ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

Iterative Optimization of Xylose Catabolism in *Saccharomyces cerevisiae* Using Combinatorial Expression Tuning

Luke N. Latimer,[1] John E. Dueber[2]

[1]Department of Chemistry, University of California, Berkeley, California

[2]Department of Bioengineering, University of California, 2151 Berkeley Way, Berkeley, California 94720; telephone: +1 510 643 4616; fax: +1 510 642 9725; e-mail: jdueber@berkeley.edu

ABSTRACT: A common challenge in metabolic engineering is rapidly identifying rate-controlling enzymes in heterologous pathways for subsequent production improvement. We demonstrate a workflow to address this challenge and apply it to improving xylose utilization in *Saccharomyces cerevisiae*. For eight reactions required for conversion of xylose to ethanol, we screened enzymes for functional expression in *S. cerevisiae*, followed by a combinatorial expression analysis to achieve pathway flux balancing and identification of limiting enzymatic activities. In the next round of strain engineering, we increased the copy number of these limiting enzymes and again tested the eight-enzyme combinatorial expression library in this new background. This workflow yielded a strain that has a ~70% increase in biomass yield and ~240% increase in xylose utilization. Finally, we chromosomally integrated the expression library. This library enriched for strains with multiple integrations of the pathway, which likely were the result of tandem integrations mediated by promoter homology.

Biotechnol. Bioeng. 2017;114: 1301–1309.

KEYWORDS: metabolic engineering; xylose utilization; combinatorial expression; synthetic biology

Introduction

The primary goal of metabolic engineering is to enable the biochemical synthesis of a variety of chemicals (Stephanopoulos et al., 1998). To accomplish this goal, particularly for longer, poorly characterized pathways, what is needed are systematic, scalable, and pathway-independent methodologies (Yadav et al., 2012). Accordingly, a number of systematic methodologies have been developed. Specific examples include modifying host genomes (Chavez et al., 2016; Cong et al., 2013; Wetmore et al., 2015), engineering enzyme properties (Guntas et al., 2010; Romero et al., 2012; Voigt et al., 2002), tuning enzyme expression (Ajikumar et al., 2010; Deaner and Alper, 2016; Lee et al., 2015; Mutalik et al., 2013), and introducing pathway regulation (Brockman and Prather, 2015; Hoynes-O'Connor and Moon, 2015; Kushwaha and Salis, 2015).

An important step in metabolic pathway engineering is identifying rate-controlling enzymes for subsequent engineering (Leonard et al., 2010; Ro et al., 2006). An effective solution to this challenge, multivariate modular metabolic engineering (MMME), was pioneered by Ajikumar et al. (2010), (reviewed in Biggs et al., 2014 and Pandey et al., 2016). Optimizing taxadiene production in *Escherichia coli*, the authors first divided the pathway into expression balanced modules based on enzymatic activities (high or low); then they varied expression of the two modules in a combinatorial fashion. The production landscape from the first screen was used to inform adjustments to expression in a second combinatorial screen, which ultimately improved taxadiene titers by ~15,000 fold. While MMME covers an expression space with a limited number of strains, it is reliant on previous characterization of enzymatic activities and does not allow for granular optimization that may be needed to address interdependencies, particularly within modules.

An alternative approach to identify rate-controlling steps is to use combinatorial expression libraries as a means of surveying the expression landscape, where enzymes that enrich for high expression are likely controlling pathway flux. Previous work from our lab took advantage of improved DNA assembly technologies to develop a combinatorial expression library designed to simultaneously optimize each enzymatic step allowing granular resolution while requiring no prior biochemical knowledge of the pathway, which is ideal for poorly characterized pathways (Lee et al., 2013). The cost for these benefits is a much larger library size. To handle this size, we used regression modelling to predict genotypes that preferentially produce different metabolites in the branched violacein pathway (Lee et al., 2013). In another instance, we employed selection for xylose utilization in *Saccharomyces cerevisiae* to search through the library (Latimer et al., 2014).

In this work, we sought to expand upon our earlier optimization experiments by demonstrating an iterative optimization scheme that uses combinatorial expression libraries to identify rate-

Correspondence to: J.E. Dueber

Contract grant sponsor: Energy Biosciences Institute
Received 26 October 2016; Revision received 8 January 2017; Accepted 2 February 2017
Accepted manuscript online 6 February 2017;
Article first published online 20 February 2017 in Wiley Online Library (http://onlinelibrary.wiley.com/doi/10.1002/bit.26262/abstract).
DOI 10.1002/bit.26262

controlling enzymes and inform further engineering of the metabolic pathway. For our demonstration, we chose the fungal xylose utilization pathway in *S. cerevisiae* as a model system. Following our proposed scheme, we created an iteratively optimized strain with improved xylose utilization capabilities. Finally, we investigated chromosomal integration of these combinatorial libraries.

Materials and Methods

Strains and Media

All single cassette plasmids were transformed in chemically competent TG1 (Lucigen, Madison, WI) cells and grown in Lysogeny Broth (LB, Difco; Becton, Dickinson and Company, Franklin Lakes, NJ) with spectinomycin (50 mg/L), chloramphenicol (34 mg/L), or ampicillin (100 mg/L). All multi-gene plasmid assemblies were transformed at 1.8 kV into TransforMax EPI300 (Epicentre, Madison, WI) electrocompetent *E. coli* and cultured with kanamycin (25 mg/L). The *S. cerevisiae* strain BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) (0.19 ± 0.01 g cell/L at $OD_{600} = 1$) (Latimer et al., 2014) was grown at 30 °C. Wild-type yeast cultures were grown in YPD (10 g/L Bacto Yeast Extract; 20 g/L Bacto Peptone; 20 g/L Dextrose). Yeast were transformed according to standard lithium acetate protocol (Gietz and Woods, 2006) and were selected for growth on synthetic drop-out media (6.7 g/L Difco Yeast Nitrogen Base w/o Amino Acids; 2 g/L Drop-out Mix Synthetic minus Leucine and Uracil, w/o Yeast Nitrogen Base [US Biological, Salem, MA]; 20 g/L Dextrose or 20 g/L Xylose [Sigma, St. Louis, MO]).

Plasmids and Combinatorial Expression Libraries

All DNA oligomers were synthesized by Integrated DNA Technologies (IDT). All *S. stipitis* cassette plasmids were used as described previously (Latimer et al., 2014). Recoded SsXRs sequences were designed using http://genedesign.jbei.org/ and IDT codon optimization tool and synthesized by IDT. *S. cerevisiae* genes were cloned by PCR from the BY4741 genome using primers listed in Supplemental Table S2 followed by either subcloning or golden gate (Lee et al., 2015). Cassette plasmids were assembled as described previously (Latimer et al., 2014). YFP fusion cassettes were cloned using standard parts from the Yeast Toolkit (Lee et al., 2015). Multi-gene plasmid libraries or plasmids with specific promoter genotypes were assembled in a BsmBI golden gate reaction using 20 fmol of each plasmid (Engler and Marillonnet, 2014). For the variable copy number of XR libraries, four libraries (pLNL336-9 corresponding to 0–3 SsXR copies, respectively) were cloned and purified separately and then mixed at equal molarity. Plasmid information is summarized in Supplemental Table S3, all plasmid sequences are deposited in the SynBERC registry (https://registry.synberc.org/login) and backbone vector construction details are available upon request.

Library plasmids were transformed as described previously (Latimer et al., 2014) then resuspended in 500 mL SX-LU under anaerobic conditions (media supplemented with 0.01 g/L ergosterol, 0.43 g/L Tween 80 and 1.4 g/L ethanol; 1 L Erlenmeyer screw cap flask flushed with $N_{2(g)}$) with an aliquot plated on SD-LU for sampling initial library coverage and diversity. For the integrated library transformation, a 50 mL culture ($OD_{600} = 2$) was prepared and transformed with 12 ug of NotI linearized pLNL616L library plasmid and 4 ug of linearized I-SceI cutter plasmid, pML1429 (Lee et al., 2015).

Fluorescence Measurements

BY4741 transformed with each YFP fusion cassette were inoculated in 300 μL of SD-Ura in a 96-well block and grown 18 h. Culture optical density and YFP fluorescence were measured using a TECAN Infinite® M1000 at 600 nm and 515/528 nm, respectively.

Shotgun Proteomics

BY4741 expressing the indicated plasmids were grown to mid-log phase anaerobically on xylose, lysed by bead beating at −20 °C (90 s total with 30 s cycles), and analyzed as described previously (Latimer et al., 2014).

Library Enrichments

Library enrichments were performed as described previously (Latimer et al., 2014). Primers used to genotype individual colonies with the TaqMan-based TRAC protocol (Lee et al., 2013) are listed in Supplementary Table S2.

Xylose Fermentations

Anaerobic xylose fermentations were performed as described previously (Latimer et al., 2014). Briefly, colonies were grown in SD-LU for 24 h then diluted into SX-LU for 48 h of aerobic growth in 24-well blocks at 750 rpm. A normalized OD_{600} was diluted into 40 mL of SX-LU in a 125 mL serum vial sealed with a rubber stopper, flushed with $N_{2(g)}$ and shaken at 100 rpm.

Metabolite Quantification

Previously frozen media samples were pelleted, and supernatant was transferred to GC/MS vials. Refractive index was measured for 10 μL sample on a Shimadzu LC20AD HPLC equipped with a Rezex RFP-fast acid H+ column (100 × 7.8 mm, 55 °C) run with 1 mL/min 0.01 N H_2SO_4 mobile phase. Metabolite concentrations were calculated using a standard curve.

PacBio Sequencing

Strain LL441AE was inoculated in 5 mL SD-LU for 24 h from glycerol stock. The saturated culture was diluted 1:100 into three 50 mL cultures in SD-LU media and grown in baffled shake flasks (250 mL, 200 rpm). When culture OD_{600} was six, the cultures were combined and genomic DNA was prepared using QIAGEN Genomic-tip 100/G according to manufacturer instructions using spooling to isolate the precipitated DNA. Genomic DNA was prepared as a PacBio Long Insert Library and sequenced by the Genomics Resource Center at the University of Maryland School

of Medicine using two SMRT cells (P6-C4). Sequencing yielded 120-fold genome coverage with 143,753 reads averaging ~10 kb with a P90 of 18,000. The genome sequence was assembled de novo using SMRT Portal (Pacific Biosciences, Menlo Park, CA) HGAP.3 yielding 53 scaffolds to determine tandem sequence at the Ura3 locus. Consensus genome sequence is available in Supplemental Data 1 and is described in the Supplementary Information.

Results and Discussion

A Workflow for Iterative Strain Engineering Using Combinatorial Expression Libraries

One of the long-standing challenges in engineering microbial metabolism is having systematic approaches to increase flux to achieve industrially viable titers, particularly for longer pathways, which often have limited biochemical characterization (Trenchard and Smolke, 2015; Yadav et al., 2012). Towards this goal, we present a methodology centered around combinatorial expression libraries as a systematic tool for evaluating the regulatory landscape of a metabolic pathway (Fig. 1A). Importantly, our workflow does not rely on prior biochemical characterization of the enzymes and simultaneously optimizes each enzymatic step, providing granular resolution for interdependencies during optimization.

The requirements for our engineering strategy are a chassis for which there are characterized promoters and genetic techniques to enable library transformations. The workflow begins with identification of coding sequences for the requisite enzymes to construct a biosynthetic pathway to produce the metabolite(s) of interest (Fig. 1A). Each enzyme is then expressed, and in vivo protein synthesis is validated. While directly assaying enzymatic activity is ideal for validation, scalable activity-independent assays, such as fluorescent reporter fusions, or shotgun proteomics, are more amenable for longer pathways and for enzymes lacking established activity assays. In the case of fluorescent protein fusions, microscopy may help determine if the protein is soluble (diffuse fluorescence) or insoluble (punctate fluorescence can be due to aggregation) (Kaganovich et al., 2008). Using acceptable coding sequences, the pathway is assembled as a combinatorial expression library employing a set of varied-strength promoters to simultaneously regulate the expression of each gene in the pathway. This library is transformed into the host of choice, the resulting strains are characterized, and high-performing strains are genotyped. Using this information, enzymatic steps that limit pathway flux can be identified by their enrichment for strong promoters (Latimer et al., 2014). These enzymatic activities can then be addressed by further overexpression via increased copy number or directed evolution of the target genes. The modifications are incorporated into another round of optimization by either recloning the library with an improved mutant enzyme or

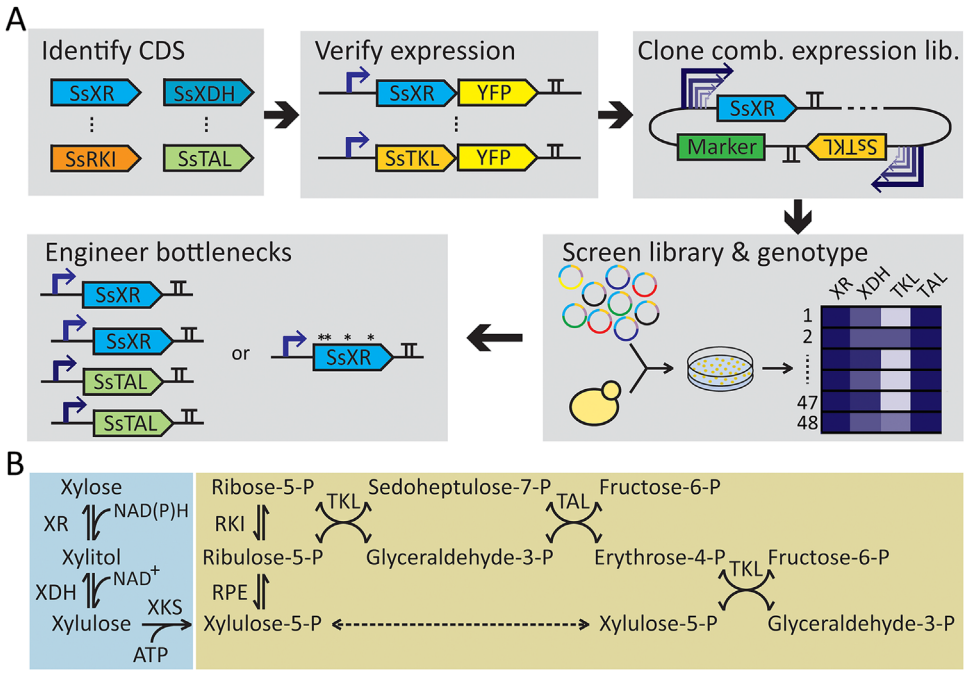

Figure 1. A methodology for iterative pathway engineering informed by expression optimization. **A)** Coding sequences (CDS) are identified and each is cloned into an expression cassette. Expression is verified (e.g. fluorescent protein fusion, activity assay, proteomics). If needed, more CDSs are screened. The metabolic pathway is then assembled with the verified enzymes as a combinatorial expression library, transformed into the chassis of choice and subjected to a screen or selection. Enriched strains are genotyped to map expression space. Enzymes enriched for high expression are modified by increased copy number and/or mutagenesis and the expression library transformation and screen is repeated. * represents point mutations **B)** The model fungal xylose utilization pathway. Heterologous (left box) xylose reductase (XR), xylitol dehydrogenase (XDH), and xylulokinase (XKS) convert xylose into pentose phosphate pathway (PPP) intermediate, xylulose-5-phosphate. The non-oxidative PPP (right box) converts xylulose-5-P into the glycolytic intermediates fructose-6-P and glyceraldehyde-3-P. Abbreviations are in accordance with the Saccharomyces Genome Database (http://www.yeastgenome.org).

reintroducing the combinatorial expression library into the strain that overexpresses additional copies of the limiting enzyme. This process could, in theory, be iterated until improvements are no longer observed. Given the large library sizes resulting from the combinatorial nature of this technique, pathways with high-throughput screens or selections are most amenable to this methodology. Alternatively, modeling has been used to predict important enzymes using limited library sampling (Lee et al., 2013).

Xylose Utilization as a Model Pathway for Iterative Optimization

To demonstrate the workflow described above, we chose to optimize xylose utilization in *S. cerevisiae*. This pathway has long been of interest for fermenting lignocellulosic hydrolysates as part of second-generation biofuel and commodity chemical production (Matsushika et al., 2009). While heterologous expression of two enzymes (xylose reductase, XR, and xylitol dehydrogenase, XDH) is sufficient for xylose utilization, rapid growth requires over-expression of additional enzymes including the non-oxidative PPP (Chu and Lee, 2007). The eight-enzyme fungal pathway (Fig. 1B) is representative of a longer pathway that our workflow is designed to improve. Also, we previously demonstrated expression tuning of the *Scheffersomyces stipitis* xylose utilization pathway in *S. cerevisiae* (Latimer et al., 2014).

Expression Characterization Identifies Superior Expression of ScRPE1 and ScTKL1 Compared to SsRPE and SsTKL

The first step in our workflow is to identify functional coding sequences. This is particularly important as subsequent analysis of the expression space may not identify limiting enzymes that are poorly expressed. During our previous characterization of the xylose catabolic pathway, we compared enzyme expression of our engineered xylose utilizing strains using shotgun proteomics (Latimer et al., 2014). Even when expressed with the strong *pTDH3* promoter, we observed very few spectral counts for peptides corresponding to two of the enzymes in the PPP: ribulose-5-phosphate epimerase (SsRPE) and transketolase (SsTKL), which we interpret as poor expression or stability of these two enzymes (Redding-Johanson et al., 2011). Using only this information as our initial screen for protein expression, we searched for alternative coding sequences that express better. While we originally selected our enzymes from the natural xylose utilizing yeast, *S. stipitis*, hypothesizing that these enzymes would have high activities, we turned to the native *S. cerevisiae* homologs for improved expression of SsRPE and SsTKL.

To characterize protein expression of these homologs, we compared two sequence-independent, scalable techniques: fluorescent reporter fusions and shotgun proteomics. First, we cloned each enzyme individually with N- and C-terminal yellow fluorescent protein (YFP) fusions, and screened for fluorescence (Fig. 2A). Notably, tagged SsTKL showed almost no fluorescence, indicating poor expression. All homolog pairs except ScTKL1 and SsPYK have substantial and comparable bulk fluorescence and diffuse fluorescence by microscopy (Supplemental Fig. S1), including

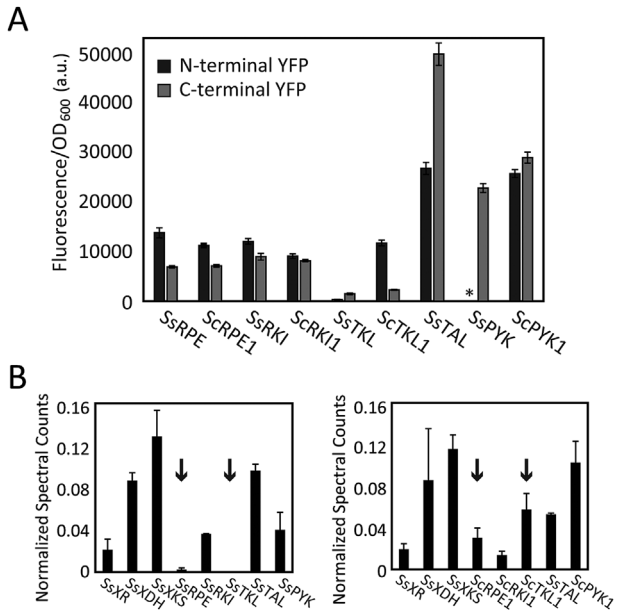

Figure 2. Fluorescent reporter fusions and proteomics identify poor expression of heterologous PPP enzymes SsRPE and SsTKL. **A**) Fluorescence values normalized by optical density for glucose-grown saturated yeast cultures expressing *S. stipitis* or *S. cerevisiae* PPP enzymes N- or C-terminal fused to fluorescent reporter YFP. N = 6. * was not tested. **B**) Shotgun proteomics peptide abundance normalized to total endogenous counts for strains expressing the indicated enzymes each regulated by *pTDH3* when grown anaerobically to mid-log on xylose. N = 3 samples for the left graph and N = 2 for the right graph. Arrows highlight low counts of SsRPE and SsTKL compared to ScRPE1 and ScTKL1.

SsRPE and ScRPE1. Previously, we observed low expression of untagged SsRPE by shotgun proteomics (Latimer et al., 2014). We hypothesize that the improved expression of SsRPE-YFP fusion compared to previous shotgun proteomics measurements is due to the YFP fusion (Janczak et al., 2015).

Second, we expressed two different pathways: (i) the entire *S. stipitis* pathway, or (ii) a chimeric pathway and analyzed anaerobic, xylose-cultured cells by shotgun proteomics (Fig. 2B). For the chimeric pathway, the downstream enzymes were *S. cerevisiae* homologs with the lone exception of SsTAL which was retained because ScTAL1 overexpression has previously been shown to be toxic (Jin et al., 2005). As seen earlier under aerobic conditions (Latimer et al., 2014), we observed little or no signal for SsRPE and SsTKL. In the chimeric strain, we measured substantial peptide counts for both ScRPE1 and ScTKL1, indicating improved expression of these homologs. Comparing the two expression characterization methods, we find them to generally be in agreement, except for SsRPE expression. Thus, tag-free approaches such as shotgun proteomics are more ideal because tagging the protein can modify the expression or activity, either in a positive manner as seen here with SsRPE but likely also in a negative manner for other proteins. However, protein fusions are adequate when instrumentation and expertise for proteomics are not available. Going forward, readily available tools for rapid characterization of heterologous in vivo protein expression/activity are needed (Redding-Johanson et al., 2011).

Chimeric PPP Expression Library Enriches Different Expression of TKL and TAL

Based on the expression characterization, we chose a chimeric xylose utilization pathway with all enzyme coding sequences taken from *S. stipitis* except *ScRPE1* and *ScTKL1*. Similar to our previous optimizations, we assembled a combinatorial expression library using the set of five constitutive promoters of varying strength spanning approximately three orders of magnitude (Supplemental Fig. S2) to drive transcription of all eight genes. The rank-ordering of the promoters was previously shown to be coding sequence independent, thus, providing a means of determining a relative expression profile for each gene in an enriched population and enabling use of the promoters for modulating any pathway in *S. cerevisiae* (Lee et al., 2013). The pathway was divided into two plasmids: the first containing the upstream *SsXR*, *SsXDH*, *SsXKS* while the second has the nonoxidative PPP and *SsPYK*. *SsPYK* was included based on Lu and Jeffries work that found overexpression of ScPYK1 improved xylose fermentation (Lu and Jeffries, 2007); we used this homolog to allow for differentiation from endogenous PYK by shotgun proteomics as well as the possibility that SsPYK is subject to less allosteric regulation. We transformed this library into the laboratory *S. cerevisiae* strain BY4741 yielding 1.8-fold library coverage and enriched the library over ten 100-fold back-dilutions by selecting for anaerobic growth supplying xylose as the sole carbon source. Using TaqMan rapid analysis of combinatorial assemblies (TRAC), a TaqMan-probe based methodology (Lee et al., 2013), we genotyped enriched colonies, which are summarized in Figure 3A.

As expected, the enrichment profile for this library is similar to the previous anaerobic enrichments of the *S. stipitis* pathway: high expression of SsXR and SsXDH, intermediate expression of SsXKS, low expression of ScRPE1 and SsRKI, and moderate expression of SsPYK (Latimer et al., 2014). Interestingly, ScTKL1 enriched for stronger promoters (i.e., no genotyped strains had the weakest *pREV1*), while previously SsTKL had enriched most strongly for *pREV1*. This difference is likely important, given that switching from SsTKL to ScTKL1 improved cell growth on xylose (Supplemental Fig. S3). While the reason for different promoter enrichment upstream of transketolase is not obvious, it highlights the importance of not relying on previous optimization results when changing coding sequences.

We also observed enrichment of weaker promoters driving expression of SsTAL compared to our previous libraries (Latimer et al., 2014). If we assume no change to expression or activity of SsTAL upon switching RPE and TKL coding sequences from *S. stipitis* to *S. cerevisiae*, then this result is unexpected. Proteomics of the ScPPP expressing strain, which expresses SsTAL, shows a decrease in SsTAL expression compared to the strain expressing the SsPPP with no change in protein sequence coverage (Fig. 2B; data not shown); there is also no change in SsXR, SsXDH and SsXKS counts between these two strains (Fig. 2B). This context-dependent enrichment of SsTAL is further evidence that systematic optimization is necessary following any major perturbation to the pathway, as has been previously speculated (Santos and Stephanopoulos, 2008), and which is accounted for in our workflow (Fig. 1A).

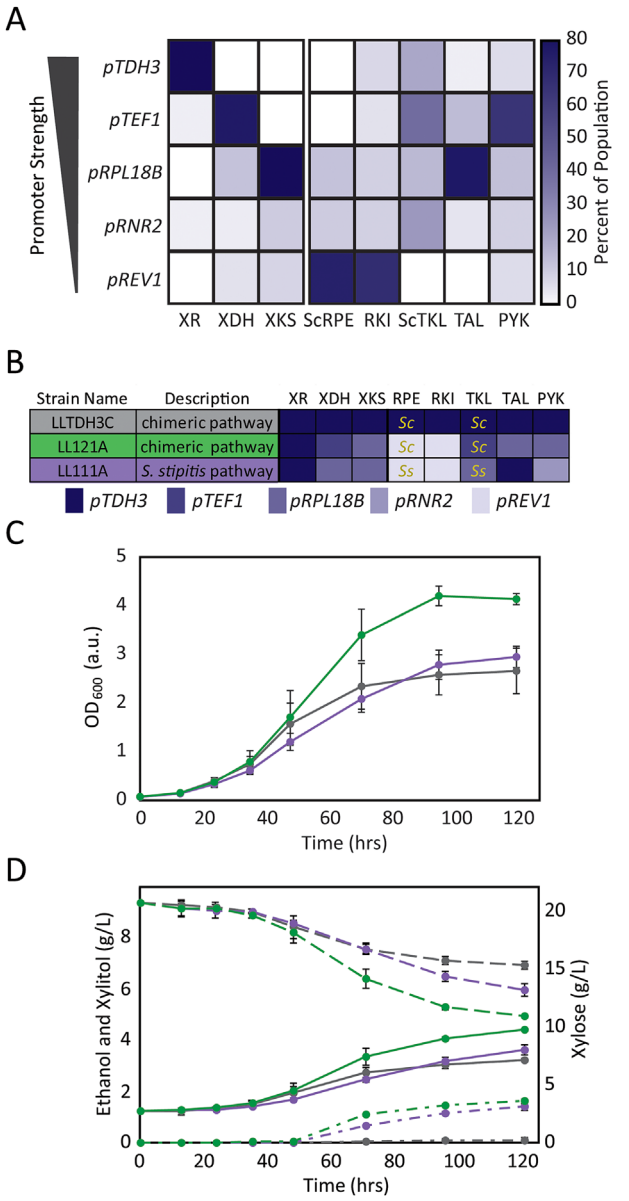

Figure 3. The chimeric xylose utilization pathway expression library enriches for higher TKL expression and lower TAL expression, improving xylose utilization. **A**) Enrichment profile heatmap generated from genotyping 48 colonies expressing the chimeric xylose catabolism pathway from an anaerobically enriched expression library. Heatmap colors correspond to the percentage of colonies with a given promoter regulating the corresponding gene. **B**) Genotypes and description of reference, all *pTDH3* strain LLTDH3C, and enriched strains. Shading of each square corresponds to promoter. RPE and TKL coding sequences are from either *S. cerevisiae* (Sc) or *S. stipitis* (Ss). LL111A was previously enriched in Latimer et al., 2014. **C**) Anaerobic growth curves in synthetic xylose media supplemented with 0.01 g/L ergosterol, 0.43 g/L Tween 80 and 1.4 g/L ethanol for strains indicated in (B). **D**) Extracellular metabolite concentrations for fermentations shown in (C). Xylose (– – –), Xylitol (– · – · –), Ethanol (———). Error bars represent SD of biological triplicates.

The Chimeric PPP Improves Xylose Utilization in *S. cerevisiae* BY4741

We aimed to determine whether the homolog substitution and subsequent expression optimization improved xylose utilization of

individual strains. Accordingly, we recloned a predominant enriched genotype from the chimeric library enrichment, LL121A, and compared it in an anaerobic fermentation to a reference, the naive high expression chimera strain, LLTDH3C, and a previously anaerobically enriched *S. stipitis* pathway expressing strain, LL111A (Latimer et al., 2014) (Fig. 3B). Previously, the all *S. stipitis*/all *pTDH3* strain showed inferior growth and xylose utilization compared to LL111A; however, the chimeric LLTDH3C showed similar growth (Fig. 3C), but slower xylose utilization at saturation compared to LL111A (Fig. 3D). Thus, the switch to the chimeric pathway appears to have improved strain performance, particularly growth. These improvements are further amplified by expression optimization: LL121A reached 42% higher culture densities, metabolized 30% more xylose, and produced 22% more ethanol compared to the previous optimal strain LL111A (Fig. 3D). The performance of LL121A is recapitulated in a second enriched genotype (i.e., LL121B) (Supplemental Fig. S4).

Iterative Optimization Enriches for Multiple Copies of XR and Improves Xylose Utilization

Based on the first round of expression optimization (Fig. 3A), we identified four targets for overexpression: SsXR, SsXDH, ScTKL1, and SsTAL. While not every one of these genes enriched for maximum expression of *pTDH3*, metabolic control theory (Fell, 2005) predicts that the pathway may not be limited solely by xylose reductase activity and these other activities may be important upon additional SsXR expression. Further, we speculated that genes which do not require additional expression should enrich for lower expression within the library to reduce unnecessary protein burden (Kafri et al., 2016). We cloned these four enzymes with strong promoters and integrated them into BY4741 to yield yJD228. We were concerned that a single extra chromosomal copy of XR may not yield a sufficient increase in xylose reductase activity based on the exclusive enrichment for *pTDH3* driving expression of SsXR as well as the low catalytic activity of SsXR compared to the other enzymes (Chen et al., 2012). To address this, we redesigned the XR-XDH-XKS library plasmid to additionally include a varying number of additional copies of *SsXR* (0, 1, 2, or 3) driven by a promoter of comparable strength to *pTDH3*, *pCCW12* (Lee et al., 2015). To each of these plasmid libraries, we also added TRAC-compatible barcodes corresponding to the *SsXR* copy number for genotyping by TRAC. To verify increased xylose reductase activity with increasing copy number, we measured activity in cell lysate and observed a linear increase in reductase activity with copy number (Supplemental Fig. S5). However, the ratio of activity improvement was less than 1:1 with copy number, suggesting that factors in transcription, translation, or folding of SsXR become limiting with increased copy number.

We transformed the modified library into yJD228, which yielded a lower 35% library coverage, a consequence of a four-fold increase in library size compared to the first library, due to the additional variable of varying *SsXR* copy number. The library was subjected to enrichment as before. For many of the genes, the enrichment profile for the anaerobically enriched library showed less dramatic enrichment for a specific promoter (Fig. 4A). For example, ScTKL1 and SsTAL showed a slight preference for lower expression. Most striking is the enrichment of SsXR, which not only enriched exclusively for high expression for the gene controlled by the

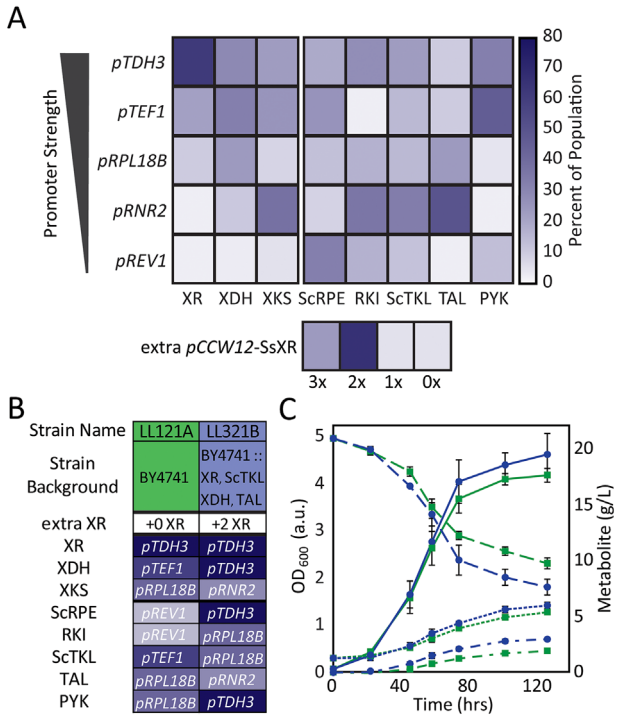

Figure 4. Iterative expression optimization enriches for four copies of *SsXR* and improves xylose utilization. **A**) Enrichment profile heatmap generated from genotyping 48 colonies from an anaerobically enriched promoter library regulating the chimeric xylose pathway with 0–3 extra copies of *pCCW12-SsXR* in strain background yJD228. **B**) Genotypes and description of enriched strains. **C**) Anaerobic fermentations in synthetic xylose media supplemented with 0.01g/L ergosterol, 0.43g/L Tween 80 and 1.4 g/L ethanol for strains indicated in (B): LL121A (square); LL321B (circle). OD600 (——) Xylose (— — —), Xylitol (— · — · —) Ethanol (- - - - - -). Error bars represent SD of biological triplicates.

variable promoter but also enriched for an additional two or three copies yielding a total of four or five copies of *SsXR* in each strain. A number of these strains also enriched for stronger promoters driving XDH and XKS expression compared with previous anaerobic enrichments, which we speculate may be necessary to balance increased xylose reductase flux. For a number of the downstream enzymes, there was weaker enrichment, which suggests that these activities are no longer critical for achieving high pathway flux. An exception is SsPYK which enriched for higher expression compared to previous libraries.

To test whether these changes in the library improved xylose utilization and ensure any improvement was not a result from strain adaptation, we reassembled an enriched genotype, LL321B in the yJD228 background. Compared with the first optimized strain, LL321B outperforms LL121A utilizing 19% more xylose and producing 15% more ethanol (Fig. 4B and C; Supplemental Table S1), demonstrating that iterative optimization informed by combinatorial expression libraries can improve strain performance. Compared to adapted XR-XDH-XKS expressing strains reported in the literature (Supplemental Table S1), LL321B shows a little less than half the xylose fermentation rates, indicating further modifications to the strain background are needed.

A Chromosomally Integrated Expression Library Selects for Multiple Integrations

Many metabolic engineering applications require genomic integration of the enzymes for expression reproducibility, stability, and selection-free fermentations (Shi et al., 2016). In our implementation of this iterative strain engineering strategy, expression optimization occurred using episomal plasmids. A common solution would be the integration of the optimized plasmid sequence to ascertain the benefits of chromosomal expression. However, strain performance can decrease upon integration since expression from a given sequence typically decreases when integrated (Jensen et al., 2014). As a result, further optimization after integration by increasing enzyme expression or copy number may be needed. Alternatively, by optimizing expression in the genome upfront, the final strain can be used directly and thereby avoid the need for these additional steps. Thus, we sought to implement chromosomal-based combinatorial expression optimization. Previously, genomic integration of these large 10^5 member libraries was impractical due to low integration efficiency, however, this efficiency has been dramatically improved with recently developed techniques (Lee et al., 2015; Ryan et al., 2014; Wingler and Cornish, 2011).

To implement the use of chromosomally-integrated combinatorial expression libraries, we redesigned our starting strain and library to: (i) reduce recombination rates; and (ii) increase expression of limiting enzymes identified earlier, anticipating an overall decrease in expression. Accordingly, we recloned the xylose utilization pathway with unique terminators using the Golden Gate scheme defined in our Yeast Toolkit (Lee et al., 2015) comprising 10 genes divided for integration into two loci (Fig. 5A). To simplify the library integration, we assigned five genes a single promoter (i.e., fixed expression) and integrated them into yRC864, a BY4741 strain with an I-SceI restriction site integrated at the Ura3 locus (Chen et al., 2015), to create yJD231. These genes included *ScRPE1* and *SsRKI* regulated by intermediate strength promoters and two extra copies of recoded (to lower the probability of undesired recombination) *SsXR* and one copy of *SsXDH* expressed by strong

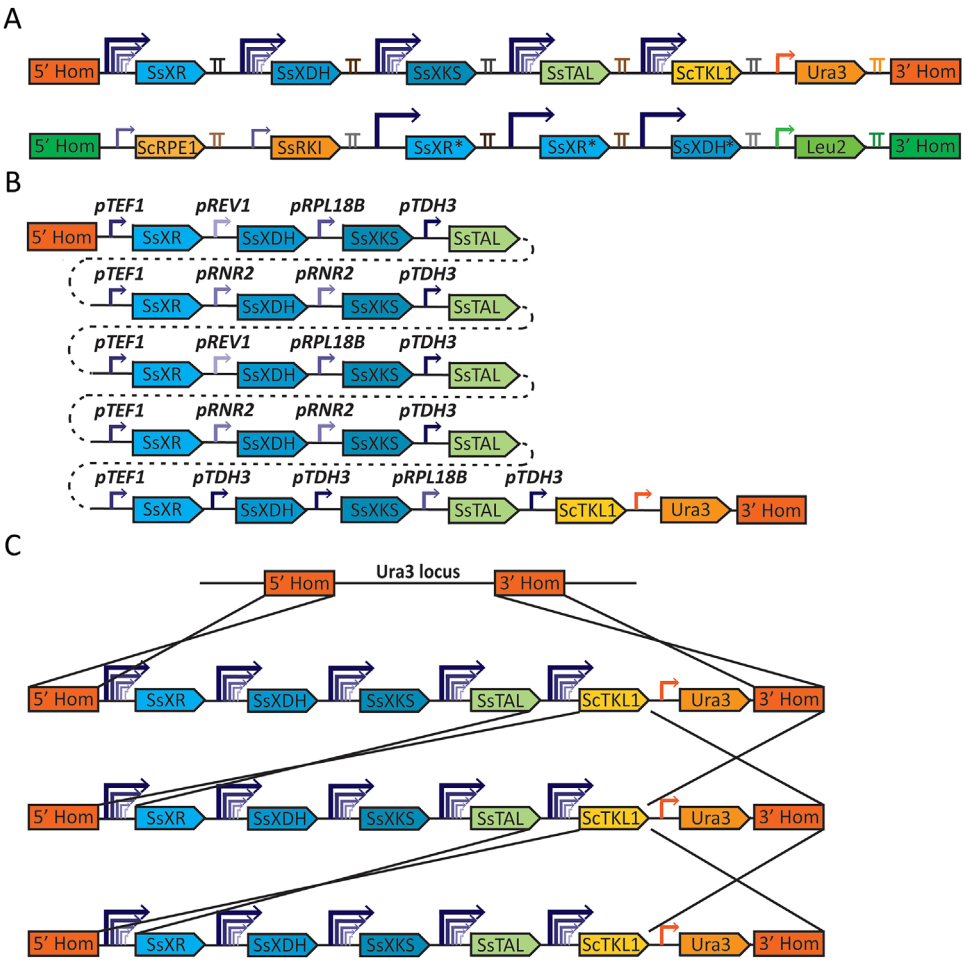

Figure 5. Enrichment of the integrated expression library yields multiple integration events. **A**) Vector design for chromosomally integrated combinatorial expression library strains. Homology regions correspond to the associated auxotrophic marker locus. *indicates recoded genes. **B**) Genome architecture at the Ura3 locus for expression library isolate, LL441AE, as determined by PacBio sequencing. Dashed lines are shown for continuity, however there is no extra DNA between rows. Terminators are omitted for clarity. **C**) Proposed mechanism for multiple integration events mediated by homologous recombination at repeated promoters in the combinatorial library.

promoters. The combinatorial library included five genes: *SsXR*, *SsXDH*, *SsXKS*, *SsTAL*, and *ScTKL1* targeted for integration into the Ura3 locus. We transformed this library along with an I-SceI cutter plasmid into yJD231 to enhance integration efficiency (Lee et al., 2015), which yielded over 20-fold library coverage. As before, we subjected the library to selection by anaerobic growth on xylose.

By comparing enrichment profiles for colonies before enrichment and after ~66 generations we observed that the library enriched for strains that had integrated the combinatorial library genes multiple times. This is indicated by multiple genes each showing probe signal in the TRAC reaction for multiple different promoters (Supplemental Fig. S6). To verify gene duplication of the library as detected by TRAC, we sequenced a strain from a later enrichment, LL441AE, using PacBio technology, which provides reads up to 30 kb, enabling detection of tandem gene duplications (Supplemental Data 1). Our de novo contig assembly showed that LL441AE has five or more tandem repeats of the first four genes in the combinatorial library expression construct (*SsXR*, *SsXDH*, *SsXKS*, and *SsTAL*), spanning 50 kb, at the Ura3 locus (Fig. 5B, Supplemental Table S3). Based on the absence of *ScTKL1* in all but the last repeat, we propose a mechanism of tandem integration mediated by homologous recombination between the promoter of *SsXR* on one DNA polymer and the promoter of *ScTKL1* on another DNA polymer (Fig. 5C). Furthermore, at least three of these repeats were likely integrated during the initial library transformation because the promoters driving *SsXDH* and *SsXKS* are different for each of these repeats. To generate these unique combinations of promoters, unique DNA fragments from the transformation mixture would be required.

Enriching for the subpopulation of strains with multiple integrations results in loss of the information provided by genotyping because the expression space of the library is no longer defined. While it is useful to learn that expression is limiting for one or more of these enzymes, the extra effort to identify the limiting enzymes partially undermines the utility of using these well-characterized promoter libraries.

When compared in xylose fermentations, these enriched multi-copy integrated strains show 80% faster growth rates as well as faster xylose consumption rates and ethanol productivities compared to the episomal expression optimized strains (Supplemental Table S1). Thus, these integrated libraries may yield high performing strains, though their enzyme expression rules are not fully understood. When optimizing other metabolic pathways where the library is enriched by screening, these multiple integrant strains should not prevent learning the expression rules so long as sufficient single-integration colonies are screened to observe enrichment trends.

Conclusion

In this work, we presented a workflow designed to iteratively and systematically improve a metabolic pathway, which should be pathway-independent, and demonstrated its use with the fungal xylose utilization pathway as a model system. The final strain resulting from iterative optimization, LL321B, grows to approximately three times the density while consuming five times more xylose during anaerobic fermentation compared to the initial, naive strain where all enzymes were from *S. stipitis* and expressed with the strong promoter, *pTDH3* (Latimer et al., 2014). Further improvements were achieved through enrichment of chromosomal integration of the pathway which resulted in selection for strains that had undergone gene duplication. To improve this workflow and other metabolic engineering efforts, better methods for enzyme characterization, ideally activity-based, are needed. Further, methods for identifying other key regulatory enzymes in central metabolism, such as inverse metabolic engineering, can be incorporated into this scheme to decrease initial biases. As more multi-gene metabolic pathways are engineered to produce molecules of interest, the chief challenge shifts from gene identification towards achieving relevant titers of these products and developing generalizable, systematic optimization strategies to enable rapidly improving pathway flux.

We thank Daniel Medina-Cleghorn, Rebecca Kohnz, and Daniel Nomura for assistance in collecting and analyzing proteomic data. We are grateful for expertise provided by Jeffrey Skerker and Zachary Russ for genomic DNA purification and analysis of the PacBio sequencing data. Lastly, we appreciate the careful review and discussions about this manuscript with Michael Lee and Jamie Cate. We have no conflict of interest to declare. Initial funding for this research was provided by the Energy Biosciences Institute.

References

Ajikumar PK, Xiao WH, Tyo KEJ, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G. 2010. Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. Science 330:70–74.

Biggs BW, Paepe BD, Santos CNS, Mey MD, Ajikumar PK. 2014. Multivariate modular metabolic engineering for pathway and strain optimization. Curr Opin Biotechnol 29:156––62.

Brockman IM, Prather KLJ. 2015. Dynamic metabolic engineering: New strategies for developing responsive cell factories. Biotechnol J 10:1360–1369.

Chavez A, Tuttle M, Pruitt BW, Ewen-campen B, Chari R, Ter-ovanesyan D, Haque SJ, Cecchi RJ, Kowal EJK, Buchthal J, Housden BE, Perrimon N, Collins JJ, Church G. 2016. Comparison of Cas9 activators in multiple species. Nat Methods 13:563–569.

Chen S-H, Hwang D-R, Chen G-H, Hsu N-S, Wu Y-T, Li T-L, Wong C-H. 2012. Engineering transaldolase in *Pichia stipitis* to improve bioethanol production. ACS Chem Biol 7:481–486.

Chen R, Rishi HS, Potapov V, Yamada MR, Yeh VJ, Chow T, Cheung CL, Jones AT, Johnson TD, Keating AE, DeLoache WC, Dueber JE. 2015. A barcoding strategy enabling higher-throughput library screening by microscopy. ACS Synth Biol 4:1205–1216.

Chu BCH, Lee H. 2007. Genetic improvement of *Saccharomyces cerevisiae* for xylose fermentation. Biotechnol Adv 25:425–441.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hus PD, Wu X, Jiang W, Marraffini LA, Zhang F. 2013. Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–822.

Deaner M, Alper HS. 2016. Promoter and terminator discovery and engineering. Adv Biochem Eng/Biotechnol 1–24.

Engler C, Marillonnet S. 2014. Golden gate cloning. Methods Mol Biol 1116:119–131.

Fell DA. 2005. Metabolic control analysis. Top Curr Genet 13:69–80.

Gietz RD, Woods RA. 2006. Yeast transformation by the LiAc/SS carrier DNA/PEG method. Methods Mol Biol 313:107–120.

Guntas G, Purbeck C, Kuhlman B. 2010. Engineering a protein—protein interface using a computationally designed library. Proc Natl Acad Sci USA 107: 19296–19301.

Hoynes-O'Connor A, Moon TS. 2015. Programmable genetic circuits for pathway engineering. Curr Opin Biotechnol 36:115–121.

Janczak M, Bukowski M, Goreck A, Dubin G, Dubin A, Wladyka B. 2015. A systematic investigation of the stability of green fluorescent protein fusion proteins. J Polish Biochem Soc 62:10–14.

Jensen NB, Strucko T, Kildegaard KR, David F, Maury J, Mortensen UH, Forster J, Nielsen J, Borodina I. 2014. Easyclone: Method for iterative chromosomal integration of multiple genes in *Saccharomyces cerevisiae*. FEMS Yeast Res 14:238–248.

Jin YS, Alper HS, Yang YT, Stephanopoulos G. 2005. Improvement of xylose uptake and ethanol production in recombinant *Saccharomyces cerevisiae* through an inverse metabolic engineering approach. Appl Environ Microbiol 71:8249.

Kafri M, Metzl-raz E, Jona G, Barkai N. 2016. The cost of protein production. Cell Rep 14:22–31.

Kaganovich D, Kopito R, Frydman J. 2008. Misfolded proteins partition between two distinct quality control compartments. Nature 454:1088–1096.

Kushwaha M, Salis HM. 2015. A portable expression resource for engineering cross-species genetic circuits and pathways. Nat Commun 6:1–11.

Latimer LN, Lee ME, Medina-Cleghorn D, Kohnz RA, Nomura DK, Dueber JE. 2014. Employing a combinatorial expression approach to characterize xylose utilization in *Saccharomyces cerevisiae*. Metab Eng 25:20–29.

Lee ME, Aswani A, Han AS, Tomlin CJ, Dueber JE. 2013. Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. Nucleic Acids Res 1–11.

Lee ME, Deloache WC, Cervantes B, Dueber JE. 2015. A highly characterized yeast toolkit for modular, multipart assembly. ACS Synth Biol 4:975–986.

Leonard E, Ajikumar PK, Thayer K, Xiao W-H, Mo JD, Tidor B, Stephanopoulos G, Prather KLJ. 2010. Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. Proc Natl Acad Sci USA 107:13654–13659.

Lu C, Jeffries T. 2007. Shuffling of promoters for multiple genes to optimize xylose fermentation in an engineered *Saccharomyces cerevisiae* strain. Appl Environ Microbiol 73:6072–6077.

Matsushika A, Inoue H, Kodaki T, Sawayama S. 2009. Ethanol production from xylose in engineered *Saccharomyces cerevisiae* strains: current state and perspectives. Appl Microbiol Biotechnol 84:37–53.

Mutalik VK, Guimaraes JC, Cambray G, Lam C, Christoffersen MJ, Mai Q-A, Tran AB, Paull M, Keasling JD, Arkin AP, Endy D. 2013. Precise and reliable gene expression via standard transcription and translation initiation elements. Nat Genet 10:354–360.

Pandey RP, Parajuli P, Koffas MAG, Sohng JK. 2016. Microbial production of natural and non-natural flavonoids: Pathway engineering, directed evolution and systems/synthetic biology. Biotechnol Adv 34:634–662.

Redding-Johanson AM, Batth TS, Chan R, Krupa R, Szmidt HL, Adams PD, Keasling JD, Lee TS, Mukhopadhyay A, Petzold CJ. 2011. Targeted proteomics for metabolic pathway optimization: Application to terpene production. Metab Eng 13(2):194–203.

Ro D-K, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J, Chang MCY, Withers ST, Shiba Y, Sarpong R, Keasling JD. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440:3–6.

Romero PA, Krause A, Arnold FH. 2012. Navigating the protein fitness landscape with gaussian processes. Proc Natl Acad Sci USA 100:E193–E201.

Ryan OW, Skerker JM, Maurer MJ, Li X, Tsai JC, Poddar S, Lee ME, DeLoache WC, Dueber JE, Arkin AP, Cate JHD. 2014. Selection of chromosomal DNA libraries using a multiplex CRISPR system. Elife 3:1–15.

Santos CNS, Stephanopoulos G. 2008. Combinatorial engineering of microbes for optimizing cellular phenotype. Curr Opin Chem Biol 12:168–176.

Shi S, Liang Y, Zhang MM, Ang EL, Zhao H. 2016. A highly efficient single-step, markerless strategy for multi-copy chromosomal integration of large biochemical pathways in *Saccharomyces cerevisiae*. Metab Eng 33:19–27.

Stephanopoulos GN, Aristidou AA, Nielsen J. 1998. Metabolic engineering principles and methodologies. London: Elsevier Science. p 1–14.

Trenchard IJ, Smolke CD. 2015. Engineering strategies for the fermentative production of plant alkaloids in yeast. Metab Eng 30:96–104.

Voigt CA, Martinez C, Wang Z-G, May SL, Arnold FH. 2002. Protein building blocks preserved by recombination. Nat Struct Biol 9:553–558.

Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, Blow MJ, Bristow J, Butland G, Arkin AP, Deutschbauer A. 2015. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. MBio 6:1–15.

Wingler LM, Cornish VW. 2011. Reiterative recombination for the in vivo assembly of libraries of multigene pathways. Proc Natl Acad Sci USA 108: 15135–15140.

Yadav VG, Mey MD, Lim CG, Ajikumar PK, Stephanopoulos G. 2012. The future of metabolic engineering and synthetic biology: Towards a systematic practice. Metab Eng 14:233–241.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site.