipolytica*. *Trends Biotechnol.* 36, 1085–1095.
- (4) Wiebe, M. G., Koivuranta, K., Penttilä, M., and Ruohonen, L. (2012) Lipid production in batch and fed-batch cultures of *Rhodospiridium toruloides* from 5 and 6 carbon carbohydrates. *BMC Biotechnol.* 12, 26.
- (5) Li, Y., Zhao, Z., and Bai, F. (2007) High-density cultivation of oleaginous yeast *Rhodospiridium toruloides* Y4 in fed-batch culture. *Enzyme Microb. Technol.* 41, 312–317.
- (6) Zhang, S., Skerker, J. M., Rutter, C. D., Maurer, M. J., Arkin, A. P., and Rao, C. V. (2016) Engineering *Rhodospiridium toruloides* for increased lipid production. *Biotechnol. Bioeng.* 113, 1056–1066.
- (7) Agbogbo, F. K., and Coward-Kelly, G. (2008) Cellulosic ethanol production using the naturally occurring xylose-fermenting yeast *Pichia stipitis*. *Biotechnol. Lett.* 30, 1515–1524.
- (8) Gao, M., Cao, M., Suastegui, M., Walker, J., Rodriguez Quiroz, N., Wu, Y., Tribby, D., Okerlund, A., Stanley, L., Shanks, J. V., and

- Shao, Z. (2017) Innovating a nonconventional yeast platform for producing shikimate as the building block of high-value aromatics. *ACS Synth. Biol.* 6, 29–38.
- (9) Satyanarayana, T., and Kunze, G. (2017) *Yeast diversity in human welfare*, pp 439–453, Springer, Singapore.
- (10) Radecka, D., Mukherjee, V., Mateo, R. Q., Stojiljkovic, M., Foulquie-Moreno, M. R., and Thevelein, J. M. (2015) Looking beyond *Saccharomyces*: the potential of non-conventional yeast species for desirable traits in bioethanol fermentation. *FEMS Yeast Res.* 15, 1.
- (11) Shao, Z., Zhao, H., and Zhao, H. (2009) DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* 37, No. e16.
- (12) Bourgeois, L., Pyne, M. E., and Martin, V. J. J. (2018) A highly characterized synthetic landing pad system for precise multicopy gene integration in yeast. *ACS Synth. Biol.* 7, 2675–2685.
- (13) Santos, C. N., Regitsky, D. D., and Yoshikuni, Y. (2013) Implementation of stable and complex biological systems through recombinase-assisted genome engineering. *Nat. Commun.* 4, 2503.
- (14) Wang, G., Zhao, Z., Ke, J., Engel, Y., Shi, Y. M., Robinson, D., Bingol, K., Zhang, Z., Bowen, B., Louie, K., Wang, B., Evans, R., Miyamoto, Y., Cheng, K., Kosina, S., De Raad, M., Silva, L., Luhrs, A., Lubbe, A., Hoyt, D. W., Francavilla, C., Otani, H., Deutsch, S., Washton, N. M., Rubin, E. M., Mouncey, N. J., Visel, A., Northen, T., Cheng, J. F., Bode, H. B., and Yoshikuni, Y. (2019) CRAGE enables rapid activation of biosynthetic gene clusters in undomesticated bacteria. *Nat. Microbiol.* 4, 2498–2510.
- (15) Juhas, M., and Ajioka, J. W. (2016) Lambda Red recombinase-mediated integration of the high molecular weight DNA into the *Escherichia coli* chromosome. *Microb. Cell Fact.* 15, 172.
- (16) Li, L., Zheng, G., Chen, J., Ge, M., Jiang, W., and Lu, Y. (2017) Multiplexed site-specific genome engineering for overproducing bioactive secondary metabolites in actinomycetes. *Metab. Eng.* 40, 80–92.
- (17) Li, L., Wei, K., Liu, X., Wu, Y., Zheng, G., Chen, S., Jiang, W., and Lu, Y. (2019) aMSGE: advanced multiplex site-specific genome engineering with orthogonal modular recombinases in actinomycetes. *Metab. Eng.* 52, 153–167.
- (18) Horwitz, A. A., Walter, J. M., Schubert, M. G., Kung, S. H., Hawkins, K., Platt, D. M., Hernday, A. D., Mahatdejkul-Meadows, T., Szeto, W., Chandran, S. S., and Newman, J. D. (2015) Efficient multiplexed integration of synergistic alleles and metabolic pathways in yeasts via CRISPR-Cas. *Cell Syst.* 1, 88–96.
- (19) Shi, S., Liang, Y., Zhang, M. M., Ang, E. L., and Zhao, H. (2016) A highly efficient single-step, markerless strategy for multicopy chromosomal integration of large biochemical pathways in *Saccharomyces cerevisiae*. *Metab. Eng.* 33, 19–27.
- (20) Bassalo, M. C., Garst, A. D., Halweg-Edwards, A. L., Grau, W. C., Domaille, D. W., Mutalik, V. K., Arkin, A. P., and Gill, R. T. (2016) Rapid and efficient one-step metabolic pathway integration in *E. coli*. *ACS Synth. Biol.* 5, 561–568.
- (21) Agard, D. A., and Sedat, J. W. (1983) Three-dimensional architecture of a polytene nucleus. *Nature* 302, 676–681.
- (22) Englaender, J. A., Jones, J. A., Cress, B. F., Kuhlman, T. E., Linhardt, R. J., and Koffas, M. A. G. (2017) Effect of genomic integration location on heterologous protein expression and metabolic engineering in *E. coli*. *ACS Synth. Biol.* 6, 710–720.
- (23) Bryant, J. A., Sellars, L. E., Busby, S. J., and Lee, D. J. (2014) Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res.* 42, 11383–11392.
- (24) Wu, X. L., Li, B. Z., Zhang, W. Z., Song, K., Qi, H., Dai, J. B., and Yuan, Y. J. (2017) Genome-wide landscape of position effects on heterogeneous gene expression in *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 10, 189.
- (25) Chen, X., and Zhang, J. (2016) The genomic landscape of position effects on protein expression level and noise in yeast. *Cell Syst.* 2, 347–354.
- (26) Sauer, C., Syvertsson, S., Bohorquez, L. C., Cruz, R., Harwood, C. R., van Rij, T., and Hamoen, L. W. (2016) Effect of genome position on heterologous gene expression in *Bacillus subtilis*: an unbiased analysis. *ACS Synth. Biol.* 5, 942–947.
- (27) Cui, Z., Jiang, X., Zheng, H., Qi, Q., and Hou, J. (2019) Homology-independent genome integration enables rapid library construction for enzyme expression and pathway optimization in *Yarrowia lipolytica*. *Biotechnol. Bioeng.* 116, 354–363.
- (28) Thompson, A., and Gasson, M. J. (2001) Location effects of a reporter gene on expression levels and on native protein synthesis in *Lactococcus lactis* and *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 67, 3434–3439.
- (29) Tsakraklides, V., Kaminen, A., Consiglio, A. L., MacEwen, K., Friedlander, J., Blitzblau, H. G., Hamilton, M. A., Crabtree, D. V., Su, A., Afshar, J., Sullivan, J. E., LaTouf, W. G., South, C. R., Greenhagen, E. H., Shaw, A. J., and Brevnova, E. E. (2018) High-oleate yeast oil without polyunsaturated fatty acids. *Biotechnol. Biofuels* 11, 131.
- (30) Friedlander, J., Tsakraklides, V., Kaminen, A., Greenhagen, E. H., Consiglio, A. L., MacEwen, K., Crabtree, D. V., Afshar, J., Nugent, R. L., Hamilton, M. A., Joe Shaw, A., South, C. R., Stephanopoulos, G., and Brevnova, E. E. (2016) Engineering of a high lipid producing *Yarrowia lipolytica* strain. *Biotechnol. Biofuels* 9, 77.
- (31) Munoz-Lopez, M., and Garcia-Perez, J. L. (2010) DNA transposons: nature and applications in genomics. *Curr. Genomics* 11, 115–128.
- (32) Li, X., Burnight, E. R., Cooney, A. L., Malani, N., Brady, T., Sander, J. D., Staber, J., Wheelan, S. J., Joung, J. K., McCray, P. B., Jr., Bushman, F. D., Sinn, P. L., and Craig, N. L. (2013) piggyBac transposase tools for genome engineering. *Proc. Natl. Acad. Sci. U. S. A.* 110, E2279–2287.
- (33) Arensburger, P., Hice, R. H., Zhou, L., Smith, R. C., Tom, A. C., Wright, J. A., Knapp, J., O'Brochta, D. A., Craig, N. L., and Atkinson, P. W. (2011) Phylogenetic and functional characterization of the *hAT* transposon superfamily. *Genetics* 188, 45–57.
- (34) Gangadharan, S., Mularoni, L., Fain-Thornton, J., Wheelan, S. J., and Craig, N. L. (2010) DNA transposon Hermes inserts into DNA in nucleosome-free regions *in vivo*. *Proc. Natl. Acad. Sci. U. S. A.* 107, 21966–21972.
- (35) Patterson, K., Yu, J., Landberg, J., Chang, I., Shavarebi, F., Bilanchone, V., and Sandmeyer, S. (2018) Functional genomics for the oleaginuous yeast *Yarrowia lipolytica*. *Metab. Eng.* 48, 184–196.
- (36) Guo, Y., Park, J. M., Cui, B., Humes, E., Gangadharan, S., Hung, S., FitzGerald, P. C., Hoe, K. L., Grewal, S. I., Craig, N. L., and Levin, H. L. (2013) Integration profiling of gene function with dense maps of transposon integration. *Genetics* 195, 599–609.
- (37) Hickman, A. B., Voth, A. R., Ewis, H., Li, X., Craig, N. L., and Dyda, F. (2018) Structural insights into the mechanism of double strand break formation by Hermes, a *hAT* family eukaryotic DNA transposase. *Nucleic Acids Res.* 46, 10286–10301.
- (38) Evertts, A. G., Plymire, C., Craig, N. L., and Levin, H. L. (2007) The hermes transposon of *Musca domestica* is an efficient tool for the mutagenesis of *Schizosaccharomyces pombe*. *Genetics* 177, 2519–2523.
- (39) Jeffries, T. W., Grigoriev, I. V., Grimwood, J., Laplaza, J. M., Aerts, A., Salamov, A., Schmutz, J., Lindquist, E., Dehal, P., Shapiro, H., Jin, Y. S., Passoth, V., and Richardson, P. M. (2007) Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*. *Nat. Biotechnol.* 25, 319–326.
- (40) Lee, J. W., Zhu, J. Y., Scordia, D., and Jeffries, T. W. (2011) Evaluation of ethanol production from corn cob using *Scheffersomyces (Pichia) stipitis* CBS 6054 by volumetric scale-up. *Appl. Biochem. Biotechnol.* 165, 814–822.
- (41) Jeffries, T. W., Grigoriev, I. V., Grimwood, J., Laplaza, J. M., Aerts, A., Salamov, A., Schmutz, J., Lindquist, E., Dehal, P., Shapiro, H., Jin, Y. S., Passoth, V., and Richardson, P. M. (2007) Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*. *Nat. Biotechnol.* 25, 319–326.
- (42) Jeffries, T. W., and Van Vleet, J. R. (2009) *Pichia stipitis* genomics, transcriptomics, and gene clusters. *FEMS Yeast Res.* 9, 793–807.
- (43) Suastegui, M., Guo, W., Feng, X., and Shao, Z. (2016) Investigating strain dependency in the production of aromatic

compounds in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* 113, 2676–2685.

(44) Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M., and Smolke, C. D. (2015) Complete biosynthesis of opioids in yeast. *Science* 349, 1095–1100.

(45) Suastegui, M., and Shao, Z. (2016) Yeast factories for the production of aromatic compounds: from building blocks to plant secondary metabolites. *J. Ind. Microbiol. Biotechnol.* 43, 1611–1624.

(46) Facchini, P. J., Bohlmann, J., Covello, P. S., De Luca, V., Mahadevan, R., Page, J. E., Ro, D. K., Sensen, C. W., Storms, R., and Martin, V. J. (2012) Synthetic biosystems for the production of high-value plant metabolites. *Trends Biotechnol.* 30, 127–131.

(47) Cao, M., Gao, M., Suastegui, M., Mei, Y., and Shao, Z. (2020) Building microbial factories for the production of aromatic amino acid pathway derivatives: From commodity chemicals to plant-sourced natural products. *Metab. Eng.* 58, 94–132.

(48) Cao, M., Gao, M., Ploessl, D., Song, C., and Shao, Z. (2018) CRISPR-mediated genome editing and gene repression in *Scheffersomyces stipitis*. *Biotechnol. J.* 13, No. 1700598.

(49) Cao, M., Gao, M., Lopez-Garcia, C. L., Wu, Y., Seetharam, A. S., Severin, A. J., and Shao, Z. (2017) Centromeric DNA facilitates nonconventional yeast genetic engineering. *ACS Synth. Biol.* 6, 1545–1553.

(50) Brewer, B. J., and Fangman, W. L. (1987) The localization of replication origins on ARS plasmids in *S. cerevisiae*. *Cell* 51, 463–471.

(51) Fiaux, J., Cakar, Z. P., Sonderegger, M., Wuthrich, K., Szyperski, T., and Sauer, U. (2003) Metabolic-flux profiling of the yeasts *Saccharomyces cerevisiae* and *Pichia stipitis*. *Eukaryotic Cell* 2, 170–180.

(52) Suastegui, M., Matthiesen, J. E., Carraher, J. M., Hernandez, N., Rodriguez Quiroz, N., Okerlund, A., Cochran, E. W., Shao, Z., and Tessonnier, J. P. (2016) Combining metabolic engineering and electrocatalysis: application to the production of polyamides from sugar. *Angew. Chem., Int. Ed.* 55, 2368–2373.

(53) Curran, K. A., Leavitt, J. M., Karim, A. S., and Alper, H. S. (2013) Metabolic engineering of muconic acid production in *Saccharomyces cerevisiae*. *Metab. Eng.* 15, 55–66.

(54) Luttik, M. A., Vuralhan, Z., Suij, E., Braus, G. H., Pronk, J. T., and Daran, J. M. (2008) Alleviation of feedback inhibition in *Saccharomyces cerevisiae* aromatic amino acid biosynthesis: quantification of metabolic impact. *Metab. Eng.* 10, 141–153.

(55) Cao, M., Seetharam, A. S., Severin, A. J., and Shao, Z. (2017) Rapid isolation of centromeres from *Scheffersomyces stipitis*. *ACS Synth. Biol.* 6, 2028–2034.

(56) Fu, J., Wenzel, S. C., Perlova, O., Wang, J., Gross, F., Tang, Z., Yin, Y., Stewart, A. F., Muller, R., and Zhang, Y. (2008) Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. *Nucleic Acids Res.* 36, No. e113.

(57) Loeschcke, A., Markert, A., Wilhelm, S., Wirtz, A., Rosenau, F., Jaeger, K. E., and Drepper, T. (2013) TREX: a universal tool for the transfer and expression of biosynthetic pathways in bacteria. *ACS Synth. Biol.* 2, 22–33.

(58) Bian, X., Tang, B., Yu, Y., Tu, Q., Gross, F., Wang, H., Li, A., Fu, J., Shen, Y., Li, Y. Z., Stewart, A. F., Zhao, G., Ding, X., Muller, R., and Zhang, Y. (2017) Heterologous production and yield improvement of epothilones in *Burkholderiales* Strain DSM 7029. *ACS Chem. Biol.* 12, 1805–1812.

(59) Gemperlein, K., Hoffmann, M., Huo, L., Pilak, P., Petzke, L., Muller, R., and Wenzel, S. C. (2017) Synthetic biology approaches to establish a heterologous production system for coronatines. *Metab. Eng.* 44, 213–222.

(60) Shao, Z., Luo, Y., and Zhao, H. (2012) DNA assembler method for construction of zeaxanthin-producing strains of *Saccharomyces cerevisiae*. *Methods Mol. Biol.* 898, 251–262.

(61) Shao, Z., and Zhao, H. (2013) Construction and engineering of large biochemical pathways via DNA assembler. *Methods Mol. Biol.* 1073, 85–106.

(62) Shao, Z., and Zhao, H. (2014) Manipulating natural product biosynthetic pathways via DNA assembler. *Curr. Protoc. Chem. Biol.* 6, 65–100.

(63) Orr-Weaver, T. L., Szostak, J. W., and Rothstein, R. J. (1981) Yeast transformation: a model system for the study of recombination. *Proc. Natl. Acad. Sci. U. S. A.* 78, 6354–6358.

(64) Trenchard, I. J., and Smolke, C. D. (2015) Engineering strategies for the fermentative production of plant alkaloids in yeast. *Metab. Eng.* 30, 96–104.

(65) Li, Y., Li, S., Thodey, K., Trenchard, I., Cravens, A., and Smolke, C. D. (2018) Complete biosynthesis of noscapine and halogenated alkaloids in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 115, E3922–E3931.

(66) Minami, H., Kim, J. S., Ikezawa, N., Takemura, T., Katayama, T., Kumagai, H., and Sato, F. (2008) Microbial production of plant benzylisoquinoline alkaloids. *Proc. Natl. Acad. Sci. U. S. A.* 105, 7393–7398.

(67) Hagel, J. M., and Facchini, P. J. (2013) Benzylisoquinoline alkaloid metabolism: a century of discovery and a brave new world. *Plant Cell Physiol.* 54, 647–672.

(68) Deng, X., Zhao, L., Fang, T., Xiong, Y., Ogutu, C., Yang, D., Vimolmangkang, S., Liu, Y., and Han, Y. (2018) Investigation of benzylisoquinoline alkaloid biosynthetic pathway and its transcriptional regulation in lotus. *Hortic. Res.* 5, 29.

(69) DeLoache, W. C., Russ, Z. N., Narcross, L., Gonzales, A. M., Martin, V. J., and Dueber, J. E. (2015) An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. *Nat. Chem. Biol.* 11, 465–471.

(70) Gold, N. D., Gowen, C. M., Lussier, F. X., Cautha, S. C., Mahadevan, R., and Martin, V. J. (2015) Metabolic engineering of a tyrosine-overproducing yeast platform using targeted metabolomics. *Microb. Cell Fact.* 14, 73.

(71) Suastegui, M., Yu Ng, C., Chowdhury, A., Sun, W., Cao, M., House, E., Maranas, C. D., and Shao, Z. (2017) Multilevel engineering of the upstream module of aromatic amino acid biosynthesis in *Saccharomyces cerevisiae* for high production of polymer and drug precursors. *Metab. Eng.* 42, 134–144.

(72) Koyanagi, T., Nakagawa, A., Sakurama, H., Yamamoto, K., Sakurai, N., Takagi, Y., Minami, H., Katayama, T., and Kumagai, H. (2012) Eukaryotic-type aromatic amino acid decarboxylase from the root colonizer *Pseudomonas putida* is highly specific for 3,4-dihydroxyphenyl-L-alanine, an allelochemical in the rhizosphere. *Microbiology* 158, 2965–2974.

(73) Mao, J., Liu, Q., Li, Y., Yang, J., Song, X., Liu, X., Xu, H., and Qiao, M. (2018) A high-throughput method for screening of L-tyrosine high-yield strains by *Saccharomyces cerevisiae*. *J. Gen. Appl. Microbiol.* 64, 198–201.

(74) Laplaza, J. M., Torres, B. R., Jin, Y.-S., and Jeffries, T. W. (2006) *Sh ble* and *Cre* adapted for functional genomics and metabolic engineering of *Pichia stipitis*. *Enzyme Microb. Technol.* 38, 741–747.

(75) Trenchard, I. J., Siddiqui, M. S., Thodey, K., and Smolke, C. D. (2015) *De novo* production of the key branch point benzylisoquinoline alkaloid reticuline in yeast. *Metab. Eng.* 31, 74–83.

(76) Gao, S., Tong, Y., Zhu, L., Ge, M., Zhang, Y., Chen, D., Jiang, Y., and Yang, S. (2017) Iterative integration of multiple-copy pathway genes in *Yarrowia lipolytica* for heterologous β -carotene production. *Metab. Eng.* 41, 192–201.

(77) Domrose, A., Weihmann, R., Thies, S., Jaeger, K. E., Drepper, T., and Loeschcke, A. (2017) Rapid generation of recombinant *Pseudomonas putida* secondary metabolite producers using yTREX. *Synth. Syst. Biotechnol.* 2, 310–319.

(78) Lu, P., Davis, B. P., Hendrick, J., and Jeffries, T. W. (1998) Cloning and disruption of the β -isopropylmalate dehydrogenase gene (*LEU2*) of *Pichia stipitis* with *URA3* and recovery of the double auxotroph. *Appl. Microbiol. Biotechnol.* 49, 141–146.

(79) Boeke, J. D., Trueheart, J., Natsoulis, G., and Fink, G. R. (1987) 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol.* 154, 164–175.

(80) Zelaya, I. A., Anderson, J. A., Owen, M. D., and Landes, R. D. (2011) Evaluation of spectrophotometric and HPLC methods for shikimic acid determination in plants: models in glyphosate-resistant and -susceptible crops. *J. Agric. Food Chem.* 59, 2202–2212.