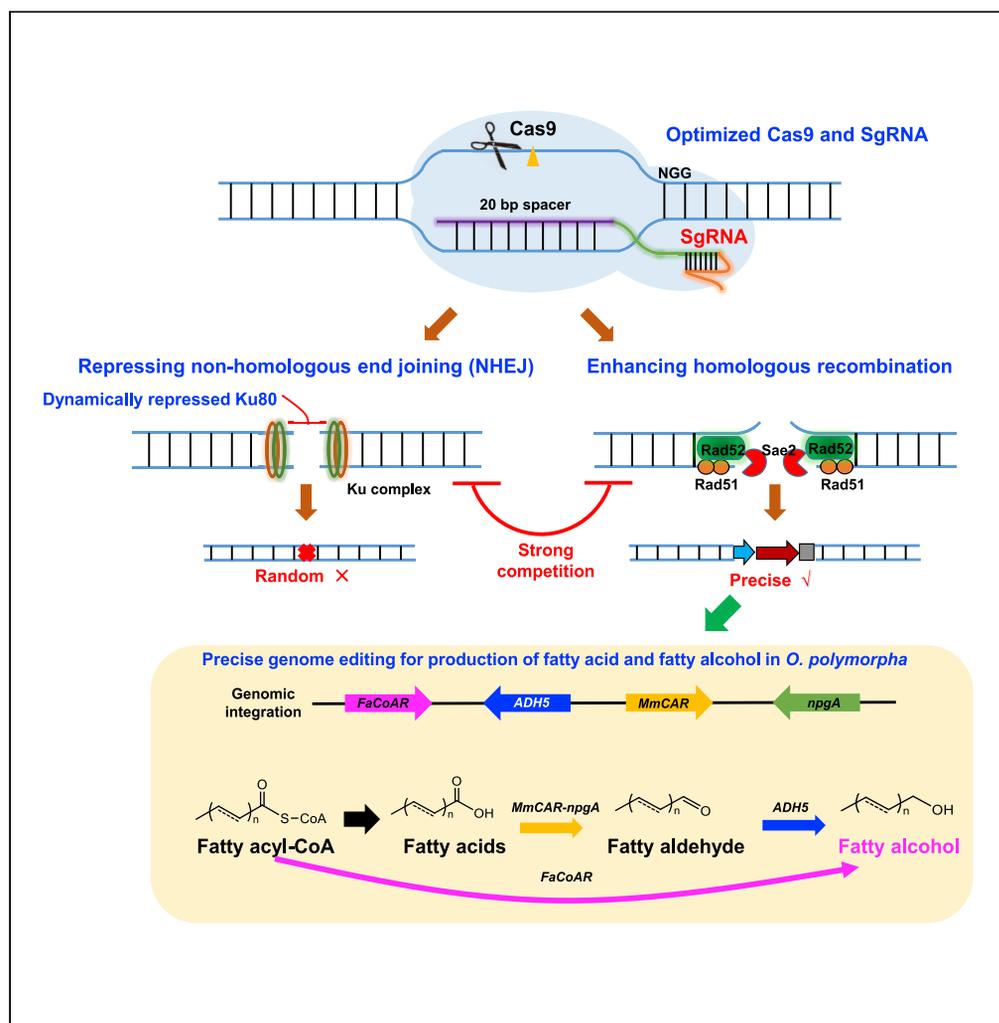


Article

Recombination machinery engineering for precise genome editing in methylotrophic yeast *Ogataea polymorpha*

Jiaoqi Gao, Ning Gao, Xiaoxin Zhai, Yongjin J. Zhou

zhouyongjin@dicp.ac.cn

HIGHLIGHTS

Establishing an efficient and convenient CRISPR-Cas9 system in *Ogataea polymorpha*

Enhancing homologous recombination for precise genome editing in *O. polymorpha*

Realizing seamless deletion and assembly of multiple fragments in *O. polymorpha*

Article

Recombination machinery engineering for precise genome editing in methylotrophic yeast *Ogataea polymorpha*Jiaoqi Gao,^{1,2,3} Ning Gao,^{1,3} Xiaoxin Zhai,^{1,3} and Yongjin J. Zhou^{1,2,3,4,5,*}

SUMMARY

Methanol biotransformation can expand biorefinery substrate spectrum other than biomass by using methylotrophic microbes. *Ogataea (Hansenula) polymorpha*, a representative methylotrophic yeast, attracts much attention due to its thermotolerance, but the low homologous recombination (HR) efficiency hinders its precise genetic manipulation during cell factory construction. Here, recombination machinery engineering (rME) is explored for enhancing HR activity together with establishing an efficient CRISPR-Cas9 system in *O. polymorpha*. Overexpression of HR-related proteins and down-regulation of non-homologous end joining (NHEJ) increased HR rates from 20%–30% to 60%–70%. With these recombination perturbation mutants, a competition between HR and NHEJ is observed. This HR up-regulated system has been applied for homologous integration of large fragments and *in vivo* assembly of multiple fragments, which enables the production of fatty alcohols in *O. polymorpha*. These findings will simplify genetic engineering in non-conventional yeasts and facilitate the adoption of *O. polymorpha* as an attractive cell factory for industrial application.

INTRODUCTION

Single-carbon (C1) feedstocks represent as attractive substrates for future biorefinery owing to their abundance and no-food competition (Clomburg et al., 2017; Zhou et al., 2018). Among these feedstocks, methanol, which can be derived from coal, natural gas, or CO₂, is an ideal substrate for bio-manufacturing owing to its liquid state for efficient mass transfer (Duan and Gao, 2018; Zhou et al., 2018). In nature, there exists a group of microorganisms named “methylotrophs,” which efficiently assimilate methanol for growth. *Ogataea (Hansenula) polymorpha*, as one of methylotrophic yeasts, possesses various advantages in the wide substrate spectrum like xylose and methanol and has extreme thermo-tolerance (over 50°C) (Saraya et al., 2012), which makes it a promising cell factory for protein expression and chemical production (Ubiyovk et al., 2011; Voronovsky et al., 2009). However, similar to other non-conventional yeasts, the difficulties in convenient and precise genome editing in *O. polymorpha* will limit its metabolic engineering toward industrial production (Cai et al., 2019; Schwartz and Wheeldon, 2017).

In recent years, the CRISPR-Cas9 system has been applied for genetic engineering in numerous organisms with high efficiency and accuracy (McGinn and Marraffini, 2018; Raschmanova et al., 2018). Precise genome editing via the CRISPR-Cas9 system depends on Cas9 protein and single guide RNA (gRNA). Cas9 protein is a kind of RNA-mediated endonuclease with two active domains, HNH and RuvC, which cuts off double-stranded DNA with the assistance of nuclear locating signal and gRNA. As another essential component, gRNA is composed of CRISPR-targeting RNA (crRNA) and trans-activating crRNA. A 20-bp spacer in crRNA comes from the targeting sequence with a specific protospacer adjacent motif at the 3' end. Cas9 protein is guided by gRNA with a specific secondary structure to perform double-strand break (DSB) (McGinn and Marraffini, 2018) (Figure 1).

Once the DSB is formed, cells immediately activate the DNA repair process to prevent genomic instability and cell death, which may eventually result in specific genome editing. Two main repair pathways may happen in the DSB repair process, including classical non-homologous end joining (NHEJ) and homologous recombination (HR) (Figure 1). NHEJ initiates by binding of the Ku70-Ku80 heterodimer to the DSB, which recruits NHEJ polymerase, nuclease, and ligase complexes, resulting in multiple rounds of

¹Division of Biotechnology, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, 457 Zhongshan Road, Dalian 116023, PR China

²CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, PR China

³Dalian Key Laboratory of Energy Biotechnology, Dalian Institute of Chemical Physics, CAS, 457 Zhongshan Road, Dalian 116023, PR China

⁴www.synbc.dicp.ac.cn

⁵Lead contact

*Correspondence: zhouyongjin@dicp.ac.cn
<https://doi.org/10.1016/j.isci.2021.102168>



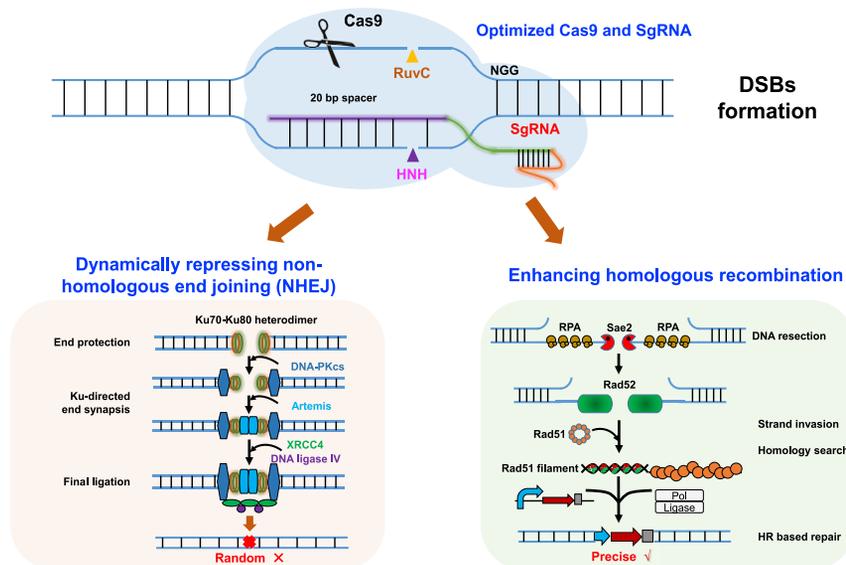


Figure 1. Optimized CRISPR/Cas9 system in *O. polymorpha* with dynamically repressed NHEJ and enhanced HR
Cas9 with the guidance of single guide RNA (sgRNA) efficiently cuts DNA to form DSBs, which are subsequently repaired by either NHEJ or HR. While NHEJ occurs, DNA end is first protected by the Ku70-Ku80 heterodimer, which performs the recruitment of DNA-dependent protein kinase catalytic subunit (DNA-PKcs), Artemis, XRCC4/DNA ligase IV complex, etc. Hence, dynamically down-regulating this key protein Ku80 is supposed to be a reasonable strategy to decrease this error-prone and random repair pathway. On the contrary, HR process starts with DNA resection, and ssDNA is generated by a complex consisting of Mre11, Sae2, etc. Subsequently, with the help of Rad52, multiple HR-related proteins (Rad51) are recruited to form the complex for strand invasion and D-loop formation. At last, through a homology search, donor DNA fragments are precisely integrated into the specific sites by DNA polymerase and ligase. Therefore, overexpressing these HR-related proteins represents another alternative way to enhance HR efficiency in *O. polymorpha*.

nucleotide deletion and insertion. This process is error-prone, uncontrollable, and template independent in a random manner (Chang et al., 2017), which is obviously not suitable for metabolic engineering. On the contrary, HR is supposed to be a preferred DNA repair pattern over NHEJ, as HR may result in the site-specific integration or deletion of a target fragment. The choice and regulation of HR pathway is extremely complicated concerning cell cycle, end resection in DSBs, and multiple HR-related proteins like Rad51 (Ceccaldi et al., 2016).

NHEJ is the dominant repair mechanism in most non-conventional yeasts including *O. polymorpha*, which leads to the relatively low genome editing efficiencies (normally <30%), especially low HR rates in the currently reported CRISPR-Cas9 systems in *O. polymorpha* (Juergens et al., 2018; Numamoto et al., 2017; Wang et al., 2018), and these systems were far behind the requirements for precise pathways engineering and thus hinder extensive genetic engineering for metabolic reprogramming (Weninger et al., 2018; Yu et al., 2018). Here, we constructed and optimized a CRISPR-Cas9 system in *O. polymorpha* to enhance the genome cutting efficiency. Then we enhanced the HR rates and repressed the NHEJ process for precise genetic manipulation. With this endeavor, we successfully established an applicable platform in non-conventional yeast *O. polymorpha*, which enabled scarless gene deletion, *in vivo* assembly, and integration of multiple DNA fragments. In particular, this genome editing platform was applied in engineering *O. polymorpha* for production of fatty alcohols. We anticipate that this efficient genetic engineering system can make this *O. polymorpha* an important workhorse for methanol and biomass-based bio-manufacturing.

RESULTS

Construction and optimization of CRISPR-Cas9 system

To achieve the expression of Cas9 protein, a human codon-optimized CAS9 gene was integrated into the genome by single crossover, which was controlled by the promoter of glyceraldehyde-3-phosphate dehydrogenase (P_{GAP}) and the terminator of alcohol oxidase 1 (T_{AOX1}), both from *Komagataella phaffii*. Strain

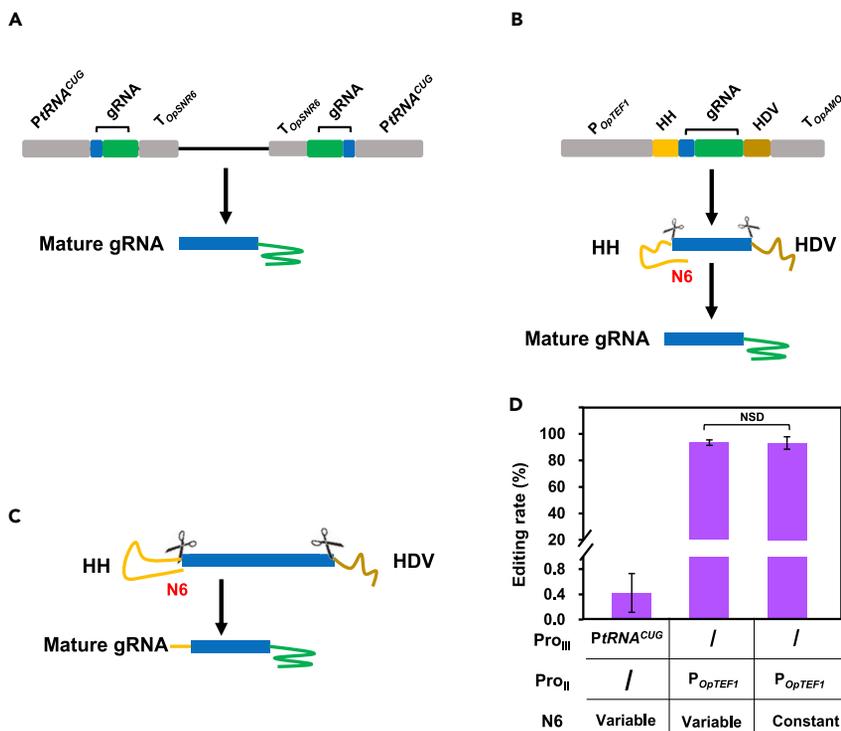


Figure 2. Optimization of gRNA expression for improving genome editing efficiency

(A) Dual direction of gRNA expression by RNA pol III promoter, generating a clean mature gRNA.

(B) gRNA expression by RNA pol II promoter (P_{OpTEF1}), generating a clean mature gRNA with the help of ribozyme HH and HDV.

(C and D) (C) Constant N6 sequence generates a mature gRNA with extra 6 bp of 5' cap; (D) gene editing rate of targeting gene *OpADE2* with different gRNA expression cassettes and N6 sequences. As *OpADE2* disruption resulted in red colony, the editing rates were calculated as percentages of red colonies. Data are presented as mean \pm SEM ($n = 3$ biologically independent samples). Statistical analysis using paired t test showed no significant difference (NSD) between variable and constant N6 sequences.

495-3 with a copy of the integrated *CAS9* gene was verified by RT-PCR (Figure S1A) and had no growth defect compared with the wild-type strain.

An episomal plasmid for gRNA expression was adopted by inserting the autonomous replication start from *Kluyveromyces lactis* (panARS) (Figure S1B) (Liachko and Dunham, 2014). Considering the advantage of RNA pol III promoter ($PtRNA^{CUG}$) in yielding a functional mature gRNA without 5' cap and 3' tail (Figure 2A), a dual direction gRNA expression cassette was constructed with the promoter of $tRNA^{CUG}$ from *O. polymorpha* (Numamoto et al., 2017). However, the mutation rate was less than 0.1% when targeting the gene *OpADE2* in strain 495-3 (Table S1). In this case, the terminator T_{SUP4} from *S. cerevisiae* was replaced by a more effective terminator like $SNR6t$ from *O. polymorpha* (Figures S1B and 2A). Unfortunately, the editing efficiency was still very low (around 0.4%) when targeting the *OpADE2* gene (Table S1 and Figure 2D). To further enhance gRNA expression, a gRNA expression cassette mediated by RNA pol II promoter P_{OpTEF1} (promoter of translation elongation factor EF-1 alpha from *O. polymorpha*) was constructed. This gRNA expression cassette enabled 93.4% editing efficiency for *OpADE2*, which was over 200-fold higher than that mediated by RNA pol III promoter (Table S1 and Figure 2D). Similarly, when targeting the gene *OpKU80*, all eight selected transformants had mutations around gRNA sites by sequencing, achieving a 100% editing efficiency (Table S1).

In the aforementioned system, ribozyme HH contains a variable 6-bp sequence that is reverse complement to the first 6 bp on gRNA sequence (Figure 2B), which makes it difficult and expensive to replace the 20-bp spacer when targeting another gene owing to the requirement of multiple long primers (Figure S1C). Hence, a constant N6 sequence reversely complementing to the last 6-bp sequence in

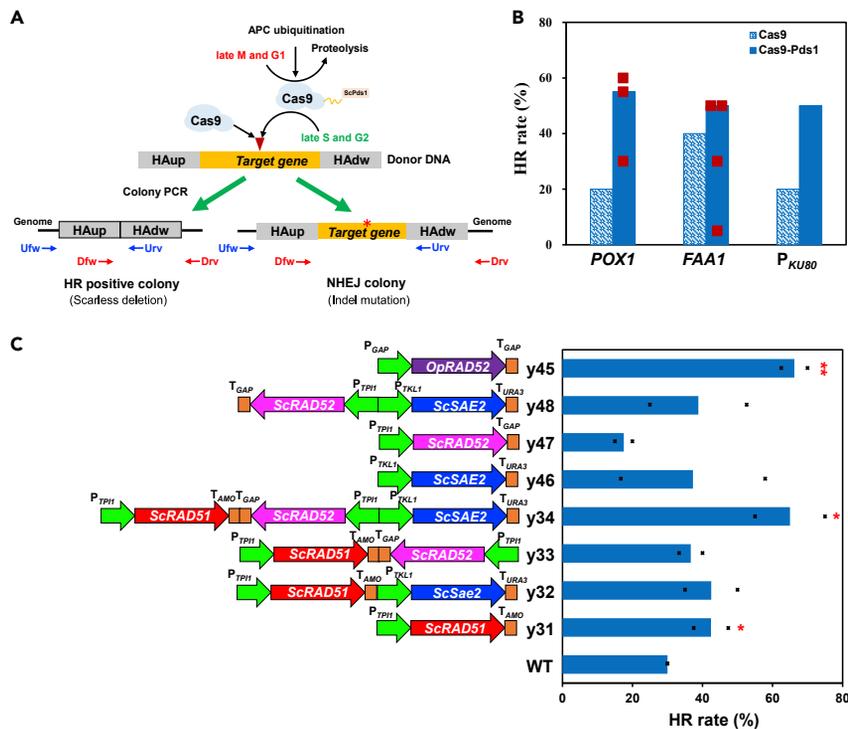


Figure 3. Enhancing HR process promotes the efficiencies of scarless gene deletion

(A) A fusion of Cas9- and ScPds1-promoted HR activity synchronized with cell cycle. Cas9-Pds1 was ubiquitinated and degraded in the M and G1 phases and functional as the normal Cas9 protein in the late S and G2 phases. HR rates were calculated as the percentages of HR clones, which were determined by colony PCR of transformants and showed a shorter DNA band when compared with NHEJ colony. Ufw and Urv and Dfw and Drv are forward and reverse primers from upstream and downstream, respectively, for colony PCR.

(B) The HR rates were calculated while targeting genes *OpPOX1*, *OpFAA1*, and *Ku80* promoter sequence (*P_{KU80}*) in strains expressing Cas9 or Cas9-Pds1. In particular, multiple experiments were performed to prove an unstable behavior in strain Cas9-Pds1 while targeting genes *OpPOX1* and *OpFAA1* (red square).

(C) Overexpressing HR-related proteins improved HR rates for scarless disruption of succinate-CoA ligase encoding gene (*OpLSC2*). Total 20 colony from each biological parallel was picked and tested by colony PCR to calculate HR rate. Data are presented as means of two biologically independent samples with displayed data points. Red asterisks indicate statistical significance as determined using paired t test (* $p < 0.05$; ** $p < 0.01$).

ribozyme HH was adopted for convenient construction of gRNA expression cassettes (Figures 2C and S1C). We then tested the possible negative effect of the constant N6 sequence on genome editing rates by targeting multiple genes (Table S1 and Figure 2D). The mutation efficiency of constant N6 sequence showed no significant differences when compared with that of variable N6 sequence while targeting both *OpADE2* and *OpKU80*. DNA sequencing showed that there were mostly indel mutations (1- or 2-bp deletion, insertion, and mutation) at editing sites and a few random insertions of the large DNA fragment (Figure S2), which suggested that the constant N6 sequence did not affect the guiding efficiency for Cas9 enzyme. Thus the constant N6 sequence was utilized for the further construction of gRNA expression cassettes.

Enhanced HR-mediated DSB repair

Low HR rate has been previously reported in *O. polymorpha*, which seriously hinders its application in extensive metabolic reprogramming (Juergens et al., 2018; Numamoto et al., 2017). We thus tried to enhance HR rate for precise genome engineering. We first explored to synchronize the expression of Cas9 protein from *O. polymorpha* in the late S and G2 phases of cell cycle, when HR activity remains the highest to meet the requirements of sister chromatid synapsis (Gutschner et al., 2016). Thus, an anaphase-promoting complex (APC)-dependent mitotic cyclin protein Pds1 from *S. cerevisiae* was fused to the C terminal of Cas9, which was functional in the S and G2 phases, and recognized, ubiquitinated, and degraded by APC in the M and G1 phases (Figure 3A) (Cohen-Fix et al., 1996). Similar to that in

mammalian cells, Cas9-Pds1 fusion significantly enhanced the scarless gene deletion from 20%–40% to 50%–60% in strain *O. polymorpha* 495-3 (Figure 3B), which suggested that mitotic cyclin protein Pds1 from *S. cerevisiae* played a positive role in promoting the HR rate of *O. polymorpha*. However, unstable promotion of HR rates (5%–60%) was observed for targeting *OpFAA1* and *HpPOX1* in numerous experiments (Figures 3B and S3B), which might be attributed to the complex regulation of cell cycle (Aird et al., 2018).

We then explored other simple and stable approaches to promote HR-mediated repair process. It has been showed that three functional proteins Sae2, Rad52, and Rad51 played the main role in efficient HR in *S. cerevisiae* (Figure 1) (Krejci et al., 2012; Mimitou and Symington, 2008). We thus learned from *S. cerevisiae* with extremely high HR rates and reconstructed its HR system in *O. polymorpha*. HR-related proteins from *S. cerevisiae* demonstrated an obvious promotion in HR rate by 10%–20% in the overexpression strains (Figure 3C). In particular, strain y34 with the combination of *S. cerevisiae* genes ScSAE2, ScRAD52, and ScRAD51 had HR rates of ~70%, when targeting succinate-CoA ligase gene *OpLSC2* (Figure 3C) and the *KU80* promoter P_{KU80} (Figure S4B), which were significantly higher than that of the control strain. Quantitative RT-PCR (qPCR) analysis confirms the functional transcription of these genes and proved that the expression level of these genes was dozens of times higher than that of endogenous gene *OpRAD52* in wild-type (Figure S5A). A moderate expression level balanced the HR rate, the colony-forming units (CFU) per unit of cell (CFU/OD₆₀₀), and cell growth. Despite the lower HR rates, some specific strains (y46, y47, and y48) showed an obviously higher CFU number due to an efficient DSB repair process (Figures S3C and S4C). Subsequently, endogenous *OpRAD51* and *OpRAD52* from *O. polymorpha* were identified and overexpressed, and unfortunately, *OpSAE2* was not successfully identified based on homology search of *O. polymorpha* genome. *OpRAD51* overexpression was lethal, and only *OpRAD52* overexpression strain (y45) was obtained. *OpRAD52* overexpression resulted in a significantly higher CFU/OD₆₀₀ (Figure S3C) and HR rate (Figure 3C) in targeting gene *OpLSC2* due to its extremely higher expression level (Figure S5A). However, targeting the promoter of *KU80* in strain y45 had a fluctuation in HR rates (Figure S4). These data suggested that y34 had the most stable and significant increase in HR rate for genome manipulation.

Down-regulation of non-homologous end joining

NHEJ plays the dominant role in DSB repair in several non-conventional organisms including *O. polymorpha* (Figure S2) (Schwartz and Wheeldon, 2017). Ku heterodimer proteins Ku70 and Ku80 are key components for NHEJ-based DNA repair by binding the DNA DSB ends (Figure 1). Previous studies showed that the disruption of *KU70* or *KU80* repressed NHEJ and resulted in a relatively higher HR-mediated DSB repair (Juergens et al., 2018; Kretzschmar et al., 2013). Here we also showed that *KU80* disruption significantly improved the relative HR-mediated DSB repair to almost 100% (Figure 4E). However, this *KU80* disruption seriously reduced the CFU number (Figure 4D) and slightly retarded the cell growth (Figure 4B), which suggested that repressing Ku heterodimer proteins caused stress on cellular fitness. Alternatively, we dynamically repressed *KU80* by replacing its native promoter with a responsive promoter of *MET3* gene (P_{OpMET3}) that was repressed by methionine (Figure S6A) (Yoo et al., 2015). With this system, *KU80* can be conditionally repressed during genetic manipulation by adding methionine in culture media and/or selection plates and de-repressed in bio-production conditions (Figure 4A). Expression level of the gene *KU80* was obviously down-regulated in strain Ku80-dw with methionine (Figure S5B), which, however, was not as much lower as the reported level (10%–20%) (Yoo et al., 2015). Optimization of methionine concentrations showed that 1.7 mM methionine was enough to down-regulate Ku80 for enhancing HR rate in rich medium like YPD (Figures S6B and S6C), and a higher methionine concentration was recommended to compensate the possible consumption and metabolism during cultivation in basic medium (Figure S6D).

The dynamic down-regulation of *KU80* (strain Ku80-dw) had a much higher CFU number when compared with that of *KU80* disruption (strain Δ Ku80) and was comparable with that of the wild-type strain (Figure 4D). Furthermore, strain Ku80-dw grew faster than the wild-type strain at the early log phase, whereas strain Ku80 Δ had poorer growth in the late stage of log and stable phases (the differences were significant, Figure 4B). In total, this dynamic down-regulation of Ku80 enabled an over 3-fold higher HR rate (~60%) with the highest positive clones (Figure 4C) when compared with the wild-type strain.

We then investigated whether the combination of overexpression of HR-related genes with down-regulation of *KU80* (Ku80-dw) could further improve HR rates (Figures 4D and 4E). Additional overexpression of

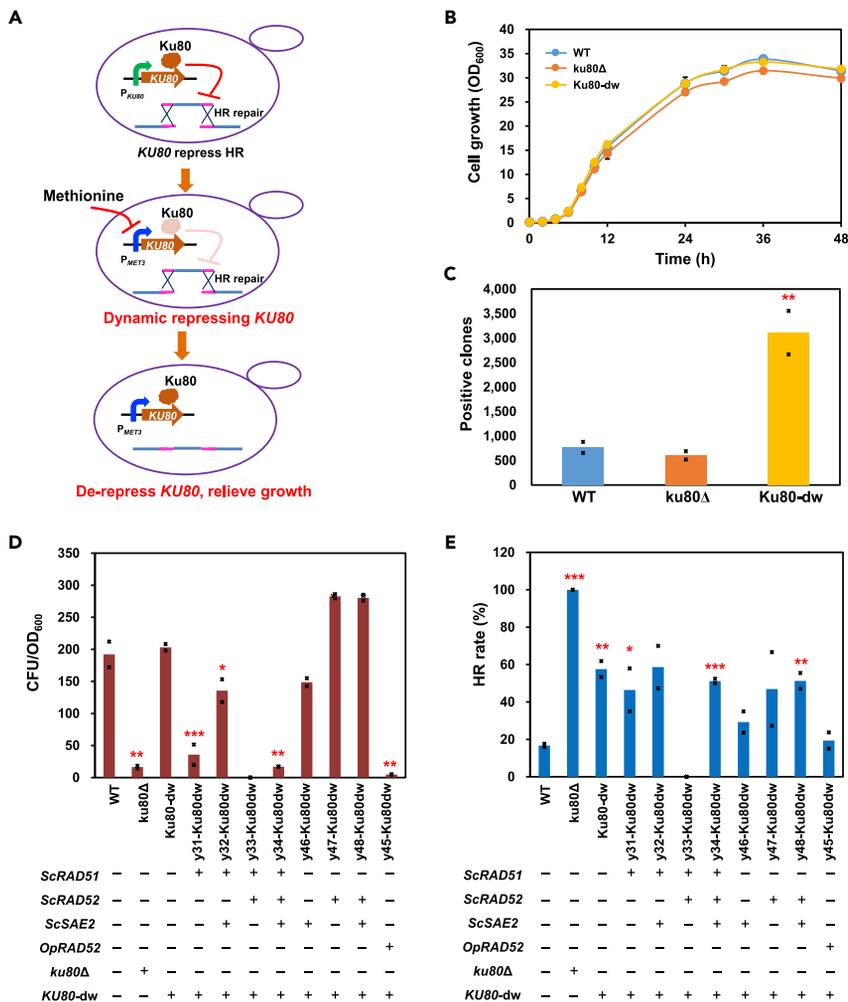


Figure 4. Down-regulation of KU80 enhanced HR rates by repressing NHEJ process

(A) *KU80* expression under the control of P_{OpMET3} promoter was down-regulated when methionine existed. A down-regulated *KU80* prevented Ku complex formation, which eventually decreased NHEJ activity, and the repressed NHEJ process forced cells to select HR to repair DNA DSB for survival. Without methionine supplementation, the *KU80* repression will be removed and enabled the normal cell growth during production condition. (B–E) (B) Cell growth behaviors in strains wild-type (WT), *KU80* disruption (*ku80Δ*), and *KU80* down-regulation (*Ku80-dw*). Data are presented as mean \pm SEM ($n = 2$ biologically independent samples). (C) Positive clones (positive rate \times transformant number) targeting gene *POX1* in strains WT, *ku80Δ*, and *Ku80-dw*. CFU/OD₆₀₀ (D) and HR rate (E) in strains with both overexpressed HR-related proteins and down-regulated *KU80* expression, when targeting gene *OpLSC2*. Total 20 colonies from each biological parallel were picked and tested by colony PCR to calculate the HR rate. Data are presented as means of two biologically independent samples with displayed data points. Red asterisks indicate statistical significance as determined using paired t test (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).

ScRAD51 with down-regulation of *KU80* resulted in significantly lower CFU/OD₆₀₀ (strain y31-Ku80dw, y32-Ku80dw, y33-Ku80dw, and y34-Ku80dw versus *Ku80-dw*). Overexpression of ScRAD52 with down-regulation of *KU80* increased CFU/OD₆₀₀, but had a marginal effect on the HR activity. Interestingly, combined overexpression of endogenous *OpRAD52* with *KU80* down-regulation resulted in much lower CFU/OD₆₀₀ and HR rate, which suggested that an extensively regulated endogenous DSB repair system brought severe cellular stress. These data clearly showed that a combination of two strategies did not further promote the HR rate in *O. polymorpha*. Considering the balance of HR rates and CFU/OD₆₀₀ numbers, y34 (overexpression of ScSAE2, ScRAD51, and ScRAD52) and *Ku80-dw* strain had similar HR rates of 60%–70% (Figures 3C and 4E).

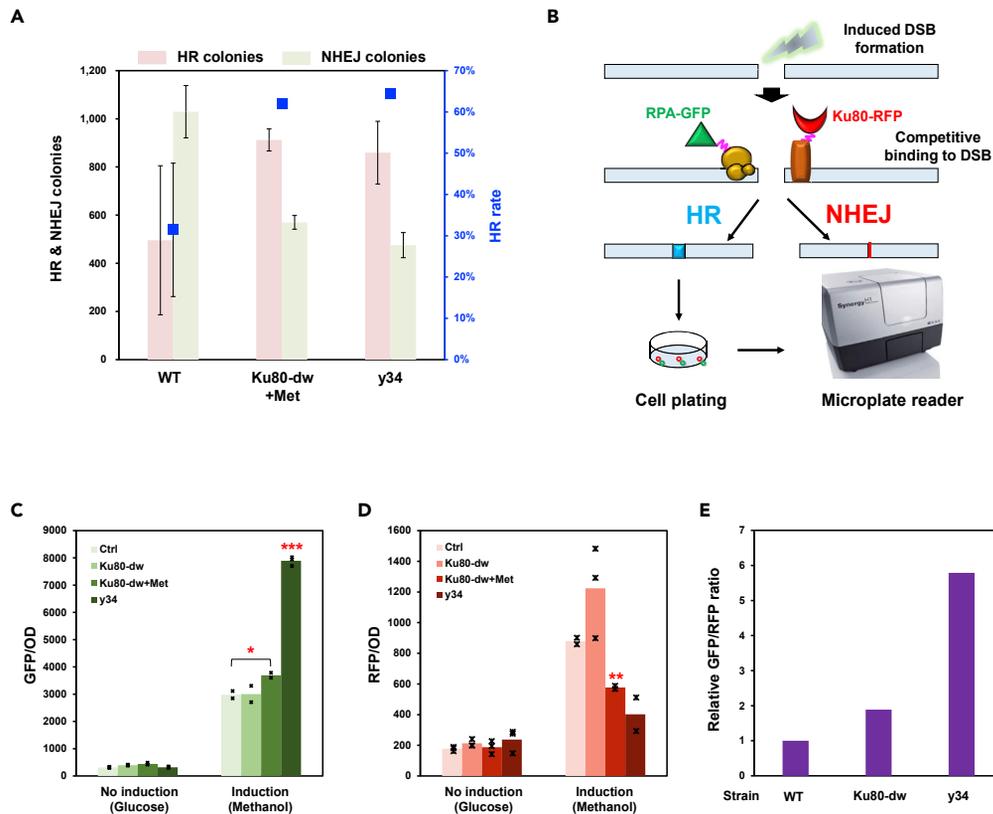


Figure 5. A competitive binding to DSBs leads to the change in HR/NHEJ ratio

(A) HR colonies and NHEJ colonies in strains of wild-type (WT), *KU80* down-regulation (Ku80-dw under repression of methionine), and overexpression of HR-related proteins (y34), which were calculated by multiplication of total colonies and HR rates (blue square).

(B) Schematic illustration of a competitive mechanism of HR (RPA) and NHEJ (Ku80). For controlling DSBs formation well, an inducible gRNA plasmid targeting *LSC2* gene was constructed by using a methanol-induced promoter P_{DAS1} . The relative abundance of RPA-GFP and Ku80-RFP was detected by fluorescence intensity.

(C) Fluorescence intensities of GFP (B) and RFP (C) were measured at 24 h in strains WT, Ku80-dw, and y34. In particular, 1.7 mM methionine was added to repress the Ku80 expression in strain Ku80-dw (Ku80-dw + Met). Cells were cultivated in Delft basic salt media containing 10 g/L methanol (Induction), or 20 g/L glucose (No induction), at 37°C, 220 rpm.

(D) The relative abundance of RPA and Ku80, which was calculated by the ratio of GFP fluorescence intensities and RFP fluorescence intensities, was highly consistent with the corresponding HR rates. Data are presented as means of three biologically independent samples with displayed data points. Red asterisks indicate statistical significance as determined using paired t test (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).

Competitive binding of RPA and Ku80 to DSB sites

Down-regulation of *KU80* (strain Ku80-dw) and overexpression of *RADs* (strain y34) both had more HR colonies and higher HR rates when compared with the wild-type strain (Figure 5A), which might be owing to a competitive binding of relating proteins in HR process and NHEJ pathway to DSB sites (O'Driscoll and Jeggo, 2006). Hence, the initial proteins RPA and Ku80 were selected to roughly profile the relative binding efficiencies to DSB sites (Heyer et al., 2010; Krejci et al., 2012). We here fused green fluorescent protein to the main component protein RPA (RPA-GFP) of the HR process and red fluorescent protein to Ku80 (Ku80-RFP) of the NHEJ pathway. After the induction of DNA damage at the *OpLSC2* site, the intensities of fluorescence of GFP and RFP were both detected at 24 h by a microplate reader (Figure 5B).

In methanol media (for inducing DSB formation), de-regulation of *KU80* (strain Ku80-dw with methionine) improved the expression of the key protein RPA (RPA-GFP), which suggested repression of Ku80 (Figure 5C). Vice versa, enhancing the expression of HR-related proteins (strain y34) repressed the expression

of NHEJ-related protein Ku80 (Figure 5D). Finally, RPA/Ku80 ratio demonstrated a relative binding efficiency to DSB sites, which eventually showed the relative strength of HR or NHEJ. The RPA/Ku80 ratio in strain Ku80-dw was double that of the control strain, and strain y34 had the highest RPA/Ku80 ratio, which was in line with the HR rates (Figure 5A) and suggested that y34 had extremely strong HR activity. Meanwhile, no fluorescence signal was detected in all strains growing in glucose media, which verified that the DSBs can only be created by methanol induction (Figures 5C and 5D). In particular, we counted 50,000 cells to calculate the relative expression level of RPA via flow cytometer (Figure S7). When setting the GFP fluorescence intensity >500 as HR-positive cells (because the GFP fluorescence intensity of y34 is around 500), the positive cells of control, Ku80-dw, and y34 accounted for 28.8%, 35.1%, 59.3%, respectively, which was in agreement with Figure 5. These data suggested that, through the competition of binding sites at DSB sites, the overexpressing HR-related proteins promoted HR-based DNA repair, which in turn decreased the NHEJ strength, and down-regulation of gene *KU80* increased DSB repair via the HR pathway by decreasing the NHEJ strength, which was also highly consistent with qPCR results (Figure S5).

Application of CRISPR-Cas9 toolkit in genetic engineering

We first applied this CRISPR-Cas9 toolkit for scarless deletion of *OpLSC2* gene and determined the minimal homology arm (HA) lengths for efficient HR in Ku80-dw and y34 strains (Figure 6A). 1,000-bp HA resulted in similar gene deletion efficiency (60%–70%) in these two strains, which were both significantly higher than that of the wild-type strain. However, HA lengths of 200 bp and 500 bp led to a dramatic decrease in HR rates in strain Ku80-dw, which was almost in line with those in the control strain, whereas strain y34 continued to have higher HR rate (~40%) even with 200 bp HA. These data indicated that the strain y34 overexpressing HR-related proteins was best for further precise and convenient genome editing (Figures 6B and 6C).

We then tested the possibility of strain y34 in site-specific integration of a large fragment. An expression cassette of gene *ScIDP2* (5.5 kb), harboring *FAA1* targeting HA with lengths of 200 bp, 500 bp, or 1,000 bp (Figure 6D), were transformed into strain y34. 1,000-bp HA enabled sufficient integration efficiency of 40%–50%, whereas shorter HA lengths resulted in lower positive rates (20% for 500 bp, 8.3% for 200 bp, Figures 6E and 6F). It should be mentioned that LiAc/ssDNA chemical transformation led to low integration efficiency (10%–20%) (Figure S8), which suggested that high transformation efficiency was required to uptake sufficient DNA fragments for HR-mediated integration.

Extensive metabolic engineering involves the construction and optimization of long biosynthetic pathways with multiple genes, which thus relies on rapid assembly of multiple DNA fragments in plasmids or genome (Shao et al., 2008). We thus explored the possibility of *in vivo* assembly of a 20-kb plasmid harboring a fatty alcohol biosynthetic pathway (Figure 6G). The plasmid was divided into 3–5 parts, and the HA lengths were set as about 500 bp. Although much lower than gene deletion, the CFU/OD₆₀₀ numbers were similar among assemblies of 3 parts, 4 parts, and 5 parts (Figure 6H). Furthermore, considerably positive rates (up to 20%) were obtained (Figure 6I), which provided a convenient strategy for *in vivo* assembly of a large plasmid with multiple genes and would be helpful for pathway optimization as it was done in *S. cerevisiae* (Zhou et al., 2012, 2016b).

We finally applied this genetic platform for genome integration of biosynthetic pathways with multiple genes, which is considered to be stable in cell factory construction. A previous optimized fatty alcohol biosynthetic pathway (Zhou et al., 2016b) was integrated to the *POX1* site in *O. polymorpha* y34 (Figures S9 and 7A). There were similar numbers of CFU/OD₆₀₀ in spite of a slight decrease while integrating over one cassette (Figure 7B). As expected, a sharp decline in positive rates was observed with the increased number of integrated cassettes (Figure 7C). When integrating only one cassette containing gene *MmCAR*, the positive rate reached 70%, which was in line with previous integration of a large fragment (Figure 6E). However, *in vivo* self-assembly of four expression cassettes reduced the positive rate to less than 10% (Figure 7C).

HFD1 gene in the aforementioned correct transformants with specific genes for fatty alcohol production were further disrupted, which was very essential for the production of fatty aldehyde-derived chemicals (Zhou et al., 2016b). Cassette integration had a marginal effect on cell growth (Figure S10A). Genome integration of fatty acid reductase gene *MmCAR* and its cofactor gene *npgA* enabled fatty alcohol biosynthesis in *O. polymorpha*. *FacoAR*, encoding fatty acyl-CoA reductase, was also beneficial for fatty

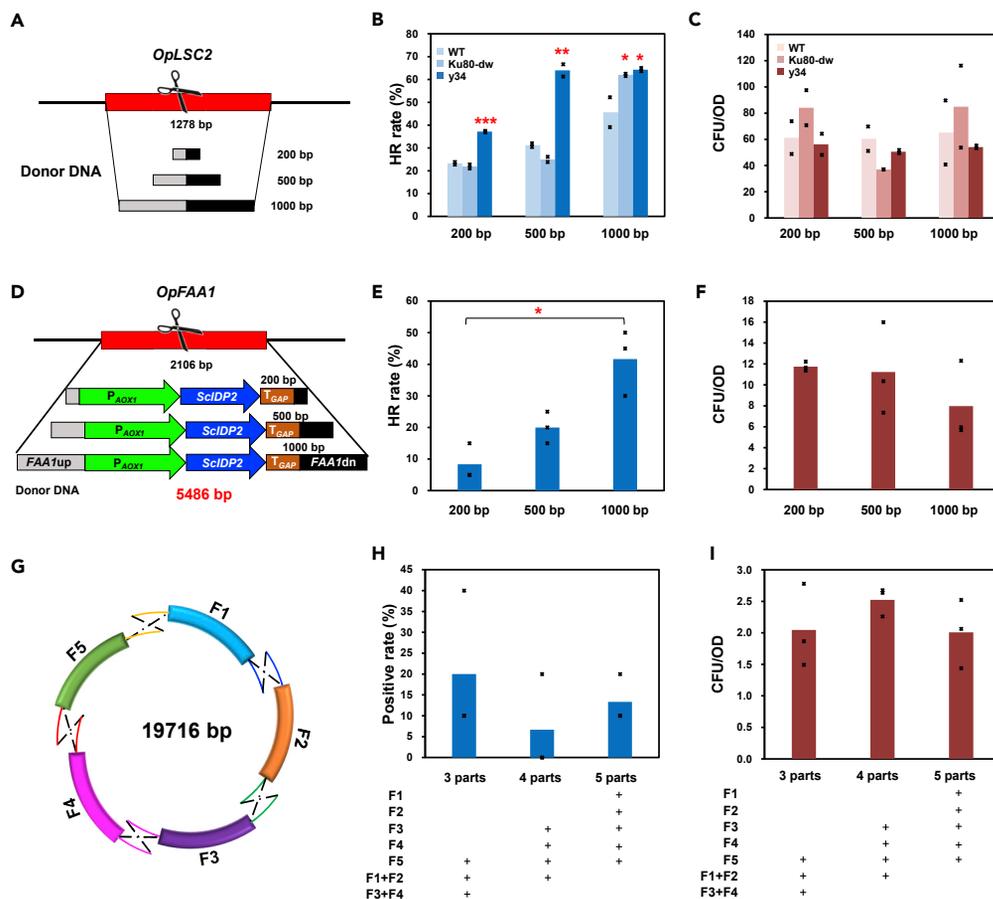


Figure 6. Genetic engineering with enhanced HR process and optimized CRISPR/Cas9 toolkit

(A) Schematic illustration of gene scarless deletion with HA lengths of 200, 500, and 1,000 bp, respectively.
 (B) HR rates for gene deletion with various HA lengths, in strains wild-type (WT), *Ku80* down-regulation (Ku80-dw), and overexpression of HR-related genes (*y34*).
 (C) CFU/OD₆₀₀ for gene deletion with various HA lengths, in strains wild-type (WT), Ku80-dw, and *y34*.
 (D) Schematic illustration of homologous integration of a large fragment (*ScIDP2* cassette at *OpFAA1* locus) with HA of length of 200, 500, and 1,000 bp, respectively.
 (E) HR rates for *ScIDP2* integration with various HA lengths in strain *y34*.
 (F) CFU/OD₆₀₀ of *ScIDP2* integration with various HA lengths in strain *y34*.
 (G) Schematic illustration of *in vivo* self-assembly of a plasmid with size 20 kb, with HA length 500 bp.
 (H) Positive rates of plasmid assembly via 3 parts, 4 parts, and 5 parts, respectively, in strain *y34*.
 (I) CFU/OD₆₀₀ of plasmid assembly via 3 parts, 4 parts, and 5 parts, respectively, in strain *y34*. Total 20 colonies from each biological parallel were picked and tested by colony PCR to calculate HR rate. Data are presented as means of two (A–C) or three (D–I) biologically independent samples. Red asterisks indicate statistical significance as determined using paired *t* test (**p* < 0.05; ***p* < 0.01; ****p* < 0.001), and no significant differences were unmarked.

alcohol biosynthesis (Figure 7D), which was consistent with that in *S. cerevisiae* (Zhou et al., 2016a, 2016b). It was interesting that the overexpression of *S. cerevisiae ADH5* gene encoding alcohol dehydrogenase resulted in a 2.5-fold higher fatty alcohol production (strain C1-2-1 versus C1-3-12 in Figure 7D), which suggested that the alcohol dehydrogenase or aldehyde reductase were not comparable with the alcohol fermentation yeast *S. cerevisiae* and need to be enhanced for fatty alcohol production. To enhance the precursor fatty acid supply, gene *FAA1* was disrupted, which eventually resulted in a dramatic increase in fatty alcohol titer to around 12 mg/L (Figure 7E). Finally, we showed that this strain C1-3 could produce 0.62 ± 0.01 mg/L and 3.26 ± 0.51 mg/L fatty alcohols from methanol when cultivated in basic salt and rich media, respectively (Figure 7F). Hence, the production of fatty acid-derived

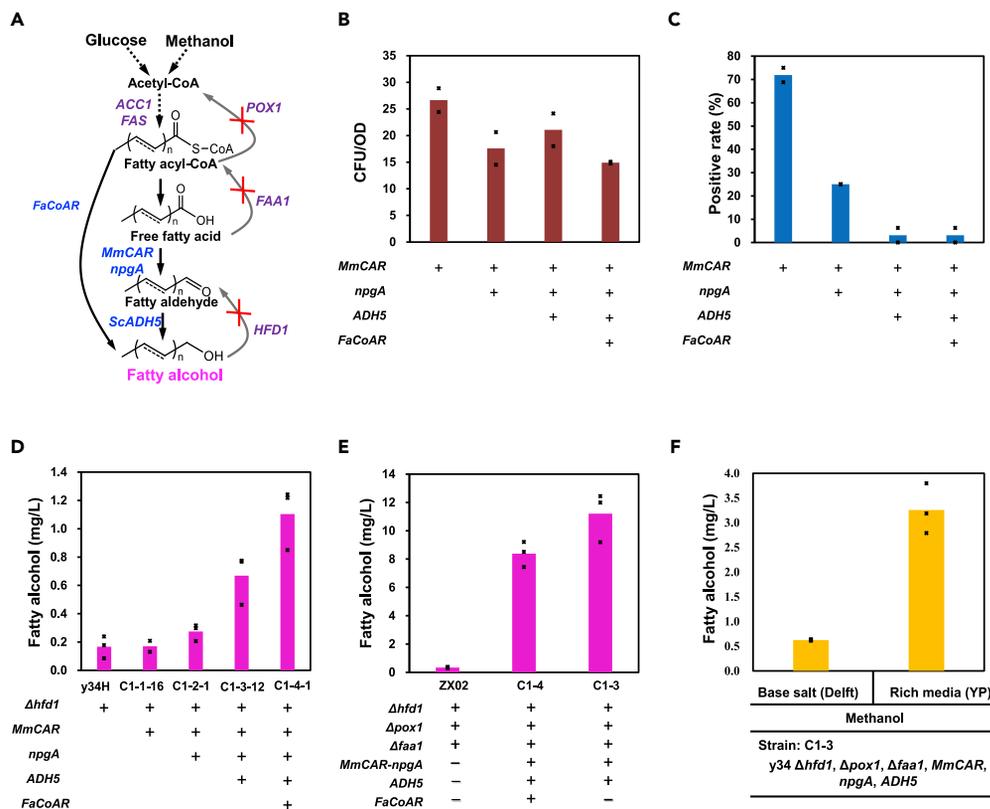


Figure 7. Genomic integration of a biosynthetic pathway for fatty alcohol production

(A) Biosynthetic pathway for fatty alcohol production. Dual deletion of *HFD1* and *POX1* is helpful for fatty alcohol biosynthesis in *O. polymorpha*. Overexpression of carboxylic acid reductase from *Mycobacterium marinum* (*MmCAR*) and its co-factor *npgA*, alcohol dehydrogenase from *S. cerevisiae* (*ADH5*), and fatty acyl-CoA reductase (*FaCoAR*) from *Marinobacter aquaeolei* VT8 (Zhou et al., 2016a, 2016b), was supposed to achieve the production of fatty alcohol. *In vivo* pathway assembly for fatty alcohol production was carried out by integrating multiple expression cassettes. (B and C) The CFU/OD₆₀₀ (B) and positive rate (C) were both calculated with integration of different numbers of genes. (D) The engineered strains were cultivated in basic salt media that containing 20 g/L glucose. (E) Disrupted gene *FAA1* further increased fatty alcohol production from glucose in basic salt media. (F) Fatty alcohol production from methanol in basic salt media or YP media; 5 g/L methanol was supplemented at 24 and 48 h. Cells were cultured for 96 h, at 37°C, 220 rpm. Total 20 colonies from each biological parallel were picked and tested by colony PCR to calculate HR rate. Data are presented as mean of two (B and C), or three (D–F) biologically independent samples.

chemicals from methanol (Figure S10) showed the great potential of *O. polymorpha* as a chassis for methanol-based bio-refinery.

DISCUSSION

Engineering methylotrophic microbes for methanol biotransformation provides a great opportunity in expanding the bio-manufacturing blueprint other than biomass-derived bio-refinery (Duan and Gao, 2018; Zhou et al., 2018). *O. polymorpha* is such a eukaryotic microbe that has been recognized as an attractive host for protein expression and ethanol fermentation due to its thermotolerance and methanol assimilation (Manfrão-Netto et al., 2019; Olson et al., 2015). However, the lack of genetic tools, especially the low HR activity, makes it challenging in expanding its product portfolio other than proteins and ethanol by engineering cellular metabolism. Here we thus established an efficient genetic platform in *O. polymorpha* by optimizing the CRISPR-Cas9 system and recombination machinery engineering (rME). We demonstrated the possible regulation of HR activity and NHEJ strength in this yeast, and the precise genome editing in the engineered strain may further promote its application as a cell factory for the biosynthesis of valuable products from methanol.

Our CRISPR-Cas9 system, composed of an integrated Cas9 protein and the optimized episomal gRNA expression vector, enabled the highest genome editing efficiency in *O. polymorpha* so far (Juergens et al., 2018; Numamoto et al., 2017; Wang et al., 2018). This system showed good convenience and time-saving pattern for extensive metabolic engineering due to the adoption of RNA pol II promoter with the aid of ribozyme and constant N6 sequence. Besides, we found that functional gRNA expression is very essential for guiding Cas9 toward targeting sites. In spite of the extensive adoption in *S. cerevisiae* and many other non-conventional yeasts (Cao et al., 2017; Horwitz et al., 2015; Mitsui et al., 2019; Nambu-Nishida et al., 2017; Schwartz et al., 2015) (Table 1), RNA pol III promoter failed to drive efficient gRNA expression, or no suitable RNA pol III promoter has been identified in *O. polymorpha* so far.

Predictable and controllable genetic engineering is very essential for extensive metabolic rewiring in the construction of robust cell factories (Chen et al., 2020). Indeed, the high HR rate in *S. cerevisiae* is considered to be one of the main advantages as a preferred host for producing a variety of chemicals, because short homology arms are sufficient to bring nearly 100% targeted repair (Mitsui et al., 2019). However, the relatively high NHEJ always in non-conventional yeasts such as *O. polymorpha* retards the precise genetic manipulation as it repairs DSB in an unpredictable manner (Saraya et al., 2012). Thus we tried to enhance the HR rate with several strategies (Figure 1). Although coordination of Cas9 expression with the high HR activity cell phase (Gutschner et al., 2016; Yang et al., 2016) enabled doubling of HR-mediated genome editing as reported in other studies (Gutschner et al., 2016), the difficulty in controlling cell cycles led to the instability of HR rate. Alternatively, overexpression of HR-related proteins in strain y34 significantly promoted HR-mediated DNA repair process, and well-balanced HR rate, CFU number per OD₆₀₀, and cell growth owing to a moderate expression level. These lessons told us that an appropriate level and even an induced, or dynamically regulated, system could be optimized for further enhancing the HR efficiency. Some similar results were also observed in mammalian cells with the overexpression of *RAD51*, *RAD 52*, or *CtIP* (Arjun et al., 2010; Charpentier et al., 2018; Di et al., 2005; Jayathilaka et al., 2008; Johnson et al., 1996; Shao et al., 2017; Vispé et al., 1998; Yáñez and Porter, 1999), which again proved that enhancing the expression of HR-related proteins effectively promoted HR-mediated DNA repair.

As it might compete with HR, we thus tried to repress the NHEJ by down-regulating the relating Ku heterodimer protein Ku80. As described in previous reports (Juergens et al., 2018; Kretzschmar et al., 2013), deletion of *KU80* gene retarded cell growth and decreased CFU/OD₆₀₀. Here, we first demonstrated a dynamically repressed CRISPR-Cas9 system in *O. polymorpha* with a methionine-repressed promoter P_{OPMET3}, which guaranteed the highest positive clones without retarding cell growth. This system might be extended as a general strategy in non-conventional yeasts (Schwartz et al., 2017). Interestingly, combining the overexpression of HR-related protein with the *KU80* repression did not further improve the HR rate, suggesting that HR and NHEJ behaved competitively in the DSB repair process. This phenomenon could be roughly explained as the competition of binding site between HR proteins and Ku heterodimer proteins. Although down-regulation of *KU80* (strain Ku80-dw) significantly improved HR efficiency, strain y34 with overexpressing HR-relating genes had better and stable performance when using shorter HA lengths and conducting complex genetic manipulation such as integration of large and multiple fragments.

This *O. polymorpha* optimized CRISPR-Cas9 system with enhanced HR rate significantly facilitated metabolic engineering to clearly realize the scareless gene deletion, genome integration of large fragment even with a short HA of 200 bp, and *in vivo* assembly of episomal plasmid with a large size up to 20 kb. Besides, we provide a feasible approach for the construction of genetically stable *O. polymorpha* for industrial process by genome integration of up to four gene expression cassettes, and the production of fatty acid-derived chemicals in *O. polymorpha* from both glucose and methanol were also achieved. Despite a lower titer (Cordova et al., 2020; D'Espaux et al., 2017; Liu et al., 2020; Mcneil and Stuart, 2017), we can expect that further engineering the fatty acid metabolism and methanol utilization would enhance fatty alcohol production as it has been done in *S. cerevisiae*.

Limitations of the study

In this study, the recombination machinery has been systematically engineered to establish an efficient and convenient CRISPR-Cas9 system with an enhanced HR rate in *O. polymorpha*. Yet, an inducible and dynamic system should be constructed to avoid any unpredictable growth defect in more harsh conditions.

Table 1. Genome editing by CRISPR-Cas9 in industrially important yeasts

| Host | Cas9 | | sgRNA | | Editing rates | HR | | | Minimal HA | Reference |
|----------------------------------|--------------------------|------------|-----------------------------|------------|----------------------|---------------------|-------------|---------------|------------|------------------------------------|
| | Promoter | Type | Promoter | Type | | Deletion | Integration | Marker needed | | |
| <i>S. cerevisiae</i> | <i>S. pyogenes</i> Cas9 | | RNA pol III promoter | | Nearly 100% | | | No | 50 bp | (Mitsui et al., 2019) |
| <i>Scheffersomyces stipitis</i> | P _{ENO1} | Episomal | P _{SNR52} | Episomal | 83%–100% | – | | | – | (Cao et al., 2017) |
| <i>K. phaffii</i> | P _{HTA1} | Episomal | P _{HTB1} | Episomal | 43%–95% | 2.4% | 24% | Yes | 1,000 bp | (Weninger et al., 2016) |
| <i>K. lactis</i> | P _{FBA1} | Integrated | P _{SNR52} | Episomal | – | – | 41%–55% | No | 500 bp | (Horwitz et al., 2015) |
| <i>Kluyveromyces marxianus</i> | P _{ScPDC1} | Episomal | P _{ScSNR52} | Episomal | – | – | 28% | No | 50 bp | (Nambu-Nishida et al., 2017) |
| <i>Yarrowia lipolytica</i> | P _{UAS1B8-TEF1} | Episomal | P _{SCR1'-tRNAGly} | Episomal | ^c 54%/92% | – | 16%–73% | No/Yes | 1,000 bp | (Schwartz et al., 2015) |
| <i>Ogataea thermomethanolica</i> | P _{AOX1} | Episomal | P _{AOX1} | Episomal | 63%–97% | – | | | | (Phithakrotchanakoon et al., 2018) |
| <i>O. parapolyomorpha</i> | P _{AaTEF1} | Episomal | P _{ScTDH3} | Episomal | ^a 0%/63% | ^b 0%/<1% | – | No | 500 bp | (Juergens et al., 2018) |
| <i>O. polymorpha</i> | P _{OpTDH3} | Episomal | P _{OpSNR6-tRNACUG} | Episomal | 17%–71% | 47% | – | Yes | 60 bp | (Numamoto et al., 2017) |
| <i>O. polymorpha</i> | P _{AaTEF1} | Episomal | P _{ScTDH3} | Episomal | ^a 0%/9% | – | – | – | – | (Juergens et al., 2018) |
| <i>O. polymorpha</i> | P _{ScTEF1} | Integrated | P _{ScSNR52} | Integrated | – | 58%–65% | 62%–66% | Yes | 500 bp | (Wang et al., 2018) |
| <i>O. polymorpha</i> | P _{KpGAP} | Integrated | P _{TEF1} | Episomal | 90%–95% | 60%–70% | 40%–70% | No | 200 bp | This study |

^aUpon transformation, none of the transformants exhibited mutated genotype; higher editing rate needed further 192 h of incubation in selective medium.

^bNo transformants showed repair pattern via HR in wild-type, and *KU80* disruption resulted in 7 of 1,900 transformants with a scarless gene deletion in *O. parapolyomorpha* DL-1.

^cEditing rates were calculated after 2 and 4 days of outgrowth in selective liquid media.

In particular, more strictly repressed promoter needs further investigation to down-regulate gene *KU80*. Moreover, an explicit mechanism on the competition of HR and NHEJ remains elusive for a comprehensive regulation in further applications. Finally, a fine regulation must be performed to significantly increase fatty alcohol production from both glucose and methanol.

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Yongjin J. Zhou (zhouyongjin@dicp.ac.cn).

Materials availability

gRNA plasmids (pHpgrRNA13 and pHpgrRNA50) generated in this study have been deposited to Addgene (Yongjin Zhou, 78587). Other materials generated in this study are available upon request from the Lead Contact with a completed Materials Transfer Agreement.

Data and code availability

The published article includes all datasets generated or analyzed during this study.

METHODS

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2021.102168>.

ACKNOWLEDGMENTS

The work was financially supported by National Natural Science Foundation of China (21808216 and 21922812), Dalian Science and Technology Innovation Funding (2019J12GX030), DMTO research grant (grant no. DICP DMTO201701), and BioChE research grant (grant no. DICP BioChE-X201801) from Dalian Institute of Chemical Physics (DICP), CAS. The authors thank the Energy Biotechnology Platform of DICP for providing facility assistance.

AUTHOR CONTRIBUTIONS

J.G. designed the research, performed most of the experiments, collected data, and prepared the manuscript. N.G. carried out partial transformation experiments and conducted analysis. X.Z. conducted the fermentation experiments for fatty alcohol production. Y.J.Z. conceived the concept, designed the experiment, and drafted the manuscript.

DECLARATION OF INTERESTS

This work has been included in patent applications in Chinese (202010628649.3 and 202010626783.X) by Dalian Institute of Chemical Physics, CAS.

Received: October 17, 2020

Revised: December 20, 2020

Accepted: February 5, 2021

Published: March 19, 2021

REFERENCES

- Aird, E.J., Lovendahl, K.N., Martin, A.S., Harris, R.S., and Gordon, W.R. (2018). Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. *Commun. Biol.* 1, 54.
- Arjun, K., Giuseppe, R., Di, P.C., Vania, L., Arturo, F., and Alvaro, G. (2010). Enhancement of gene targeting in human cells by intranuclear permeation of the *Saccharomyces cerevisiae* Rad52 protein. *Nucleic Acids Res.* 38, e149.
- Cai, P., Gao, J., and Zhou, Y. (2019). CRISPR-mediated genome editing in non-conventional yeasts for biotechnological applications. *Microb. Cell Fact.* 18, 63.
- 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.
- Ceccaldi, R., Rondinelli, B., and D'Andrea, A.D. (2016). Repair pathway choices and

consequences at the double-strand break. *Trends Cell Biol.* 26, 52–64.

Chang, H.H., Pannunzio, N.R., Adachi, N., and Lieber, M.R. (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Bio* 18, 495.

Charpentier, M., Khedher, A., Menoret, S., Brion, A., Lamribet, K., Dardillac, E., Boix, C., Perrouault, L., Tesson, L., and Geny, S. (2018). CtlP fusion to Cas9 enhances transgene integration by homology-dependent repair. *Nat. Commun.* 9, 1133.

Chen, R., Yang, S., Zhang, L., and Zhou, Y.J. (2020). Advanced strategies for production of natural products in yeast. *iScience* 23, 100879.

Clomburg, J.M., Crumbley, A.M., and Gonzalez, R. (2017). Industrial biomanufacturing: the future of chemical production. *Science* 355, aag0804.

Cohen-Fix, O., Peters, J.-M., Kirschner, M.W., and Koshland, D. (1996). Anaphase initiation in *Saccharomyces cerevisiae* is controlled by the APC-dependent degradation of the anaphase inhibitor Pds1p. *Gene Dev.* 10, 3081–3093.

Cordova, L.T., Butler, J., and Alper, H.S. (2020). Direct production of fatty alcohols from glucose using engineered strains of *Yarrowia lipolytica*. *Metab. Eng. Comm.* 10, e00105.

D'Espaux, L., Ghosh, A., Runguphan, W., Wehrs, M., Xu, F., Konzock, O., Dev, I., Nhan, M., Gin, J., and Reider Apel, A. (2017). Engineering high-level production of fatty alcohols by *Saccharomyces cerevisiae* from lignocellulosic feedstocks. *Metab. Eng.* 42, 115–125.

Di, P.C., Alvaro, G., Tiziana, C., Monica, Z., and Giuseppe, R. (2005). Potentiation of gene targeting in human cells by expression of *Saccharomyces cerevisiae* Rad52. *Nucleic Acids Res.* 33, 4639–4648.

Duan, X., Gao, J., and Y.J., Z. (2018). Advances in engineering methylotrophic yeast for biosynthesis of valuable chemicals from methanol. *Chin. Chem. Lett.* 29, 681–686.

Gutschner, T., Haemmerle, M., Genovese, G., Draetta, G.F., and Chin, L. (2016). Post-translational regulation of Cas9 during G1 enhances homology-directed repair. *Cell Rep.* 14, 1555–1566.

Heyer, W.-D., Ehmsen, K.T., and Liu, J. (2010). Regulation of homologous recombination in eukaryotes. *Annu. Rev. Genet.* 44, 113–139.

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., et al. (2015). Efficient multiplexed integration of synergistic alleles and metabolic pathways in yeasts via CRISPR-Cas. *Cell Syst.* 1, 88.

Jayatilaka, K., Sheridan, S.D., Bold, T.D., Bochenska, K., Logan, H.L., Weichselbaum, R.R., Bishop, D.K., and Connell, P.P. (2008). A chemical compound that stimulates the human homologous recombination protein RAD51. *Proc. Natl. Acad. Sci. U S A* 105, 15848–15853.

Johnson, B.L., Thyagarajan, B., Krueger, L., Hirsch, B., and Campbell, C. (1996). Elevated

levels of recombinational DNA repair in human somatic cells expressing the *Saccharomyces cerevisiae* RAD52 gene. *Mutat. Res./DNA Repair* 363, 179–189.

Juergens, H., Varela, J.A., Gorter de Vries, A.R., Perli, T., Gast, V.J., Gyurchev, N.Y., Rajkumar, A.S., Mans, R., Pronk, J.T., and Morrissey, J.P. (2018). Genome editing in *Kluyveromyces* and *Ogataea* yeasts using a broad-host-range Cas9/gRNA co-expression plasmid. *FEMS Yeast Res.* 18, foy012.

Krejci, L., Altmannova, V., Spirek, M., and Zhao, X. (2012). Homologous recombination and its regulation. *Nucleic Acids Res.* 40, 5795–5818.

Kretzschmar, A., Otto, C., Holz, M., Werner, S., Hübner, L., and Barth, G. (2013). Increased homologous integration frequency in *Yarrowia lipolytica* strains defective in non-homologous end-joining. *Curr. Genet.* 59, 63–72.

Liachko, I., and Dunham, M.J. (2014). An autonomously replicating sequence for use in a wide range of budding yeasts. *FEMS Yeast Res.* 14, 364–367.

Liu, D., Geiselman, G.M., Coradetti, S., Cheng, Y., and Gladden, J. (2020). Exploiting nonionic surfactants to enhance fatty alcohol production in *Rhodospiridium toruloides*. *Biotechnol. Bioeng.* 117, 1418–1425.

Manfrão-Netto, J.H., Gomes, A.M., and Parachin, N.S. (2019). Advances in using *Hansenula polymorpha* as chassis for recombinant protein production. *Front. Bioeng. Biotech.* 7, 94.

McGinn, J., and Marraffini, L.A. (2018). Molecular mechanisms of CRISPR-Cas spacer acquisition. *Nat. Rev. Microbiol.* 17, 7.

Mcein, B.A., and Stuart, D.T. (2017). Optimization of C16 and C18 fatty alcohol production by an engineered strain of *Lipomyces starkeyi*. *J. Ind. Microbiol. Biotechnol.* 45, 1–14.

Mimitou, E.P., and Symington, L.S. (2008). Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature* 455, 770.

Mitsui, R., Yamada, R., and Ogino, H. (2019). CRISPR system in the yeast *Saccharomyces cerevisiae* and its application in the bioproduction of useful chemicals. *World J. Microbiol. Biotechnol.* 35, 111.

Nambu-Nishida, Y., Nishida, K., Hasunuma, T., and Kondo, A. (2017). Development of a comprehensive set of tools for genome engineering in a cold- and thermo-tolerant *Kluyveromyces marxianus* yeast strain. *Sci. Rep.* 7, 1–8.

Numamoto, M., Maekawa, H., and Kaneko, Y. (2017). Efficient genome editing by CRISPR/Cas9 with a tRNA-sgRNA fusion in the methylotrophic yeast *Ogataea polymorpha*. *J. Biosci. Bioeng.* 124, 487–492.

O'Driscoll, M., and Jeggo, P.A. (2006). The role of double-strand break repair—insights from human genetics. *Nat. Rev. Genet.* 7, 45–54.

Olson, D.G., Sparling, R., and Lynd, L.R. (2015). Ethanol production by engineered thermophiles. *Curr. Opin. Biotechnol.* 33, 130–141.

Phithakrotchanakoon, C., Puseenam, A., Wongwisansri, S., Eurwilaichitr, L., Ingsriswang, S., Tanapongpipat, S., and Roonsawang, N. (2018). CRISPR-Cas9 enabled targeted mutagenesis in the thermotolerant methylotrophic yeast *Ogataea thermomethanolica*. *FEMS Microbiol. Lett.* 365, fry105.

Raschmanova, H., Weninger, A., Glieder, A., Kovar, K., and Vogl, T. (2018). Implementing CRISPR-Cas technologies in conventional and non-conventional yeasts: current state and future prospects. *Biotechnol. Adv.* 36, 641–665.

Saraya, R., Krikken, A.M., Kiel, J.A., Baerends, R.J., Veenhuis, M., and van der Klei, I.J. (2012). Novel genetic tools for *Hansenula polymorpha*. *FEMS Yeast Res.* 12, 271–278.

Schwartz, C., Frogue, K., Ramesh, A., Misa, J., and Wheelodon, I. (2017). CRISPRi repression of nonhomologous end-joining for enhanced genome engineering via homologous recombination in *Yarrowia lipolytica*. *Biotechnol. Bioeng.* 114, 2896–2906.

Schwartz, C., and Wheelodon, I. (2017). Genome and metabolic engineering in non-conventional yeasts: current advances and applications. *Synth. Syst. Biotechnol.* 2, 198–207.

Schwartz, C.M., Hussain, M.S., Blenner, M., and Wheelodon, I. (2015). Synthetic RNA polymerase III promoters facilitate high-efficiency CRISPR-Cas9-mediated genome editing in *Yarrowia lipolytica*. *ACS Synth. Biol.* 5, 356.

Shao, S., Ren, C., Liu, Z., Bai, Y., Chen, Z., Wei, Z., Wang, X., Zhang, Z., and Xu, K. (2017). Enhancing CRISPR/Cas9-mediated homology-directed repair in mammalian cells by expressing *Saccharomyces cerevisiae* Rad52. *Int. J. Biochem. Cell Biol.* 92, 43–52.

Shao, Z., Zhao, H., and Zhao, H. (2008). DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* 37, e16.

Ubiyovk, V.M., Ananin, V.M., Malyshev, A.Y., Kang, H.A., and Sibirny, A.A. (2011). Optimization of glutathione production in batch and fed-batch cultures by the wild-type and recombinant strains of the methylotrophic yeast *Hansenula polymorpha* DL-1. *BMC Biotechnol.* 11, 8.

Vispé, S., Cazaux, C., Lesca, C., and Defais, M. (1998). Overexpression of Rad51 protein stimulates homologous recombination and increases resistance of mammalian cells to ionizing radiation. *Nucleic Acids Res.* 26, 2859–2864.

Voronovsky, A.Y., Rohulya, O.V., Abbas, C.A., and Sibirny, A.A. (2009). Development of strains of the thermotolerant yeast *Hansenula polymorpha* capable of alcoholic fermentation of starch and xylan. *Metab. Eng.* 11, 234–242.

Wang, L., Deng, A., Zhang, Y., Liu, S., Liang, Y., Bai, H., Cui, D., Qiu, Q., Shang, X., and Yang, Z. (2018). Efficient CRISPR–Cas9 mediated multiplex genome editing in yeasts. *Biotechnol. Biofuels* 11, 277.

Weninger, A., Fischer, J.E., Raschmanová, H., Kniely, C., Vogl, T., and Glieder, A. (2018). Expanding the CRISPR/Cas9 toolkit for *Pichia*

pastoris with efficient donor integration and alternative resistance markers. *J. Cell. Biochem.* 119, 3183–3198.

Weninger, A., Hatzl, A.-M., Schmid, C., Vogl, T., and Glieder, A. (2016). Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*. *J. Biotechnol.* 235, 139–149.

Yáñez, R.J., and Porter, A.C.G. (1999). Gene targeting is enhanced in human cells overexpressing hRAD51. *Gene Ther.* 6, 1282–1290.

Yang, D., Scavuzzo, M.A., Chmielowiec, J., Sharp, R., Bajic, A., and Borowiak, M. (2016). Enrichment of G2/M cell cycle phase in human pluripotent stem cells enhances HDR-mediated gene repair

with customizable endonucleases. *Sci. Rep.* 6, 21264.

Yoo, S.J., Chung, S.Y., Lee, D.-j., Kim, H., Cheon, S.A., and Kang, H.A. (2015). Use of the cysteine-repressible HpMET3 promoter as a novel tool to regulate gene expression in *Hansenula polymorpha*. *Biotechnol. Lett.* 37, 2237–2245.

Yu, T., Zhou, Y.J., Huang, M., Liu, Q., Pereira, R., David, F., and Nielsen, J. (2018). Reprogramming yeast metabolism from alcoholic fermentation to lipogenesis. *Cell* 174, 1549–1558.

Zhou, Y.J., Buijs, N.A., Zhu, Z., Gomez, D.O., Boonsombuti, A., Siewers, V., and Nielsen, J. (2016a). Harnessing yeast peroxisomes for biosynthesis of fatty-acid-derived biofuels and chemicals with relieved side-pathway

competition. *J. Am. Chem. Soc.* 138, 15368–15377.

Zhou, Y.J., Buijs, N.A., Zhu, Z., Qin, J., Siewers, V., and Nielsen, J. (2016b). Production of fatty acid-derived oleochemicals and biofuels by synthetic yeast cell factories. *Nat. Commun.* 7, 11709.

Zhou, Y.J., Gao, W., Rong, Q., Jin, G., Chu, H., Liu, W., Yang, W., Zhu, Z., Li, G., and Zhu, G. (2012). Modular pathway engineering of diterpenoid synthases and the mevalonic acid pathway for miltiradiene production. *J. Am. Chem. Soc.* 134, 3234–3241.

Zhou, Y.J., Kerkhoven, E.J., and Nielsen, J. (2018). Barriers and opportunities in bio-based production of hydrocarbons. *Nat. Energy* 3, 925–935.

iScience, Volume 24

Supplemental information

**Recombination machinery engineering
for precise genome editing in methylotrophic
yeast *Ogataea polymorpha***

Jiaoqi Gao, Ning Gao, Xiaoxin Zhai, and Yongjin J. Zhou

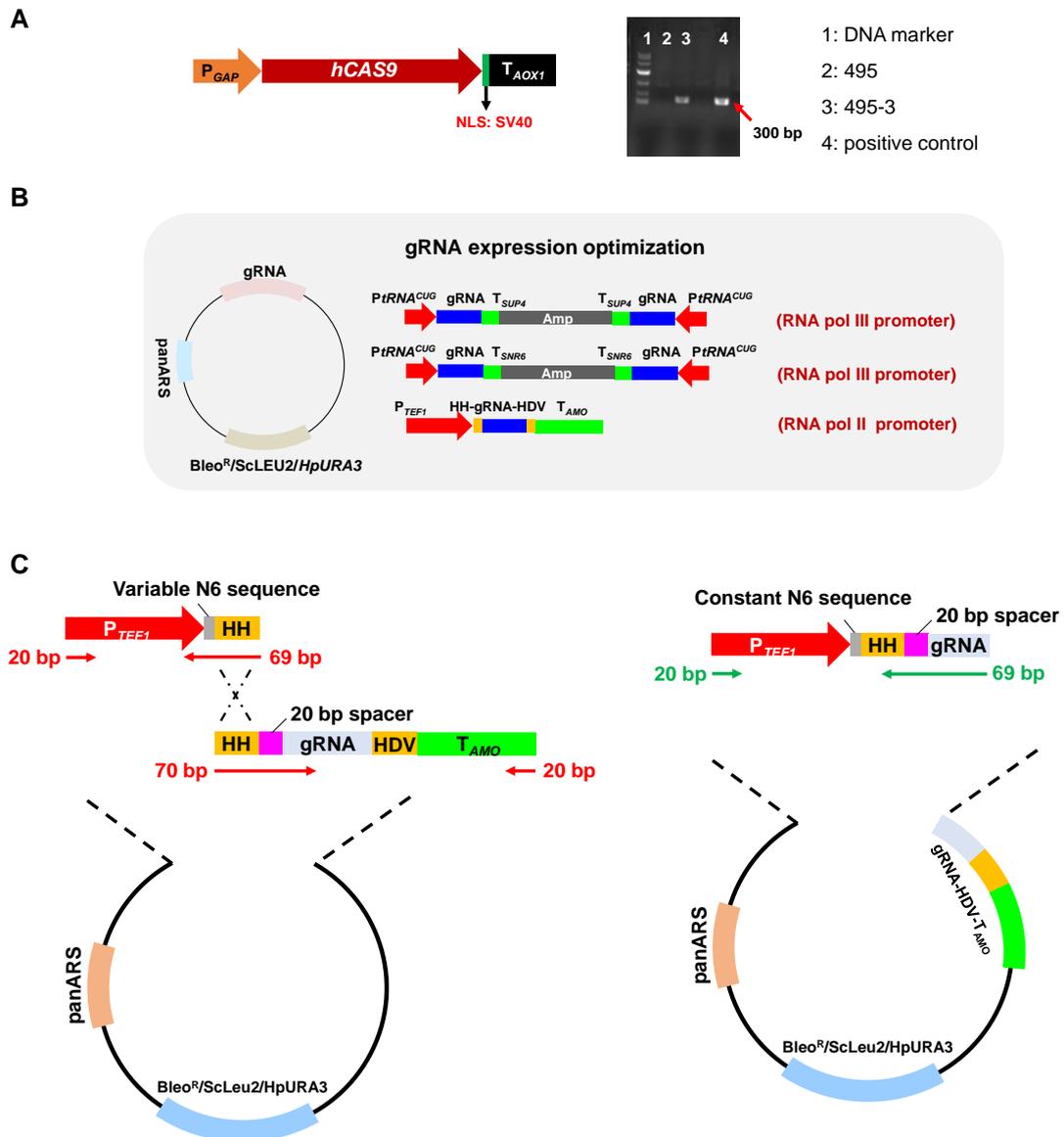


Figure S1. Construction and optimization of CRISPR/Cas9 system in *O. polymorpha*, related to Figure 2.

(A) Transcription of the integrated hCas9 protein controlled by P_{GAP} and T_{AOX1} from *P. pastoris* was verified by RT-PCR. Strain 495-3 with a copy of the integrated Cas9 and the positive control showed the obvious bands containing partial Cas9 sequence, compared with the original strain 495; (B) Optimization of gRNA expression vectors. An episomal plasmid with panARS and selective markers was obtained based on pPICZ A for gRNA expression. Expression cassettes were optimized by adoption of either $tRNA^{CUG}$ promoter, or $TEF1$ promoter. (C) Comparison of gRNA plasmid construction with variable and constant N6 sequence.

A

| | | Protospacer | PAM |
|-------------------------------------|--|--------------------------|---------------|
| <i>ADE2</i> reference from NCYC 495 | | AGAGCGCTTGAACCCCACACGCGT | CGGTTTAC |
| Colony 1 | | AGAGCGCTTGAACCCCACAC | CGT CGGTTTAC |
| Colony 2 | | AGAGCGCTTGAACCCCACAC | CGT CGGTTTAC |
| Colony 3 | | AGAGCGCTTGAACCCCACAC | #CGT CGGTTTAC |
| Colony 4 | | AGAGCGCTTGAACCCCACACG | GT CGGTTTAC |

#=GGTCAAATGTAGTGTACTACTACCTTTAAGTAAGACCCATAGGGCTCTGGTCAAAAGTAGTGC
ACTACTATAGACCAAGTCCCATAGGGCTCTGGTCAAAAGTAATGCACTACTATAGACCAAGTCC
CATAGGGCTCTG

B

| | | Protospacer | PAM |
|-------------------------------------|--|--------------------------|--------------|
| <i>KU80</i> reference from NCYC 495 | | CTGGCCATCGTTCGCAGAAGATCA | TGGAGATG |
| Colony 1 | | CTGGCCATCGTTCGCAGAAGA | CAT TGGAGATG |
| Colony 2 | | CTGGCCATCGTTCGCAGAAGA | CT TGGAGATG |
| Colony 3 | | CTGGCCATCGTTCGCAGAAGA | CT TGGAGATG |
| Colony 4 | | CTGGCCATCGTTCGCAGAAGA | CAT TGGAGATG |
| Colony 5 | | CTGGCCATCGTTCGCAGAAGAT | TGGAGATG |
| Colony 6 | | CTGGCCATCGTTCGCAGAAGAT | TGGAGATG |
| Colony 7 | | CTGGCCATCGTTCGCAGAAGATG | TGGAGATG |
| Colony 8 | | CTGGCCATCGTTCGCAGAAGA | CAT TGGAGATG |

C

| | | Protospacer | PAM |
|-------------------------------------|----------|-----------------------|---------------|
| <i>KU80</i> reference from NCYC 495 | | CTGGCCATCGTTCGCAGAAGA | TCATGGAGATG |
| 1 bp insertion | Colony 1 | CTGGCCATCGTTCGCAGAAGA | ATCA TGGAGATG |
| No mutation | Colony 2 | CTGGCCATCGTTCGCAGAAGA | TCATGGAGATG |
| 1 bp deletion | Colony 3 | CTGGCCATCGTTCGCAGAAGA | CAT TGGAGATG |
| 2 bp deletion | Colony 4 | CTGGCCATCGTTCGCAGAA | TCATGGAGATG |
| 1 bp insertion | Colony 5 | CTGGCCATCGTTCGCAGAAGA | ATCA TGGAGATG |
| Random insertion | Colony 6 | CTGGCCATCGTTCGCAGAAGG | TCATGGAGATG |
| 2 bp deletion | Colony 7 | CTGGCCATCGTTCGCAGAA | TCATGGAGATG |
| 1 bp insertion | Colony 8 | CTGGCCATCGTTCGCAGAAGA | ATCA TGGAGATG |
| No mutation | Colony 9 | CTGGCCATCGTTCGCAGAAGA | TCATGGAGATG |

#=TCCAGTGAGTTAATAAAATGACAAGGTGTTGCCTGTCAGCGTTGTAGACTACTGCCTCACAAGCCCGGGCTTTAG
TGGCACTGGCATACTCGCTTGCTCTGCACTGAAGGAGCTACCGAACCGTCCGCGTTGCAGCTCTCACTTTCTCCC
GCTACAGCATGTCTACCATGACATATTTGGTGGGATTGCCTCCAACAATAGGATTGGATTGTGACTAAATCTAT
GAATCATCCGATATTGCAATGACTATTATTACAAATATTTCCCCGAAGATCTAGGGTTCATCTATGGCCGATAGGT
GAGATAATTAATTGATCCGACACGTCAGGCTTACATGTCATAAGTAAGTAACCC...

Figure S2. DNA repair pattern of DSB by the CRISPR/Cas9 system via DNA sequencing, related to Figure 2.

(A) Targeting gene *OpADE2* with variable N6 sequence in gRNA, 4 red colons were selected for DNA sequencing; (B) Targeting gene *OpKU80* with variable N6 sequence in gRNA, 8 colons were randomly selected for DNA sequencing; (C) Targeting gene *OpKU80* with constant N6 sequence in gRNA, 9 colons were randomly selected for DNA sequencing.

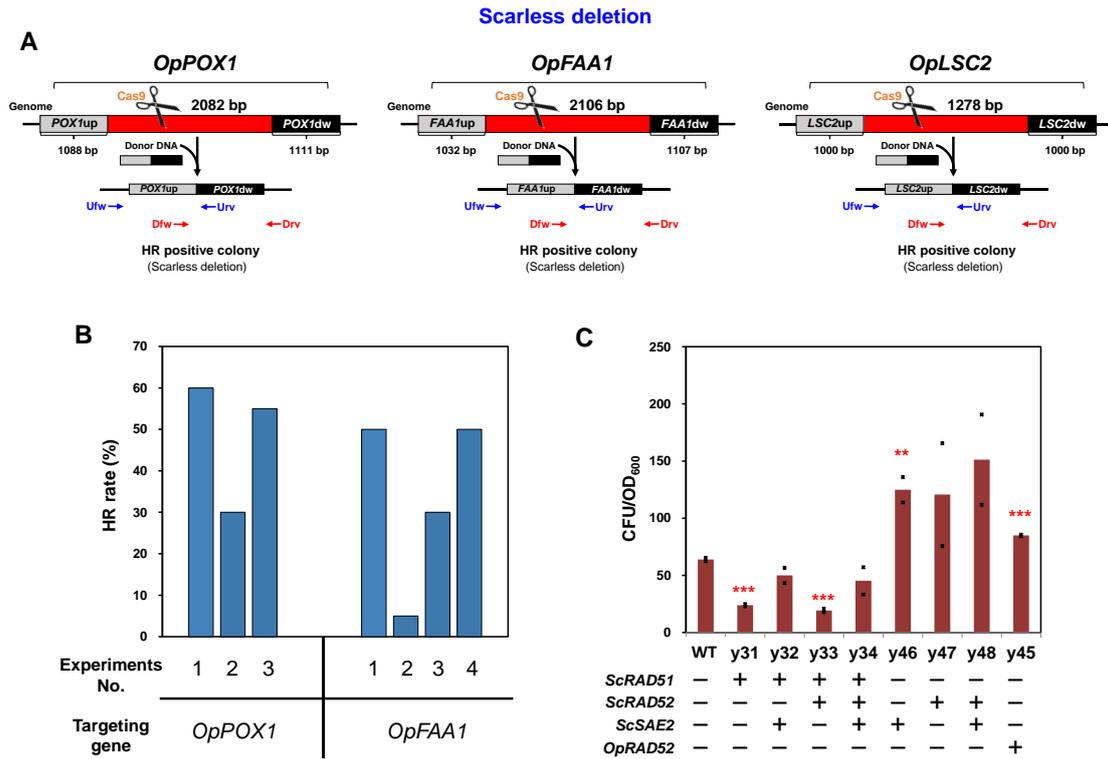


Figure S3. HR rates in *O. polymorpha* was represented by rates of gene scarless deletion, related to Figure 3.

(A) Schematic illustration of scarless deletion for gene *OpFAA1*, *OpPOX1*, and *OpLSC2*. Correct transformants demonstrated a shorter band in agarose gel electrophoresis from PCR verification. Ufw and Urv, Dfw and Drv are forward and reverse primers from upstream and downstream, respectively, for colony PCR; (B) Repeated experiments on HR rates of gene scarless deletion in strain Cas9-Pds1. (C) CFU/OD₆₀₀ in strains with overexpressed HR-related proteins when targeting *OpLSC2*. Strain y31 and y33 decreased the efficiency, and strain y32 and y34 showed a similar rates with the control strain. Unexpectedly, an increased transformation rates in other strains (y45, y46, y47, and y48) was observed. The corresponding HR rates in these strains were illustrated in Figure 3C. Total 20 colony from each biological parallel was picked and tested by colony PCR to calculate HR rate. Data are presented as means of two biologically independent samples with displayed data-points. Red asterisks indicated statistical significance as determined using paired t-test (**P < 0.01; ***<0.001).

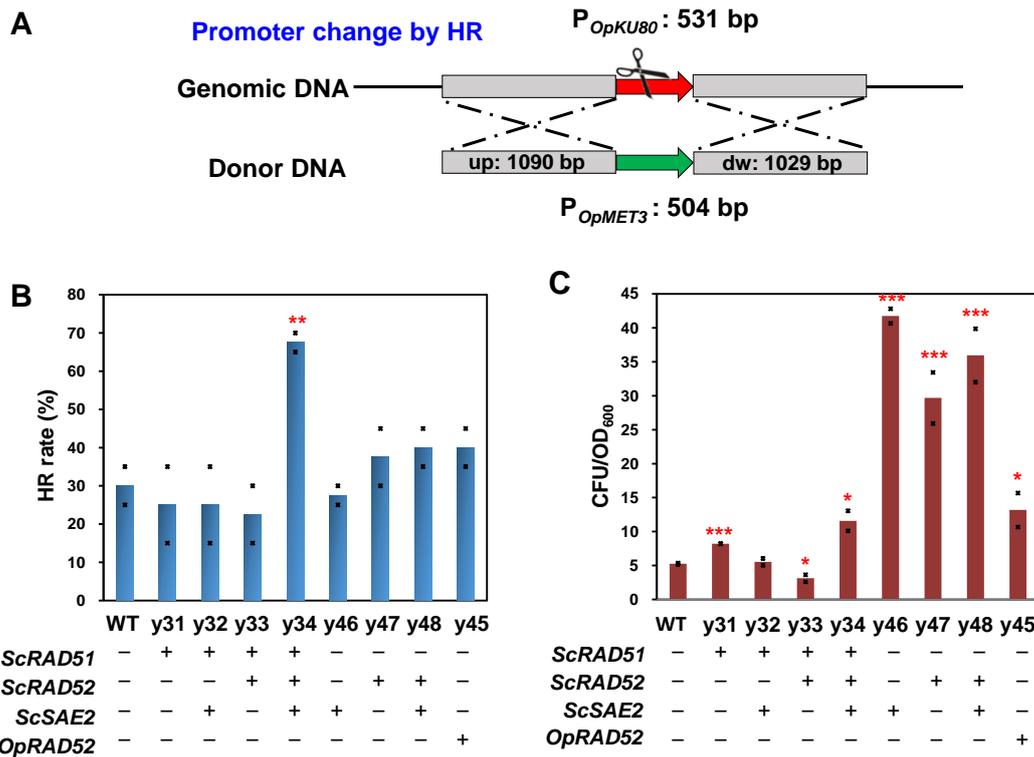


Figure S4. *In situ* substitution of a specific DNA segment in a series of strains with overexpressed HR-related proteins, related to Figure 3.

Substitution of Gene *KU80* promoter was replaced by gene *MET3* promoter (P_{OpMET3}), with a length of 504 bp (A). Transformation experiments were performed in strains with overexpressed HR-related proteins, using our CRISPR/Cas9 system. Transformants were observed and verified to detect the differences in HR rate (B) and CFU/OD₆₀₀ (C). Total 20 colony from each biological parallel was picked and tested by colony PCR to calculate HR rate. Data are presented as means of two biologically independent samples with displayed data-points. Red asterisks indicate statistical significance as determined using paired t-test (* $P < 0.05$; ** $P < 0.01$; *** < 0.001).

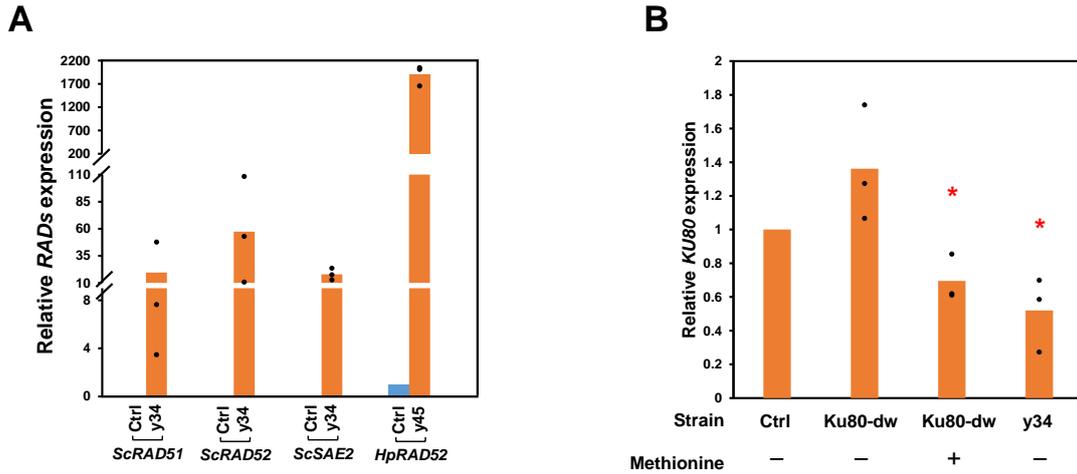


Figure S5. Quantitation of HR-related genes and *KU80* gene by qPCR experiments, related to Figure 3, Figure 4, and Figure 5.

Cells were cultivated in YPD medium at 37°C, 220 rpm for around 24 h, and then collected, and washed twice with ddH₂O for the extraction of total RNA. cDNA reversely transcribed from the total RNA was used as the template for a two-step qPCR reaction. (A) Relative expression level of HR-related genes in wild-type (Ctrl), y34, and y45. Gene *ScRAD51*, *ScRAD52*, *ScSAE2* in strain y34 had been proved to transcribe at a level of dozens of times higher than that of the endogenous gene *OpRAD52* in wild type. However, gene *OpRAD52* in strain y45 showed the extreme over-expression due to the strong constructive promoter *P_{GAP}*. (B) Relative expression level of gene *KU80* in strains of wild-type (Ctrl), Ku80-dw, Ku80-dw+Met, and y34. Methionine down-regulated up to 30% of *KU80* expression in strain Ku80-dw, and y34 also showed 50% discounted level due to HR competition. Data are presented as means of three biologically independent samples with displayed data-points. Red asterisks indicate statistical significance as determined using paired t-test (**P* < 0.05).

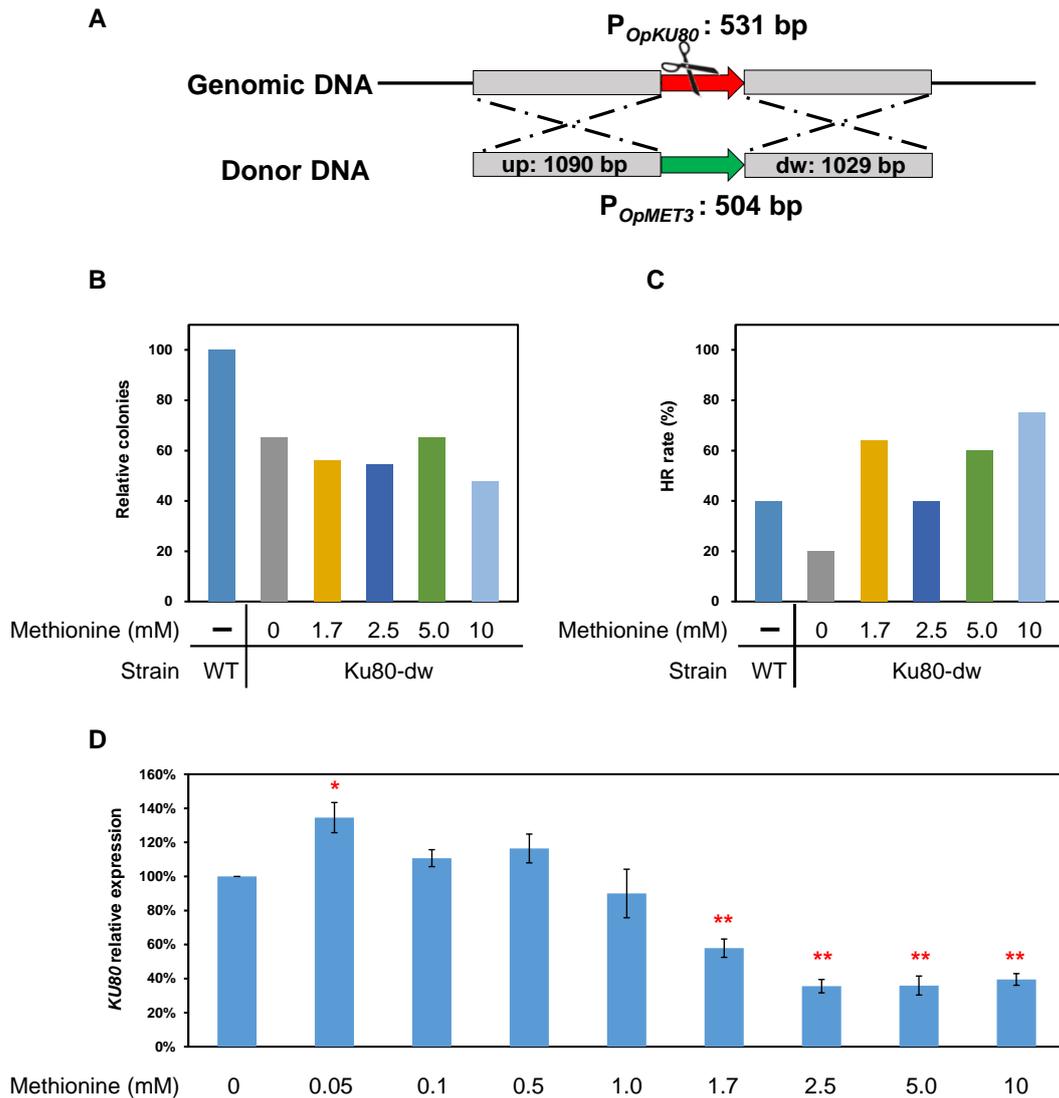


Figure S6. Ku80 expression was down-regulated by promoter P_{OpMET3} with the existence of methionine, related to Figure 4.

(A) Promoter of gene *KU80* was replaced by gene *MET3* promoter (P_{OpMET3}), with a length of 504 bp. Methionine concentrations were optimized for better performances in both transformation efficiency (B) and HR rate (C). Five concentrations of methionine, 0, 1.7, 2.5, 5.0, 10 mM, were selected to conduct the transformation experiments, and methionine was both added in YPD medium that is used for recovery procedure during transformation, and the final screening plates during the transformation process. Total 20 colony from each biological parallel was picked and tested by colony PCR to calculate HR rate. (D) qPCR was carried out to test expression levels of gene *KU80* at different methionine concentrations. Strain Pre-cultured strain Ku80-dw was inoculated in Delft basic salt medium containing 0, 0.05, 0.1, 0.5, 1.0, 1.7, 2.5, 5, and 10 mM methionine. Cell were harvested after 8-10 h of cultivation at 37°C, 220 rpm. Data are represented as mean \pm SEM. Statistical tests are presented as red asterisks (* $P < 0.05$; ** $P < 0.01$).

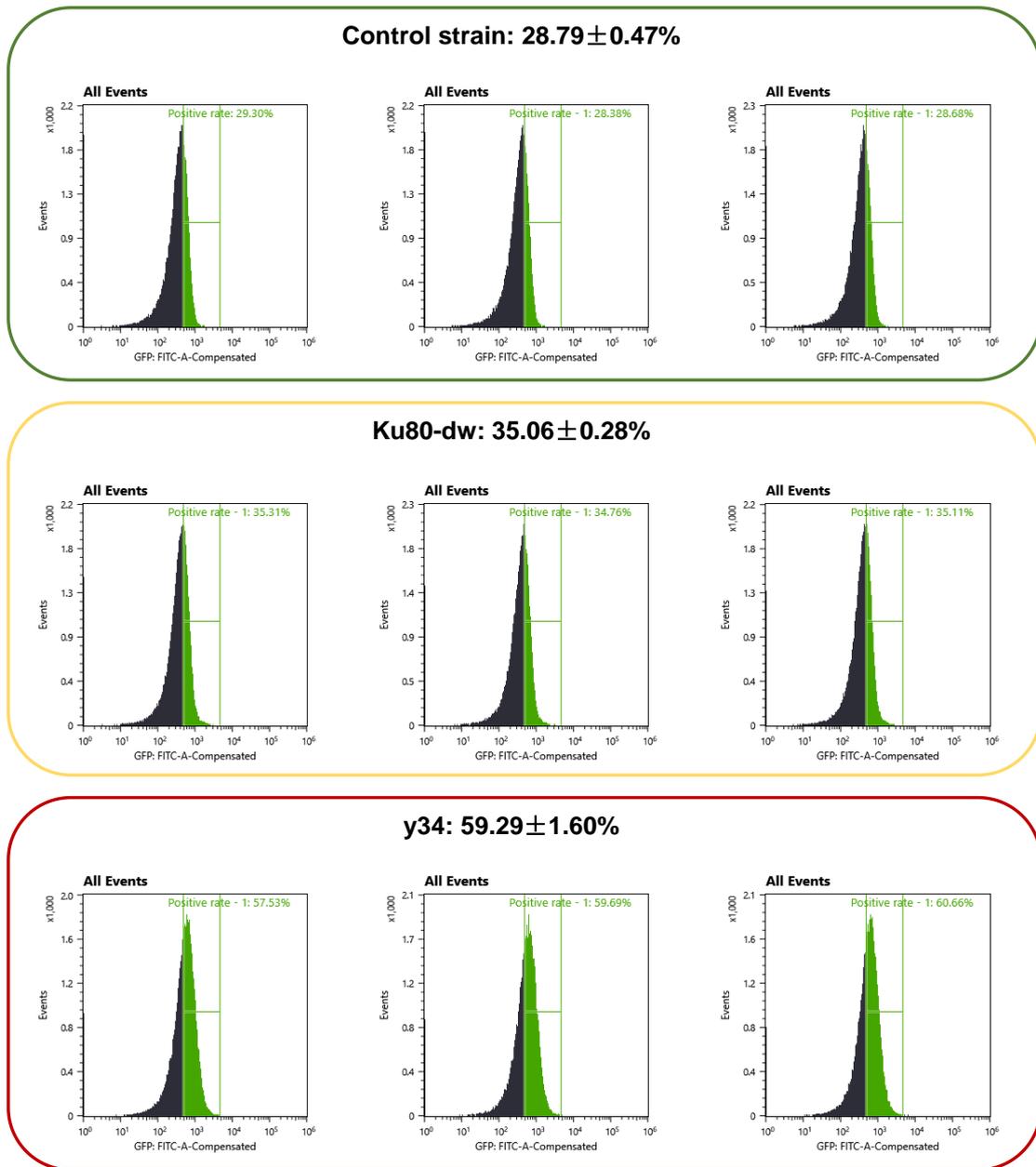


Figure S7. Positive HR cells counting by flow cytometer, related to Figure 5.

Cells were cultivated and collected as shown in the section of "Competition assay between HR and NHEJ" of "Methods and Materials". Cells were washed and resuspended in PBS buffer, and then detected by flow cytometer (Sony SH800S). 50,000 cells were counted to test the GFP fluorescence intensity. When setting the intensity >500 as HR positive cells, since the GFP fluorescence intensity of y34 is around 500, achieving a HR rate of 60%. Thus, the positive cells of control, Ku80-dw, and y34 accounted for $28.79 \pm 0.47\%$, $35.06 \pm 0.28\%$, $59.29 \pm 1.60\%$, respectively, which was in agreement of Figure 5.

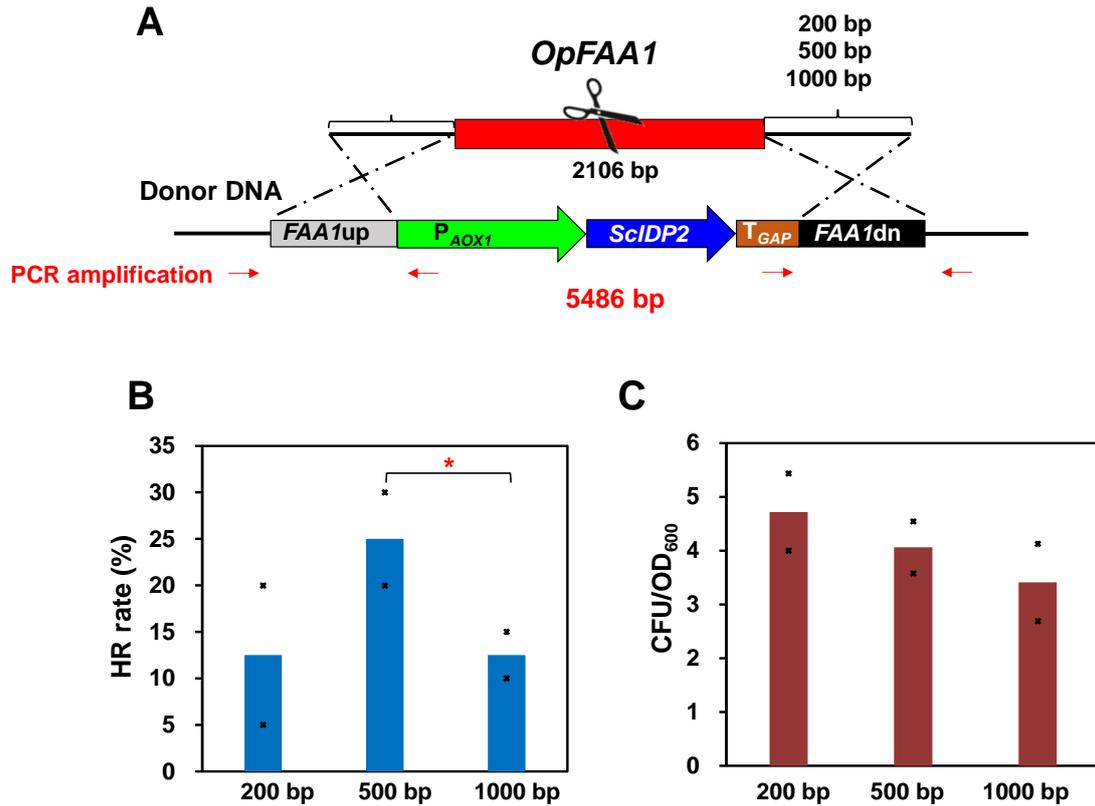


Figure S8. Homologous integration of a large fragment by LiAc/ssDNA method, related to Figure 6.

(A) Schematic illustration of homologous integration of a large fragment (*ScIDP2* cassette at *OpFAA1* locus) with HA length of 200 bp, 500 bp, and 1000 bp, respectively. (B) HR rates for *ScIDP2* integration with various HA lengths in strain y34 with overexpressing HR relating proteins; (C) CFU/OD₆₀₀ for *ScIDP2* integration with various HA lengths in strain y34 with overexpressing HR-related proteins. Total 20 colony from each biological parallel was picked and tested by colony PCR to calculate HR rate. Data are presented as means of two biologically independent samples with displayed data-points. Red asterisks indicate statistical significance as determined using paired t-test (*P<0.05).

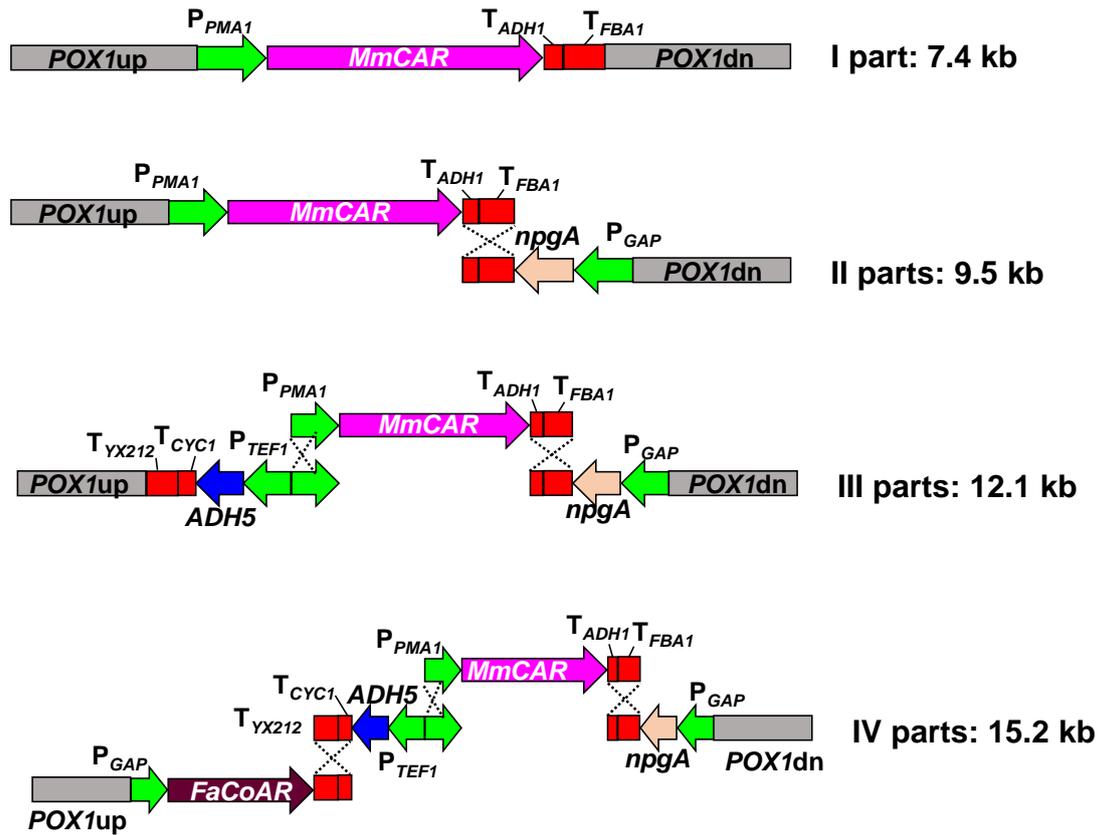


Figure S9. Schematic illustration of *in vivo* pathway assembly for fatty alcohol production in *O. polymorpha*, related to Figure 7.

Each expression cassette contains one essential gene for fatty alcohol production was separately integrated into *POX1* site with HA length of 500-1000 bp. At most, *in vivo* assembly of four cassettes (*MmCAR*, *npgA*, *ADH5*, and *FaCoAR*) reached up to 15.2 kb in length.

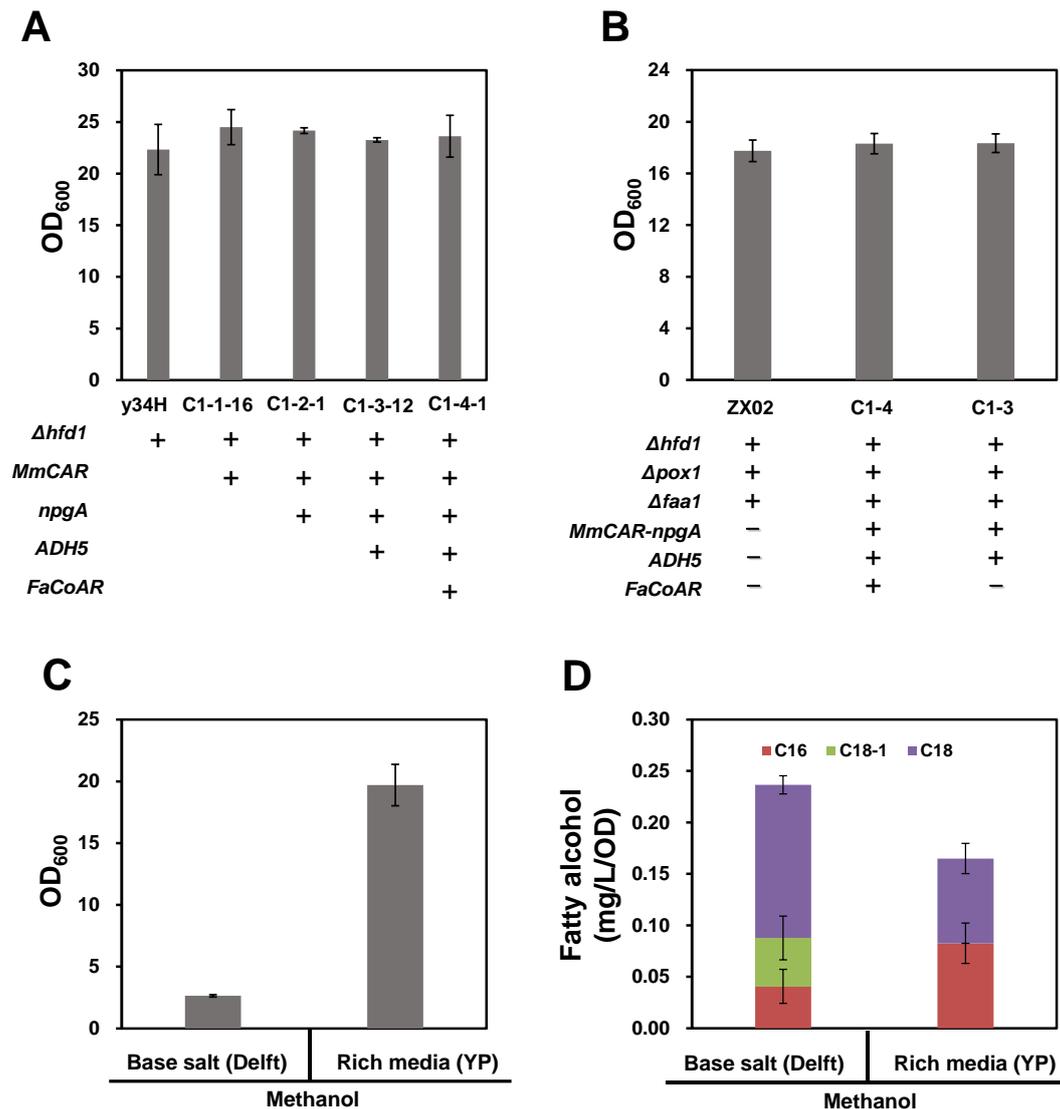


Figure S10. Fatty alcohol production from both glucose and methanol containing media, related to Figure 7.

(A) The optical density at 600 nm of multiple fatty alcohol producing strains. Multiple expression cassettes were integrated at *POX1* site as shown in Figure S9. Strains were cultivated in basic salt medium with 20 g/L glucose at 37°C, 220 rpm for 96 h, and the corresponding fatty alcohol titers were illustrated in Figure 7D. To further increase fatty acids supply, gene *FAA1* was deleted to obtain strain C1-3 (*MmCAR*, *npgA*, *ADH5*, *Δhfd1*, *Δpox1*, *Δfaa1*), which was also cultivated in basic salt medium with 20 g/L glucose at 37°C, 220 rpm for 96 h. Fatty alcohol titer was dramatically increased to 12m g/L (Figure 7E) without the influences in cell growth (B). Strain C1-3 was also cultivated in methanol-containing media, including Delft basic salt medium+5 g/L methanol and YP medium+5 g/L methanol. 5 g/L of methanol was supplemented at 24 h and 48 h. Samples were taken at 96 h to analyze biomass (C) and fatty alcohol per OD₆₀₀ (D). Rich media promoted cell growth but not fatty alcohol production. Data are presented as mean ± s.e.m. (n=3 biologically independent samples).

Table S1 Optimization of gRNA expression plasmids^a, related to Figure 2.

| Type | gRNA promoter | gRNA terminator | N6 sequence | Target gene | Efficiency |
|------|---------------------------------------|---------------------------|-------------|---------------|----------------------|
| / | / | / | / | <i>OpADE2</i> | 0% (0/654) |
| III | <i>P_{tRNA^{CUG}}</i> | <i>T_{ScSUP4}</i> | / | <i>OpADE2</i> | <0.1% |
| III | <i>P_{tRNA^{CUG}}</i> | <i>T_{OpSNR6}</i> | / | <i>OpADE2</i> | 0.46% (4/866) |
| II | <i>P_{OpTEF1}</i> | <i>T_{OpAMO}</i> | Variable | <i>OpADE2</i> | 95.6% (1243/1290) |
| II | <i>P_{OpTEF1}</i> | <i>T_{OpAMO}</i> | Constant | <i>OpADE2</i> | 93.1% (95/102) |
| II | <i>P_{OpTEF1}</i> | <i>T_{OpAMO}</i> | Variable | <i>OpKU80</i> | 100% (8/8) |
| II | <i>P_{OpTEF1}</i> | <i>T_{OpAMO}</i> | Constant | <i>OpKU80</i> | 77.8 (7/9) |

Note: ^aWhile targeting gene *OpADE2*, editing rates were calculated as the ratio of red colony and total colony. For gene *OpKU80*, multiple colony was picked, and the positive colony was determined by DNA sequencing at editing sites.

Table S2 Strains used in this study, related to Transparent Methods.

| Strains | Genotype | Resource Reference | or |
|---------------------|---|--------------------|----|
| DH5 α | F-, ϕ 80d/lacZ Δ M15, Δ (lacZYA-argF)U169, deoR, recA1, endA1, hsdR17(rk-, mk+), phoA, supE44, λ -, thi-1, gyrA96, relA1 | Takara | |
| NCYC 495 | MATa; leu1.1 | CGMCC | |
| 495-3 | MATa; leu1.1; (P _{GAP-hCAS9-T_{AOX1}}) | This study | |
| 495-3 Δ ura3 | MATa; leu1.1; ura3 Δ ; (P _{GAP-hCAS9-T_{AOX1}}) | This study | |
| Δ ku80 | MATa; leu1.1; ura3 Δ ; (P _{GAP-hCAS9-T_{AOX1}}); ku80 Δ | This study | |
| Ku80-dw | MATa; leu1.1; ura3 Δ ; (P _{GAP-hCAS9-T_{AOX1}}); P _{OpKu80} ::P _{OpMET3} | This study | |
| Cas9-Pds1 | MATa; leu1.1; (P _{GAP-hCAS9~Pds1-T_{AOX1}}) | This study | |
| y31 | MATa; ura3 Δ ; (P _{GAP-hCAS9-T_{AOX1}}); (P _{TPI1-ScRAD51-T_{AMO}}) | This study | |
| y32 | MATa; ura3 Δ ; (P _{GAP-hCAS9-T_{AOX1}}); (P _{TPI1-ScRAD51-T_{AMO}}); (P _{TKL1-ScSAE2-T_{URA3}}) | This study | |
| y33 | MATa; ura3 Δ ; (P _{GAP-hCAS9-T_{AOX1}}); (P _{TPI1-ScRAD51-T_{AMO}}); (P _{PGI1-ScRAD52-T_{GAP}}) | This study | |
| y34 | MATa; ura3 Δ ; (P _{GAP-hCAS9-T_{AOX1}}); (P _{TPI1-ScRAD51-T_{AMO}}); (P _{PGI1-ScRAD52-T_{GAP}}); (P _{TKL1-ScSAE2-T_{URA3}}) | This study | |
| y45 | MATa; ura3 Δ ; (P _{GAP-hCAS9-T_{AOX1}}); (P _{GAP-OpRAD52-T_{GAP}}) | This study | |
| y46 | MATa; ura3 Δ ; (P _{GAP-hCAS9-T_{AOX1}}); (P _{TKL1-ScSAE2-T_{URA3}}) | This study | |
| y47 | MATa; ura3 Δ ; (P _{GAP-hCAS9-T_{AOX1}}); (P _{PGI1-ScRAD52-T_{GAP}}); | This study | |
| y48 | MATa; ura3 Δ ; (P _{GAP-hCAS9-T_{AOX1}}); (P _{PGI1-ScRAD52-T_{GAP}}); (P _{TKL1-ScSAE2-T_{URA3}}) | This study | |
| y31-Ku80dw | y31 P _{OpKu80} ::P _{OpMET3} | This study | |
| y32-Ku80dw | y32 P _{OpKu80} ::P _{OpMET3} | This study | |
| y33-Ku80dw | y33 P _{OpKu80} ::P _{OpMET3} | This study | |
| y34-Ku80dw | y34 P _{OpKu80} ::P _{OpMET3} | This study | |
| y45-Ku80dw | y45 P _{OpKu80} ::P _{OpMET3} | This study | |
| y46-Ku80dw | y46 P _{OpKu80} ::P _{OpMET3} | This study | |
| y47-Ku80dw | y47 P _{OpKu80} ::P _{OpMET3} | This study | |
| y48-Ku80dw | y48 P _{OpKu80} ::P _{OpMET3} | This study | |
| y34H | y34, hfd1 Δ | This study | |
| C1-1-16 | y34, hfd1 Δ , pox1 Δ ::(P _{PMA1-MmCARper2-T_{ADH1}}) | This study | |
| C1-2-1 | y34, hfd1 Δ , pox1 Δ ::(P _{PMA1-MmCARper2-T_{ADH1}+T_{FBA1-npgAper1-P_{GAP}}}) | This study | |
| C1-3-12 | y34, hfd1 Δ , pox1 Δ ::(T _{CYC1-ADH5per1-P_{TEF1}+P_{PMA1-MmCARper2-T_{ADH1}+T_{FBA1-npgAper1-P_{GAP}}}}) | This study | |
| C1-4-1 | y34, hfd1 Δ , pox1 Δ ::(P _{GAP-FaCoARper2-T_{YX212}+T_{CYC1-ADH5per1-P_{TEF1}+P_{PMA1-MmCARper2-T_{ADH1}+T_{FBA1-npgAper1-P_{GAP}}}})} | This study | |
| ZX02 | y34, hfd1 Δ , pox1 Δ , faa1 Δ | | |
| C1-3 | C1-3-12 faa1 Δ | | |
| C1-4 | C1-4-1 faa1 Δ | | |

Table S3 Plasmids used in this study^a, related to Transparent Methods.

| Plasmid | Description | Resource |
|----------------------------|--|------------|
| pPICZ A | 2 μm, <i>Bleo</i> ^R , P _{AOX1} , T _{AOX1} | Invitrogen |
| For gene expression | | |
| pHp03 | pPICZ A-(P _{GAP} - hCAS9 -T _{AOX1})- <i>OpURA3</i> | This study |
| pHp26 | pPICZ A-(P _{GAP} - hCAS9-ScPDS1 -T _{AOX1})- <i>OpURA3</i> | This study |
| pHp31 | 2 μm, <i>Amp</i> ^R , <i>ScLEU2</i> , P _{TPI1} - ScRAD51 -T _{AMO} | This study |
| pHp32 | 2 μm, <i>Amp</i> ^R , <i>ScLEU2</i> , (P _{TPI1} - ScRAD51 -T _{AMO})+(P _{TKL1} - ScSAE2 -T _{URA3}) | This study |
| pHp33 | 2 μm, <i>Amp</i> ^R , <i>ScLEU2</i> , (P _{TPI1} - ScRAD51 -T _{AMO})+(P _{PGI1} - ScRAD52 -T _{GAP}) | This study |
| pHp34 | 2 μm, <i>Amp</i> ^R , <i>ScLEU2</i> , (P _{TPI1} - ScRAD51 -T _{AMO})+(P _{PGI1} - ScRAD52 -T _{GAP})+(P _{TKL1} - ScSAE2 -T _{URA3}) | This study |
| pHp44 | 2 μm, <i>Amp</i> ^R , <i>ScLEU2</i> , P _{GAP} - OpRAD51 -T _{GAP} | This study |
| pHp45 | 2 μm, <i>Amp</i> ^R , <i>ScLEU2</i> , P _{GAP} - OpRAD52 -T _{GAP} | This study |
| pHp46 | 2 μm, <i>Amp</i> ^R , <i>ScLEU2</i> , P _{TKL1} - ScSAE2 -T _{URA3} | This study |
| pHp47 | 2 μm, <i>Amp</i> ^R , <i>ScLEU2</i> , P _{PGI1} - ScRAD52 -T _{GAP} | This study |
| pHp48 | 2 μm, <i>Amp</i> ^R , <i>ScLEU2</i> , (P _{PGI1} - ScRAD52 -T _{GAP})+(P _{TKL1} - ScSAE2 -T _{URA3}) | This study |
| pHp63 | 2 μm, <i>Amp</i> ^R , panARS, <i>OpURA3</i> , (P _{AOX1} - MmCARper2 -T _{ADH1})+(P _{CAT1} - npgAper1 -T _{FBA1})+(P _{FGH} - ADH5per1 -T _{CYC1})+(P _{DAS1} - FaCoARper2 -T _{YX212})+(P _{GAP} - ScIDP2 -T _{GAP}) | This study |
| For gRNA expression | | |
| pHpgRNA7 | pPICZ A-(<i>Bleo</i> ^R :: <i>ScLEU2</i>)-panARS-(P _{tRNA^{CUG}} -gRNA/ ADE2 -T _{SUP4} - <i>Amp</i> ^R -T _{SUP4} -gRNA/ ADE2 -P _{tRNA^{CUG}}) | This study |
| pHpgRNA9 | pPICZ A-(<i>Bleo</i> ^R :: <i>ScLEU2</i>)-panARS-(P _{tRNA^{CUG}} -gRNA/ ADE2 -T _{SNR6} - <i>Amp</i> ^R -T _{SNR6} -gRNA/ ADE2 -P _{tRNA^{CUG}}) | This study |
| pHpgRNA11 | pPICZ A-(<i>Bleo</i> ^R :: <i>ScLEU2</i>)-panARS- <i>Amp</i> ^R -(P _{TEF1} -HH-gRNA/ ADE2 -HDV-T _{AMO}) | This study |
| pHpgRNA13 | pPICZ A-(<i>Bleo</i> ^R :: <i>ScLEU2</i>)-panARS- <i>Amp</i> ^R -(P _{TEF1} -HH [*] -gRNA/ ADE2 -HDV-T _{AMO}) ^b | This study |
| pHpgRNA14 | pPICZ A-(<i>Bleo</i> ^R :: <i>ScLEU2</i>)-panARS- <i>Amp</i> ^R -(P _{TEF1} -HH [*] -gRNA/ KU80-1 -HDV-T _{AMO}) | This study |
| pHpgRNA16 | pPICZ A-(<i>Bleo</i> ^R :: <i>ScLEU2</i>)-panARS- <i>Amp</i> ^R -(P _{TEF1} -HH-gRNA/ KU80-1 -HDV-T _{AMO}) | This study |
| pHpgRNA23 | pPICZ A-(<i>Bleo</i> ^R :: <i>ScLEU2</i>)-panARS- <i>Amp</i> ^R -(P _{TEF1} -HH [*] -gRNA/ FAA1 -HDV-T _{AMO}) | This study |
| pHpgRNA24 | pPICZ A-(<i>Bleo</i> ^R :: <i>ScLEU2</i>)-panARS- <i>Amp</i> ^R -(P _{TEF1} -HH [*] -gRNA/ POX1 -HDV-T _{AMO}) | This study |
| pHpgRNA25 | pPICZ A-(<i>Bleo</i> ^R :: <i>ScLEU2</i>)-panARS- <i>Amp</i> ^R -(P _{TEF1} -HH [*] -gRNA/ pKU80 -HDV-T _{AMO}) | This study |
| pHpgRNA27 | pPICZ A-(<i>Bleo</i> ^R :: <i>ScLEU2</i>)-panARS- <i>Amp</i> ^R -(P _{TEF1} -HH [*] -gRNA/ URA3 -HDV-T _{AMO}) | This study |
| pHpgRNA39 | pPICZ A-(<i>Bleo</i> ^R :: <i>OpURA3</i>)-panARS-- <i>Amp</i> ^R -(P _{TEF1} -HH [*] -gRNA/ pKU80 -HDV-T _{AMO}) | This study |
| pHpgRNA42 | pPICZ A-(<i>Bleo</i> ^R :: <i>OpURA3</i>)-panARS- <i>Amp</i> ^R -(P _{TEF1} -HH [*] -gRNA/ FAA1 -HDV-T _{AMO}) | This study |
| pHpgRNA43 | pPICZ A-(<i>Bleo</i> ^R :: <i>OpURA3</i>)-panARS- <i>Amp</i> ^R -(P _{TEF1} -HH [*] -gRNA/ POX1 -HDV-T _{AMO}) | This study |
| pHpgRNA50 | pPICZ A-(<i>Bleo</i> ^R :: <i>OpURA3</i>)-panARS- <i>Amp</i> ^R -(P _{TEF1} -HH [*] -gRNA/ LSC2 -HDV-T _{AMO}) | This study |
| pHpgRNA60 | pPICZ A-(<i>Bleo</i> ^R :: <i>OpURA3</i>)-panARS- <i>Amp</i> ^R -(P _{DAS1} -HH [*] -gRNA/ LSC2 -HDV-T _{AMO}) | This study |
| pHpgRNA61 | pPICZ A-(<i>Bleo</i> ^R :: <i>OpURA3</i>)-panARS- <i>Amp</i> ^R -(P _{TEF1} -HH [*] -gRNA/ RPAdn -HDV-T _{AMO}) | This study |

pHpgRNA62 pPICZ A-(*Bleo^R*::*OpURA3*)-panARS-*Amp^R*-(*P_{TEF1}*-HH**-gRNA*/***KU80dn*** This study
-HDV-T_{AMO})

^aExpressed genes, or targeted genes, are all indicated in bold.

^bHH represents for variable N6 sequence, and HH* for constant N6 sequence.

Table S5 DNA fragments used in this study by OE-PCR, related to Transparent Methods.

| No. | Fragment name | Primer Fw | Primer Rv | ^a Template | Size |
|-----|---|-----------|-----------|-------------------------|---------|
| 1 | <i>FAA1up-FAA1dn</i> | p154 | p157 | 1, 2 | 2139 bp |
| 2 | <i>POX1up-POX1dn</i> | p160 | p163 | 3, 4 | 2199 bp |
| 3 | <i>KU80up-P_{OpMET3}-KU80dn</i> | p167 | p139 | 5, 7, 6 | 2623 bp |
| 4 | <i>URA3up-URA3dn</i> | p13 | p18 | 8, 9 | 2195 bp |
| 5 | <i>LSC2up-LSC2dn</i> | p243 | p246 | 10, 11 | 2000 bp |
| 6 | <i>KU80up-KU80dn</i> | p249 | p252 | 12, 13 | 2016 bp |
| 7 | <i>RPA-GFP</i> | p321 | p328 | 14, 15, 16, 17 | 3291 bp |
| 8 | <i>KU80-RFP</i> | p330 | p337 | 18, 19, 20, 21 | 3359 bp |
| 9 | <i>RAD51</i> | p199 | p106 | 22, 23, 24 | 2542 bp |
| 10 | <i>RAD51-RAD52</i> | p199 | p208 | RAD51, 25, 26, 27 | 5604 bp |
| 11 | <i>RAD51-SAE2</i> | p199 | p213 | RAD51, 28, 29 | 4606 bp |
| 12 | <i>RAD51-RAD52-SAE2</i> | p199 | p213 | RAD51- RAD52, 30, 29 | 7668 bp |
| 13 | <i>OpRAD51</i> | p236 | p237 | 31, 32, 33 | 2641 bp |
| 14 | <i>OpRAD52</i> | p236 | p237 | 34, 35, 36 | 2698 bp |
| 15 | <i>P_{CAT1}-npgA-MmCAR</i> | p352 | p353 | 37, 38 | 6277 bp |
| 16 | <i>P_{AOX1}-P_{FGH}-ADH5-FaCpAR</i> | p356 | p363 | 39, 40, 41, 42 | 5785 bp |
| 17 | I part | p160 | p163 | 43, 44, 45 | 7714 bp |
| 18 | II pars-F2 | p370 | p163 | 46, 47 | 3379 bp |
| 19 | III parts-F1 | p160 | p448 | 48, 49, 50 | 4015 bp |
| 20 | IV parts-F1 | p160 | p380 | 43, 51, 52 | 4360 bp |

^aNumbers referred to Table S3.

Table S6 Synthesized sequences of gRNA blocks, related to Transparent Methods.

| Name | Sequence |
|---|---|
| HA(P _{TEF1})-HH-gRNA/ADE2-HDV-HA(T _{AMO}) | <p>TATAAAGAGGAGACATTCCCACATGAGATTTTTCTGATCTTTAATTAGTACATTTCGTA (upstream homologous arm sequence for plasmid construction)</p> <p>TCAAGC (variable N6 sequence)</p> <p>CTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCTCAGAT (HH)</p> <p>GCTTGAAACCCACACGCGT (20 bp spacer targeting sequence)</p> <p>GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCTTTT (stgRNA2)</p> <p>GGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGAC (HDV)</p> <p>GTATATAGTACACGACAATCTAGTAATCTCCACTATTGACGAGCTCGTCGA</p> <p>ACTGCGAAA (downstream homologous arm sequence for plasmid construction)</p> |
| HA(P _{TEF1})-HH-gRNA/Ku80-1-HDV-HA(T _{AMO}) | <p>TATAAAGAGGAGACATTCCCACATGAGATTTTTCTGATCTTTAATTAGTACATTTCGTA</p> <p>ACGATGCTGATGAGTCCGTGAGGACGAAACGAGTAAGCTCGTCTCAGAT</p> <p>CATCGTTCTGCAGAAGATCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCAGTCGGTGCTTTTGGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGACGTATATAGTACACGACAATCTAGTAATCTCCACTATTGACGAGCTCGTCAACTGCGAAA</p> |

Table S7 Fragments used for *in vivo* assembly of plasmids, related to Figure 6.

| Fragment | F1 | F2 | F3 | F4 | F5 | F1+F2 | F3+F4 |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Primers | p302+p380 | p377+p376 | p385+p369 | p370+p303 | p386+p387 | p302+p376 | p385+p303 |
| 3 parts | / | / | / | / | 6566 bp | 5794 bp | 7890 bp |
| 4 parts | / | / | 4694 bp | 3785 bp | 6566 bp | 5794 bp | / |
| 5parts | 3683 bp | 3665 bp | 4694 bp | 3785 bp | 6566 bp | / | / |

Table S8 Fragments used for *in vivo* pathway assembly for fatty alcohol production, related to Figure 7.

| | F1 | | F2 | | F3 | | F4 | |
|-----------|---------------|------------|---------------|------------|---------------|------------|---------------|---------|
| | Primer | Size | Primer | Size | Primer | Size | Primer | Size |
| I part | p160, p163 | 7714 bp | | | | | | |
| II parts | p160, p369 | 5916 bp | p370, p163 | 3379 bp | | | | |
| III parts | p160, p448 | 4015 bp | p449, p369 | 4828 bp | p370, p163 | 3379 bp | | |
| IV parts | p160, p380 | 4360 bp | p377, p448 | 2927 bp | p449, p369 | 4828 bp | p370, p163 | 3379 bp |

Table S9 *O. polymorpha* codon-optimized sequences of gene *MmCAR* and *npgA*, related to Transparent Methods.

| | |
|----------|--|
| O | ATGTCTCCAATTACCCGCGAGGAGAGACTGGAACGCAGAATCCAAGACTTGTAC |
| M | GCCAATGACCCACAGTTCGCCGCCGCCAAGCCAGCCACTGCTATCACCCGCCGC |
| m | CATTGAGAGACCGGGACTGCCACTGCCACAGATCATCGAGACCGTTATGACCGG |
| C | CTATGCCGACAGACCAGCCTTGGCCAGAGATCTGTTGAGTTTCGTGACCGACGC |
| A | TGGAACCTGGCCATACCACCTTGAGATTGCTGCCTCACTTCGAGACCATCTCGTAC |
| R | GGAGAGCTGTGGGACCGCATTCTGCTCTGGCTGACGTGCTGTCTACCGAGCAA |
| | ACCGTTAAACCGGGCGATCGCGTGTGCTTGTGGGCTTCAACTCGGTTGACTAC |
| | GCCACCATCGACATGACTCTGGCTAGATTGGGCGCTGTTGCCGTTCCACTGCAG |
| | ACCTCTGCCGCTATCACTCAGTTGCAGCCTATCGTGGCTGAGACTCAGCCTACC |
| | ATGATTGCCGCCTCTGTTGATGCTCTGGCCGATGCCACCGAGCTGGCTCTGTCT |
| | GGACAAACCGCCACCAGAGTGTGTTTTTCGACCACCACAGACAAGTTGACGCT |
| | CACAGAGCCGCCGTTGAGTCGGCCAGAGAGAGATTGGCCGGCTCTGCTGTGGT |
| | TGAGACTCTGGCCGAGGCTATTGCTAGAGGCGATGTGCCTAGAGGAGCCTCGG |
| | CCGGATCTGCTCCGGGCACCGACGTTTTTCGACGACTCGCTGGCTCTGCTGATCT |
| | ACACCTCGGGCTCGACCGAGCCCCAAAAGGAGCCATGTACCCACGCCGCAAT |
| | GTGGCCACCTTTTGGAGAAAGCGCACTTGGTTTGAAGGAGGATACGAGCCATCG |
| | ATTACTCTGAACCTCATGCCAATGTCGCACGTGATGGGCCGCCAGATTCTGTATG |
| | GAACCTCTGTGCAACGGCGGCACTGCCTATTTTGTGGCAAGTCGGATCTGTGCA |
| | CTCTGTTTGAGGATCTGGCCTTGGTTAGACCTACCGAGTTGACTTTTCGTGCCTCG |
| | CGTTTGGGACATGGTGTTCGACGAGTTCAGTCTGAAGTGGACAGAAGACTGGT |
| | GGATGGAGCCGACCGCGTTGCTCTGGAAGCCCAAGTGAAGGCCGAGATCAGAA |
| | ACGACGTGCTGGGAGGCAGATATACCTCGGCTCTGACTGGATCTGCCCAATCT |
| | CGGACGAGATGAAAGCTTGGGTTGAAGAGCTGCTGGACATGCATCTGGTGGAG |
| | GGATACGGATCGACCGAAGCCGGCATGATTCTGATTGATGGCGCCATTTCGAGA |
| | CCAGCCGTGCTGGACTATAAGCTGGTGGATGTGCCAGATCTGGGCTACTTTCTG |
| | ACCGATCGCCACACCCAAGAGGCGAGCTGCTGGTAAAACCGACTCGCTGTTCC |
| | CGGGGCTACTACCAACGCGCCGAGGTGACTGCCGACGTTTTTCGACGCTGACGG |
| | CTTCTATAGAACC GGCGATATCATGGCCGAGGTTGGACCAGAGCAGTTCGTGTA |
| | TCTGGATCGCCGCAACAACGTGTTGAAGCTGTGCAAGGCCGAGTTTGTACCCT |
| | GTGCAAGCTGGAGGCTGTGTTCCGAGACTCGCCTCTGGTGCGCCAGATCTACAT |
| | CTACGGAAACTCTGCCCGCGCCTATCTGTTGGCCGTTATCGTGCCAACCCAAGA |
| | GGCTCTGGACGCCGTGCCAGTTGAGGAACTGAAAGCCCGCTTGGGAGACTCTCT |
| | GCAAGAAGTGGCTAAGGCCCGCCGCTTGAATCGTACGAGATCCCACGCGACTT |
| | CATCATTGAGACCACCCCATGGACTCTGGAGAACGGACTGCTGACCGGCATCAG |
| | AAAGCTGGCCAGACCTCAGCTGAAGAAGCACTACGGCGAGCTGTTGGAGCAGAT |
| | TACACTGATCTGGCCACGGCCAAAGCCGACGAACTGAGATCTCTGCGCCAATC |
| | GGGAGCTGACGCTCCAGTGTGTTGACCGTGTGTAGAGCCGCTGCTGCCTTGC |
| | TGGGCGGATCTGCTTCGGACGTGCAACCAGACGCCATTTTACCGACTTGGGAG |
| | GAGACTCGCTGTGGCTCTGTGTTACCAATCTGTTGCACGAGATCTTCGACAT |
| | TGAGGTTCCAGTGGGCGTGATTGTGTCTCCAGCTAACGATCTGCAAGCTCTGGC |
| | TGACTACGTTGAAGCTGCCCGCAAACCGGGCTCTTCGAGACCTACCTTCGCTTC |
| | TGTGCATGGCGCCTCGAACGGCCAAGTGACCGAAGTTCACGCTGGCGATCTGTC |
| | TCTGGACAAATTCATCGACGCCGCTACTCTGGCCGAGGCTCCAAGACTGCCAGC |
| | TGCCAACACCCAAGTGCACACTGTGCTGTTGACCGGAGCCACCGGCTTTCTGGG |
| | CAGATATCTGGCTCTGGAGTGGCTGGAGAGAATGGATCTGGTGGACGGCAAGCT |
| | GATCTGTCTGGTTCCGCGCAAATCTGATACCGAGGCTCGCGCCAGACTGGACAA |
| | GACCTTTGACTCGGGCGACCCAGAGCTGTTGGCTCACTACCGCGCTCTGGCTGG |
| | AGACCACTTGAAGTGTTGGCCGGCGATAAAGGCGAGGCCGACTTGGGATTGG |
| | ATAGACAGACTTGGCAGAGACTGGCCGACACCGTTGACTTGATTGTGGACCCAG |
| | CCGCTCTGGTTAACCACGTGCTGCCTTATTCGCAGCTGTTTGGACCTAACGCTCT |
| | GGGAACCTGCCGAGCTGCTGAGACTGGCTCTGACCTCGAAGATCAAACCATACTC |
| | GTACACTTCGACCATCGGCGTGGCCGACCAGATTCTCCATCGGCCTTCACCGA |
| | GGATGCCGACATCAGAGTGATTTCCGCCACTAGAGCCGTGGACGACTCTTACGC |
| | CAATGGCTACTCGAACTCGAAGTGGGCCGGAGAGGTTCTGCTGCGCGAAGCTCA |
| | TGATCTGTGCGGACTGCCAGTTGCTGTGTTCCGCTGCGACATGATTCTGGCTGA |
| | CACCACTTGGGCTGGACAGTTGAACGTGCCAGACATGTTCACTCGCATGATCTT |
| | GTGCTGGCTGCCACTGGAATCGCCCCGGGCTCGTTCTATGAGCTGGCTGCTGA |
| | CGGCGCCAGACAGAGAGCCATTACGACGGCTTGCCAGTGGAGTTCATCGCTG |

| | |
|---|--|
| | <p>AGGCCATCTCGACCTTGGGCGCTCAGTCTCAAGACGGATTTACACCTACCACG TGATGAACCCTTACGACGACGGCATTGGACTGGACGAGTTTGTGACTGGCTGA ACGAATCGGGCTGCCAATTCAGCGCATTGCCGACTACGGAGACTGGTTGCAGA GATTCGAGACCGCCTTGAGAGCTCTGCCAGATAGACAGCGCCACTCTTCTTTGC TGCCACTGCTGCACAACCTACCGCCAGCCAGAAAGACCAGTGCGCGGCTCTATCG CCCCAACTGACAGATTTAGAGCCGCCGTGCAAGAGGCTAAGATCGGCCAGACA AAGACATTCCACACGTGGGCGCCCAATTATCGTGAAGTACGTGTCGGATCTGA GACTGTTGGGACTGCTGGGAGGAGGATCTGCTGCTGTGAAGCTGTCGCAAGCC AAGTCGAAGCTGTAA</p> |
| <p>O N P G A</p> | <p>ATGGTGCAAGATACCTCTTCGGCTTCGACCTCGCCTATTCTGACTCGCTGGTACA TCGACACCAGACCATTGACCGCCTCTACTGCTGCTTTGCCTTTGTTGGAGACTCT GCAACCAGCCGACCAGATCTCTGTGCAGAAGTATTACCACTTGAAGGACAAGCA TATGTCGCTGGCCTCGAATCTGCTGAAGTATCTGTTTCGTGCACCGCAACTGTCGC ATCCCATGGTCGTCTATCGTTATCTCGCGCACCCCAGACCCACACCGCAGACCA TGCTATATCCCTCCTTCTGGCTCGCAAGAGGACTCGTTCAAGGACGGCTACACC GGAATCAATGTGGAGTTCAACGTGTGCGACCAAGCCTCGATGGTGGCTATCGCC GGAACCGCCTTACCCCTAACTCGGGCGGCGACTCTAAGCTGAAGCCAGAGGT GGGCATCGACATCACTTGCGTGAACGAACGCCAAGGCAGAAATGGCGAAGAAC GCTCGCTGGAGTCGCTGCGCCAGTACATTGATATCTTCTCGGAGGTGTTCTCTAC CGCTGAGATGGCCAACATTCGCAGATTGGACGGCGTTTCTTCGTCTTCTCTGTCT GCCGATCGCTTGGTGGATTACGGCTACAGACTGTTCTACACTTACTGGGCTCTGA AGGAGGCCTACATCAAGATGACTGGCGAGGCCTTGCTGGCTCCTTGGTTGCGCG AGTTGGAGTTCTCTAACGTGGTTGCTCCAGCCGCTGTGGCTGAGTCGGGCGATT CGGCTGGCGACTTTGGCGAGCCTTATACCGGCGTGCGCACCACTCTGTATAAAA ATCTGGTGGAAGATGTGCGCATTGAGGTGGCCGCTCTGGGCGGAGATTATCTGT TCGCTACCGCCGCCAGAGGAGGAGGAATCGGAGCCTCTTCTAGACCGGGCGGC GGACCAGACGGATCGGGCATCAGATCGCAAGACCCATGGCGCCCTTTTAAAAAG CTGGACATCGAGAGAGATATCCAGCCATGCGCCACCGGCGTGTGTAAGTCTTG TCGGGCGGCGGCTCTTCGAAGTTGTAA</p> |

Transparent Methods

Strains and cultivation

All strains used in this study were listed in [Table S2](#). *Ogataea polymorpha* NCYC 495 *leu1.1* was purchased from China General Microbiological Culture Collection Center (CGMCC), and other strains were stored in our lab or constructed in this study. Without otherwise specified, yeast strains were cultivated in YPD medium containing 20 g/L glucose, 20 g/L peptone, and 10 g/L yeast extract. For screening transformants, synthetic dropout (SD) media (6.7 g/L yeast nitrogen base without amino acids and 20 g/L glucose) was utilized with supplementing essential amino acids. To remove gRNA plasmids, transformants were plated on SC plates with 1 g/L 5-fluoroorotic acid (5-FOA). Strains with an integrated Cas9 protein were selected on YPD containing 100 mg/L Zeocin. Delft basic salt medium (2.5 g/L (NH₄)₂SO₄, 14.4 g/L KH₂PO₄, 0.5 g/L MgSO₄•7H₂O, 1 mL/L Vitamin solution, 2 mL/L Trace metal solution) (Zhou et al., 2016a; Zhou et al., 2016b), was used for cell cultivation and fermentation with glucose or methanol as carbon sources. *Escherichia coli* DH5α was grown in LB medium (10 g/L tryptone, 10 g/L NaCl, and 5 g/L yeast extract), and for plasmid construction, LB medium with the specific antibiotics, such as ampicillin (100 mg/L) and Zeocin (25 mg/L) was adopted. All strains were cultivated at 37°C, 220 rpm in a shake incubator (Zhichu Shaker ZQZY-CS8). Pre-cultures were cultivated in 15 mL tube with a working volume of 3 mL, and cells for transformation and fermentation were both cultured in 100 mL flasks with a working volume of 20 mL.

Scarless gene deletion and integration

All plasmids constructed and primers in this study were listed in [Table S3](#) and [Table S4](#), respectively. DNA manipulation like PCR amplification, enzyme digestion, and ligation was performed by standard procedure. Donor DNA and gene expression cassette were constructed by overlap extension PCR (Zhou et al., 2012) ([Table S5](#)). For scarless gene deletion, upstream homologous arm was directly linked with downstream homologous arm via overlap extension PCR. The purified DNA fragments with an amount of 500 ng was applied in transformation, together with a specific gRNA plasmid. Similarly, for the site-specific integration, each DNA fragment like upstream and downstream homologous arms, promoter, gene, and terminator was prepared to construct the donor DNA by overlap extension PCR. Other steps were the same with those for scarless gene deletion. *E. coli* was transformed as previously described (Cohen et al., 1972). *O. polymorpha* was transformed by either LiAc/ssDNA method (Faber et al., 1994) (experiments in [Figure 2](#), [Figure 3](#), [Figure 4](#), [Figure S3](#), [Figure S4](#), [Figure S5](#) and [Figure S8](#)), or electroporation (Qian et al., 2009) (experiments in [Figure 5](#), [Figure 6](#) and [Figure 7](#)). All DNA sequences are referred to the genome of *O. polymorpha* NCYC 495 *leu1.1* at NCBI database. In particular, gene *ADE2* (GenBank: NW_017264700.1), *KU80* (GenBank: NW_017264699), *POX1* (GenBank: NW_017264699), *FAA1* (GenBank: NW_017264698), *LSC2* (GenBank: NW_017264704) were adopted for genetic manipulation.

CAS9 gene integration

For genome integration of *CAS9*, a completed Cas9 expression cassette was constructed, containing P_{GAP} promoter from *K. phaffii*, the human codon-optimized *CAS9* gene, and *AOX1* terminator ([Figure S1A](#)). For transformant selection, Zeocin resistance markers (*Bleo*^R) under control of P_{TEF1} promoter from *K. phaffii* in the commercial plasmid pPICZ A was introduced. The resulting pHp03 was linearized by *Apal*, and integrated at *OpURA3* locus. To achieve the fusion of Cas9 and Pds1 from *S. cerevisiae*, pHp26 was constructed. The linearized plasmid by *Apal* was also integrated at *OpURA3* locus. Transformants were verified by PCR and sequencing, and the correct transcription and expression of *CAS9* gene was tested by RT-PCR ([Figure S1A](#)).

Construction of gRNA expression plasmids

Episomal gRNA expression plasmids were constructed based on pPICZ A, including an autonomous replication start from *K. lactis* (panARS), resistance marker genes (*LEU2* gene from *S. cerevisiae*, or *URA3* gene from *O. polymorpha*, under control of their own promoter and terminator), and the gRNA expression cassettes. As shown in [Table S3](#), some initial plasmids for gRNA expression were selected by *ScLEU2*, which turned out to be difficult in plasmid

removal. In this case, *O. polymorpha* with a disrupted *URA3* gene was obtained (495-3 Δ ura3) as control strain, and resistance marker was replaced by *OpURA3* gene in the following experiments, which can be dropped-out by 5-FOA. gRNA was expressed under the bidirectional promoter of RNA pol III. A promoter of *tRNA^{CU6}* from *O. polymorpha* was amplified with primers p19-p20 and p23-p19, which was then placed at each end of the gRNA cassettes, respectively. Subsequently, ptRNA^{CU6}-1 fragment was fused by Amp-1 fragment (p21 and p36) with primers p24 and p36, yielding ptRNA-Amp-1 fragment. Similarly, ptRNA-Amp-2 fragment was obtained by combining ptRNA^{CU6}-2 fragment and Amp-1 fragment (p37 and p22), using primers p37 and p25. Vector backbone was obtained by PCR amplification using primers p26 and p27. At last, two fused fragments and the backbone fragment were assembled by Gibson Assembly[®] Master Mix according to its manipulating instruction. The products were transformed into *Escherichia coli* DH5 α , achieving plasmid pHpgRNA7. Plasmid pHpgRNA9 was constructed to replace the terminator of gRNA with T_{SNR6}. Terminator T_{SNR6} was amplified with primers of p95 and p96, together with fragment Amp-1 (p97-p36) and Amp-2 (p37-p98). Subsequently, fragment tSNR6-Amp-1 and Amp-2-tSNR6 were fused with primers of p95-p36 and p37-p95, respectively. These two final fragments were cyclized by Gibson Assembly[®] Master Mix, with the vector backbone that was amplified with primer p94.

gRNA expression controlled by RNA pol II was also constructed as following procedure. gRNA blocks HA(pTEF1)-HH-gRNA/ADE2-HDV-HA(AMOt) and HA(pTEF1)-HH-gRNA/Ku80-1-HDV-HA(AMOt) were synthesized by Sangon Biotech Co., Ltd (Table S6). Taking this as a template, gRNA cassettes were amplified with primers of p116 and p117, which was subsequently fused with promoter P_{TEF1} (p103-p104) and terminator (p105-p106). The resulting fragment was linked by Gibson Assembly[®] Master Mix, with the vector backbone that was amplified with primers p267 and p268. Correct plasmids were named pHpgRNA11 and pHpgRNA 16, respectively.

To replace the variable N6 sequence with the constant sequence of ATCTGA, a DNA fragment was constructed by fusing promoter P_{TEF1} (amplified with p103-p104) and fragment gRNA-AMOt (amplified with p126-p102), and then integrating into plasmid pHpgRNA13. While substituting 20 bp gRNA spacer targeting other genes or sites, gRNA expression cassettes were fused by gRNA fragment 1 (p101 and p127) and gRNA fragment 2 (pX and p102), which was subsequently assembled with backbone fragment. A series of gRNA plasmids were successfully constructed in Table S3. The most important gRNA expression plasmids pHpgRNA13 and pHpgRNA50 were deposited to Addgene (Yongjin Zhou, 78587).

Overexpression of HR-related proteins

Expression cassettes were constructed by OE-PCR according to Table S5, and promoters with intermediate strength were chosen to avoid a negative effect on cell growth. The obtained fragments were inserted into plasmid backbone via simple enzyme digestion. The resulting plasmids were named as pHp31-pHp34 and pHp44-pHp45. In particular, pHp46-pHp48 were obtained by enzyme digestion and plasmid cyclization based on pHp32-pHp34. The linearized plasmids were introduced into strain 493-3 Δ ura3 by single crossover, resulting the corresponding strains in Table S2.

Down-regulation of KU80

To dynamically down-regulate *KU80* gene, its native promoter was replaced by the promoter P_{OpMET3}. To fulfill this *in-situ* replacement, a gRNA targeting 5'-UTR of *KU80* gene and the corresponding donor DNA with the HA length of 1000 bp were both constructed. Five concentrations of methionine, 0, 1.7, 2.5, 5.0, 10 mM, were added in YPD medium that is used for recovery procedure during transformation, and/or the final screening plates to repress *KU80* gene.

Competition assay between HR and NHEJ

To verify our proposed hypothesis, RPA and Ku80 were selected as the targets to represent for HR and NHEJ, respectively, and their relative abundance was detected by fluorescence intensities of GFP and RFP. In this case, the fusion of RPF-GFP and Ku80-RFP were both achieved in strains of 493-3 Δ ura3, Ku80-dw, and y34, respectively. gRNA plasmids and donor DNA were listed and constructed as Table S3 and Table S5. For well controlling DSBs formation, an inducible gRNA plasmid targeting *LSC2* gene was constructed by replacing constitutive P_{TEF1} promoter with methanol-induced P_{DAS1} promoter (pHpgRNA60). Afterwards, pHpgRNA60

was transformed into above strains with fused proteins, and the existence of plasmids and unedited *LSC2* sites were both confirmed by PCR and sequencing. To test the fluorescence intensities of GFP and RFP, cells were pre-cultured in Delft basic salt medium with 20 g/L glucose, and then switched to Delft basic salt medium with 10 g/L methanol, or 20 g/L glucose at an initial OD₆₀₀ of 0.2. For repressing *KU80* gene expression, 1.7 mM methionine was added in the medium when cultivating strain Ku80-dw. After 24 h cultivation, cells were collected to detect both fluorescence and OD₆₀₀ by Bio-Tek Synergy H1 multimode microplate reader. The fluorescence intensity of GFP was measured with excitation at 485 nm and emission at 528 nm, and RFP was detected with excitation at 580 nm and emission at 614 nm. Three parallel experiments were adopted, and duplicate analysis was applied to each sample. Fluorescence was finally normalized to the corresponding cell density (OD₆₀₀).

Quantitative real-time PCR

To test expression levels of gene *KU80*, *ScRAD51*, *ScRAD52*, *ScSAE2*, and *OpRAD52*, strains wild-type (495-3 Δ ura3), Ku80-dw, y34, and y45 were cultivated in YPD medium at 37°C, 220 rpm for 24 h, and in particular, strain Ku80-dw was grown in YPD medium w/o 10 mM methionine. To test the expression level of gene *KU80* at different methionine concentrations. Strain Ku80-dw was pre-cultured in YPD medium, and then inoculated into in Delft basic salt medium containing 0, 0.05, 0.1, 0.5, 1.0, 1.7, 2.5, 5, and 10 mM methionine, respectively. Cell were harvested after 8-10 h of cultivation at 37°C, 220 rpm.

Cells were collected, and washed twice with ddH₂O, and the total RNA was extracted by RNAsimple Total RNA Kit (DP419, TIANGEN, Beijing China). The total RNA of each sample (1 μ g) was reversely transcribed to cDNA using the PrimeScript[®] RT reagent Kit (Takara Bio Inc.) according to the manufacturer's protocol. A two-step PCR reaction was employed, and the system and condition were described in the manufacturer's protocol of SYBR[®] Premix Ex Taq[™] II (Takara Bio Inc.). Gene *GAPDH* and actin was adopted as the endogenous reference gene, and the data analysis was determined by the method of $2^{-\Delta\Delta C_T}$ as described by Livak and Schmittgen (Livak and Schmittgen, 2001). Primers were listed in Table S4, and all strains with three biologically independent parallel samples were adopted to guarantee the reproducibility of all the results.

Multiple fragment assembly in episomal plasmid or genome

In vitro construction of episomal plasmid (pHp63) was achieved by OE-PCR and Gibson Assembly[®] Master Mix (Table S5). Genes for fatty alcohol production (*MmCAR* and its co-factor *nggA*, *ADH5*, and *FaCoAR*) was obtained from previous studies (Zhou et al., 2016a; Zhou et al., 2016b), and methanol-inducible promoters (*P_{AOX1}*, *P_{CAT1}*, *P_{FGH}*, and *P_{DAS1}*) where applied to control gene expression. *In vivo* assembly of pHp63 was performed in strain y34 with an enhanced HR activity. Fragments used were obtained by PCR as shown in Table S7. The intact pHp63 was divided into 3-5 parts, and the homologous region was set as promoters, or terminators with the length of ~500 bp. Concentrations of each fragment used for transformation were set as around 500 ng. Transformants were verified by PCR at the connected region. Genes for fatty alcohol production were integrated into *POX1* site in strain y34 as each complete expression cassette (Figure S9). More than one gene integrated was achieved by *in vivo* assembly with the 500-1000 bp HA at regions of promoters, or terminators. Fragments used were obtained by PCR as shown in Table S8. Sequences of gene *MmCAR* and *nggA* were codon-optimized based on *O. polymorpha* as shown in Table S9. Concentrations of gRNA plasmid and I part fragment used for transformation were set as around 500 ng, and concentrations of other fragments were set as the same moles as I part fragment.

Quantitation of HR rate versus NHEJ strength

DNA repair pattern was tested by PCR. HR rate was defined as the ratio of positive colony to total picked colony. For scarless gene deletion, a HR positive colony would show a shorter DNA band at both upstream and downstream as shown in Figure 3A and Figure S3A, and upstream and downstream were also both identified while integrating gene cassettes at a specific site (Figure 5A). In particular, for *in vivo* plasmid assembly as shown in Figure 6 G-I, colony PCR was conducted to test the correct link between neighboring DNA fragments, and a positive colony was defined as the transformants containing all identified linkers.

To calculate HR rate and CFU/OD₆₀₀, each strain with two biologically independent parallel

samples was adopted to conduct transformation experiments. During the transformation, OD₆₀₀ of competent cells were measured to calculate the CFU/OD₆₀₀. After around 2-3 days of cultivation, total colony was counted, and over 20 colony was picked into selective media with 300 µL working volume in 1.5 mL Eppendorf tubes. 100 µL cultures were used to extract crude genome for PCR according to reported protocols by Lööke et al. (Lööke et al., 2011) The HR positive colony was identified, determined as shown in [Figure 3A](#), [Figure S3A](#), and [Figure S7A](#) to calculate HR rate.

Fatty alcohol production

For fatty alcohol production from glucose, Delft basic salt medium contain 20 g/L glucose was used as fermentation medium in 100 mL flasks with the working volume of 20 mL. Correct transformants were pre-cultured in 3 mL Delft basic salt medium at 37°C, 220 rpm for 24 h, and the pre-cultures were transferred to fermentation medium, and cultivated at 37°C, 220 rpm for 96 h. For fatty alcohol production from methanol, the strain c1-3-12 (*MmCAR*, *npgA*, *ADH5*, *Δhfd1*) were cultivated in methanol-containing media, including Delft basic salt medium+5 g/L methanol and YP medium+5 g/L methanol. 5 g/L of methanol was supplemented at 24 h and 48 h. Samples were taken at 96 h for quantification of biomass and fatty alcohols.

Extraction and quantification of fatty alcohol were performed as described previously (Zhou et al., 2016a; Zhou et al., 2016b) with some modifications. 2 mL cell cultures were collected in 15 mL glass tubes and freeze dried for 48 h (SCIENTZ-18N). The cell pellets were pounded to pieces and added 4 mL chloroform:methanol (v/v, 2:1) containing 10 mg/L pentadecanol as internal standards. Then fatty alcohols were extracted by using microwave digestion/extraction system (Milestone DRN 41). The samples were vortexed vigorously and placed in the microwave reaction vessel containing 10 mL of Mili-Q water then sealed with a TFM screw cap. The temperature programming of microwave extraction was ramped to 60 °C (from room temperature, using 800 W for 24 vessels) within 6 min and kept constant for 10 min. After the samples were cooled down to room temperature, 1 mL NaCl solution (0.73%, w/v) was added and then the samples were vortexed vigorously. Thereafter, the samples were centrifuged at 1000 g for 10 min allowing for phase separation and the organic phase was transferred into a new clean extraction tube. The extracted fractions were dried by using a centrifugal vacuum concentrator (miVac Quattro High, GeneVac). 100 µl BSTFA was added into the vial for silylation (80°C, 30 min) and 400 µl hexane was added. Quantification of silylated fatty alcohols was performed by gas chromatography (Focus GC, ThermoFisher Scientific) equipped with a Zebron ZB-5MS GUARDIAN capillary column (30 m * 0.25 mm * 0.25 µm, Phenomenex). The GC program was set as follows: initial temperature of 50°C, hold for 2.5 min; then ramp to 230°C at a rate of 15°C per min and hold for 2 min; finally reached 325°C at a rate of 25°C per min and hold for 4 min. The temperature of inlet and detection were kept at 250°C and 300 °C, respectively. The flow rate of the carrier gas (nitrogen) was set to 1.0 mL/min.

Statistical Analysis

Continuous variables are expressed as mean of two or three biologically independent samples with displayed data-points. In all cases, significance was defined as $p \leq 0.05$. Statistical analysis was carried out using paired t-test with heteroscedasticity testing method by Office Excel Software.

Reference

Cohen, S.N., Chang, A.C., and Hsu, L. (1972). Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. *P. Natl. Acad. Sci.* *69*, 2110-2114.

Faber, K.N., Haima, P., Harder, W., Veenhuis, M., and Geert, A. (1994). Highly-efficient electrotransformation of the yeast *Hansenula polymorpha*. *Curr. Genet.* *25*, 305-310.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2^{(-Delta Delta C(T))} Method. *Methods* *25*, 402.

Löoke, M., Kristjuhan, K., and Kristjuhan, A. (2011). Extraction of genomic DNA from yeasts for PCR-based applications. *Biotechniques* *50*, 325-328.

Qian, W., Song, H., Liu, Y., Zhang, C., Niu, Z., Wang, H., and Qiu, B. (2009). Improved gene disruption method and Cre-loxP mutant system for multiple gene disruptions in *Hansenula polymorpha*. *J. Microbiol. Meth.* *79*, 253-259.

Zhou, Y.J., Gao, W., Rong, Q., Jin, G., Chu, H., Liu, W., Yang, W., Zhu, Z., Li, G., and Zhu, G. (2012). Modular pathway engineering of diterpenoid synthases and the mevalonic acid pathway for miltiradiene production. *J. Am. Chem. Soc.* *134*, 3234-3241.