J Ind Microbiol Biotechnol
DOI 10.1007/s10295-016-1776-0

Combinatorial metabolic pathway assembly in the yeast genome with RNA-guided Cas9

Steve F. EauClaire[1] · Jianzhong Zhang[1] · Corban Gregory Rivera[1] · Lixuan L. Huang[1]

Received: 9 February 2016 / Accepted: 11 April 2016

Abstract The yeast *Saccharomyces cerevisiae* is an important industrial platform for the production of grain and cellulosic ethanol, isobutanol, butanediol, isoprenoids, and other chemicals. The construction of a successful production strain usually involves multiple gene knockouts and chromosomal integration of expression cassettes to redirect the metabolic fluxes for the conversion of sugars and other feed stocks into the desired product. RNA-guided Cas9 based genome editing has been demonstrated in many prokaryotic and eukaryotic hosts including *S. cerevisiae*, in which it has been additionally exploited as a tool for metabolic engineering. To extend the utilization of RNA-guided Cas9 as a metabolic pathway building tool, we demonstrated the direct assembly and chromosomal integration of up to 17 overlapping DNA fragments encoding the beta-carotene biosynthetic pathway. Furthermore, we generated a combinatorial strain library for the beta-carotene biosynthetic pathway, directly integrated into the yeast genome to create a diverse library of strains. This enabled the screening of combinatorial libraries in stable chromosomally integrated strains for rapid improvements of product titers. This combinatorial approach for pathway assembly will significantly accelerate the current speed of metabolic engineering for *S. cerevisiae* as an industrial platform, and increase the number of strains that can be simultaneously evaluated for enzyme screening, expression optimization and protein engineering to achieve the titer, rate and yield necessary for the commercialization of new industrial fermentation products.

Keywords RNA-guided Cas9 · *S. cerevisiae* · Metabolic engineering · β-Carotene biosynthesis · Combinatorial pathway assembly

Introduction

The yeast *Saccharomyces cerevisiae* has been a eukaryotic model organism and an important industrial fermentation host for many decades. It has been used in the production of fuels and chemicals, such as grain and cellulosic ethanol, isobutanol, butanediol, isoprene as well as various higher value-specialty chemicals, nutraceuticals and pharmaceuticals, e.g., artemisinin and amorphadiene, β-carotene, astaxanthin, flavonoids, and short chain fatty acids [4, 17, 20, 32, 35]. Some of the most time consuming steps in the metabolic engineering process are the screening and selection of appropriate enzymes, and optimization of the pathway fluxes to redirect the native metabolic pathways towards high-level production of the target fermentation products. For the construction of stable and robust industrial yeast strains, it is highly desirable to have all the genetic elements for the engineered metabolic pathway integrated into the yeast genome. Although *S. cerevisiae* has an efficient homologous recombination mechanism that allows the integration of linear DNA fragments, the chromosomal integration efficiency is still relatively low, and therefore, depends on the use of selection markers to screen for strains that contain the DNA fragments integrated at the targeted loci [30]. Usually individual gene expression cassettes are pre-assembled via a variety of approaches such

Electronic supplementary material The online version of this article (doi:10.1007/s10295-016-1776-0) contains supplementary material, which is available to authorized users.

✉ Lixuan L. Huang
Lisa.L.Huang@dupont.com

[1] Industrial Biosciences, E. I. Du Pont Company, Wilmington, DE, USA

Published online: 02 May 2016

as subcloning ([21], SOEing PCR [27], Gibson assembly [11], and Golden Gate assembly [24]. These expression cassettes are further combined into multiple gene expression cassettes in vitro or subcloned in vivo to encode a multi-step enzyme pathway, as well as a positive selection marker such as a *URA3* or kanamycin resistance gene (e.g., KanMX) before integrating into a target locus in the yeast genome [1]. Integration of a multi-step metabolic pathway may involve a lengthy process with several rounds of chromosome integration and recycling of selection markers.

A variety of approaches have been developed for modular DNA assemblies taking advantage of the homologous recombination in *S. cerevisiae*. Gibson et al. carried out a one-step assembly of 25 overlapping DNA fragments and created the complete circular 592 kb synthetic *Mycoplasma genitalium* genome as a self-replicating plasmid in yeast [11]. Shao et al. generated an eight gene cassette (19 kb) either on a plasmid or in yeast genome with high efficiencies (70–100 %) by homologous recombination and multiple copy chromosomal integration into the δ regions of the genome [30]. For the integration of target pathway genes, native homologous recombination efficiency was significantly improved upon the generation of a double strand DNA break in the genome using restriction enzymes such as the homing endonuclease I-SceI, which can be repaired using gene expression cassettes as donor DNAs for homologous recombination [19].

CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated proteins) systems were first identified in bacteria as part of an immunity system towards bacteriophage and plasmids [13]. It has been used to improve the phage resistance of some lactic acid bacteria through natural adaptation and selection [3]. Over the last few years, RNA-guided CRISPR-associated protein (RNA-guided Cas9) technology has been used to introduce gene editing in many organisms, e.g., bacteria, *S. cerevisiae* yeast, the filamentous fungus *Trichoderma reesei*, *Caenorhabditis elegans*, mammalian cells, and plants [3, 5, 8–10, 21, 22]. Dicarlo et al. reported the first example of sgRNA/Cas9 mediated gene disruption in *S. cerevisiae* [7]. Ryan et al. used the sgRNA/Cas9 system to create multiple DNA breaks simultaneously across the genome of diploid or polyploid *S. cerevisiae* strains, while inserting either short dsDNA 'barcodes' or up to three overlapping dsDNA fragments into these dsDNA breaks without the use of a selection marker [29].

Recently, several studies have been published on multiplex metabolic pathway engineering in *S. cerevisiae* using RNA-guided Cas9 systems (summarized in Table 2). Horwitz et al. integrated six DNA constructs encoding 11 genes of the muconic acid biosynthetic pathway simultaneously into 3 genomic loci by co-expression of three guide RNAs targeting the different loci [14]. Jakočiūnas et al. applied a multiplex sgRNA/Cas9 system for simultaneous knockout of five genes [15]; they also integrated three groups of DNA fragments, each with five overlapping DNA sequences at three loci simultaneously via sgRNA/Cas9 [16]. Mans et al. combined *ACS1/ACS2* double gene deletion with the simultaneous in vivo assembly and integration of six overlapping DNA fragments into the *ACS1* locus [23]. Ronda et al. integrated three gene cassettes simultaneously into three intergenic loci using multiplex sgRNA targeting [28]. Tsai et al. reconstructed a xylose utilizing *S. cerevisiae* yeast strain using sgRNA/Cas9, and demonstrated that it had similar performance as a strain constructed through traditional integration approaches using selectable markers [33]. The number of multiplex gene targeting ranges from 2 to 6 loci, with integration cassettes containing ~45–60 bp homology recombination arms between various overlapping integration fragments or with genomic sequences proximal to the sgRNA target site. Assuming the homologous recombination is efficient for the multi-fragment assemblies, the different sgRNA/Cas9 targeting efficiencies are likely due to the intrinsic sgRNA/Cas9 targeting efficiency of each genomic locus.

These recent studies have demonstrated that genome editing facilitated by the homologous recombination system of *S. cerevisiae* has been made more efficient by generating chromosomal double strand breaks using RNA-guided Cas9, allowing high frequency chromosomal integration of multiple DNA fragments. So far it has been shown that a 3–6 piece DNA assembly can be directly integrated into a single target site via homologous recombination, but there has not been any attempt to assemble larger number of DNA fragments in the same genome locus. Although recently a combinatorial approach of transcriptional balancing was applied for the biosynthesis of violacein through a five step pathway [18], it is not known whether multiple combinations of gene expression levels for a given pathway assembly can be simultaneously generated using sgRNA/Cas9. To make the sgRNA/Cas9 tool box more versatile for the assembly and integration of large numbers of DNA fragments in the yeast genome, we demonstrated the efficient homologous recombination-directed assembly of up to 17 DNA fragments in a single transformation step in the *S. cerevisiae* genome via sgRNA/Cas9. As a proof of concept for further optimization of the gene expression levels to improve flux for the engineered metabolic pathway, we generated a small combinatorial library of three carotenoid pathway genes directly in the genome. Such combinatorial strain libraries allow the simultaneous optimization of multiple enzyme expression levels for metabolic pathway assembly.

Materials and methods

Strains, plasmids, growth media and reagents

E. coli strain DH5α (New England Biolabs, Ipswich, MA) and TOP10 (Life Technologies, Grand Island, NY) were

used for routine subcloning and plasmid transformation. *S. cerevisiae* strain *BP1548* was used as the host strain for gene expression and pathway integration. L-canavanine was obtained from Sigma-Aldrich (St. Louis, MO, USA); β-carotene was from MP Biomedicals (Solon, OH, USA).

Yeast strains were cultivated in either YPD (10 g/L yeast extract, 10 g/L peptone, 20 g/L glucose), or synthetic complete (SC) media with 6.7 g/L yeast nitrogen base (DifCo/BD, Franklin Lakes, New Jersey) with an appropriate amino acid drop out mix (Formedium, Norfolk, UK). The canavanine selection medium is SC-Arg-Lys containing 60 mg/L L-canavanine.

The *crtE*, *crtB*, *crtI*, and *crtY* genes were PCR amplified from the carotenoid pathway in *Yarrowia lipolytica* expression plasmids (pYPS102, pYCRTEBI) [31].

PCR products were purified by agarose gel electrophoresis or using DNA Clean and Concentrator Columns (Zymo Research Corporation, Irvine, CA) before being used as templates for the DNA assembly and yeast transformation.

Cas9 target site selection

Initial sgRNA target sites were enumerated across an *S. cerevisiae* S288c reference genome using the pattern N^{20}NGG encompassing the PAM sequence NGG. The candidates were further refined to ensure the uniqueness of the target site in the genome, by eliminating the sequence N^{20}NAG which contains a single base mismatch at the PAM site. As described by DiCarlo et al. [7], duplicate target sites were identified and removed using Bowtie2. Candidate sgRNA target sites were removed if they contained T^6 which can cause premature transcriptional termination. Remaining candidate sgRNA target sites were loaded into a DuPont proprietary genome browser along with the yeast gene model for reference and indexed to enable gene search.

Plasmid construction

The yeast 2μ plasmid-based vector pSE005 was constructed for expression of the Cas9 gene in plasmid pRS426-PDC1-PmeI-SfiI which contains PDC1 promoter, PmeI and SfiI restriction sites, and ILV5 terminator (300 bp of the 3′ downstream UTS sequence of *ILV5* gene) subcloned into the polylinker of pRS426 (2μ vector with *URA3* selection marker) (34). pSE005 contains the *PDC1* promoter, the Cas9 gene with 3′NLS (Nuclear Localization Sequence) based on Cong et al. (5), and the *ILV5* terminator. The Cas9 gene was codon optimized for expression in *Y. lipolytica*, synthesized by Genscript (Piscataway, New Jersey), and PCR amplified using 5′ primer SE004 and 3′ primer SE005 (adding PmeI and SfiI sites, respectively). The DNA sequences of all the primers and gBlock DNAs in this study are included in Table S3 in the Supplementary Files online.

The guide RNA expression cassettes including the *SNR52* promoter and the *SUP4* terminator were synthesized as gBlock DNA by IDT (Integrated DNA Technologies, Coralville, Iowa) and subcloned into pRS423 (2μ vector with *HIS3* selection marker) or pRS424 (2μ vector with *TRP1* marker). pSE008 contains the gBlock DNA encoding for sgRNA.CAN1.Y [7] (Table S4, Supplementary Files) cloned into XhoI and SacI sites in the polylinker of pRS423; pSE028 contains the gBlock DNA for sgRNA.HO.F cloned into the NotI and SacI sites of pRS423. pSE029 contains the the gBlock DNA for sgRNA.HO.G cassette between NotI and SacI sites of pRS423. pSE045 is the *HIS3* marker containing derivative of pSE029, constructed by ligating the NotI-SacI cassette from pSE029 into the polylinker of pRS423.

Yeast transformation

Yeast transformations with plasmids and linear DNAs were performed using the Frozen-EZ yeast transformation II Kit™ (Zymo Research, Irvine, CA, USA) or electroporation. To perform electroporation, overnight yeast cultures were diluted 1:10 in 50 mL fresh YPD or SC-uracil media, grown to an OD600 of 0.8–0.9, and harvested by centrifugation at 700*g* for 5 min. The cells were resuspended in 30 mL of 0.1 M Lithium acetate, 10 mM Dithiothreitol, 10 mM-Tris–HCl, pH 7.5, 1 mM EDTA and incubated at room temperature for 1 h. The cells were harvested again at 700 g for 5 min, washed twice in 40 mL sterile water followed by washing in 10 mL of ice cold 1 M sorbitol, before resuspending in 300 μL of 1 M ice cold sorbitol. To transform DNA by electroporation, 1 μg of DNA (in up to 5 μL) and 50 μL of cells were mixed in microfuge tube and incubated on ice for 5 min. The yeast cells/DNA mixture was added to a 0.1 cm chilled electroporation cuvette, and electroporation was performed at 1.5 kV, 25 μF (GenePulser II, BioRad, Hercules, CA, USA). The cells were incubated in YPD or SC media for 1 h at 30 °C, plated on selective media and incubated at 30 °C for 3 days.

Integration of the β-carotene pathway into the *CAN1* locus

A total of 17 DNA fragments (PCR#1-#17, Table S1, Supplementary Files) were generated by PCR as components for the overlapping assembly of 5 expression cassettes at the *CAN1* locus. Table S1 contains a list of the DNA fragments including template and primer names, the sizes of the PCR products, as well as the overlapping DNA

sequences for each fragment. The carotenoid pathway (*crt*) genes were PCR amplified from plasmids pYPS102 and pYcrtEBI [31]. The five terminator fragments were synthesized as gBlock DNAs and PCR amplified using 5′ primers with 50 bp overlap with the corresponding ORF, and 3′ PCR primers with 50 bp overlap with the unique linker sequence that marks the junction between each terminator and the next promoter for the following gene cassette (L2–L6). These 15 fragments are flanked by two 250 bp fragments, GB10 and GB11, with homology to the sequences upstream and downstream of the sgRNA/Cas9 target site CAN1.Y. All PCR products were purified and DNA concentrations were determined using a NanoDrop ND1000 machine (NanoDrop Products, Wilmington, DE, USA). The PCR fragments were combined in equal molar amounts to a total of 2 μg of DNA in a total volume of 22, and 1 μl was analyzed by electrophoresis. Finally, the combined DNA fragments were lyophilized and redissolved in 8 μL ddH_2O.

Yeast strain *BP1548Cas9* was constructed by transformation of pSE005 into *BP1548*. For the in vivo assembly of the 17 genes of the β-carotene pathway at the *CAN1* locus, 1 μg of the various DNA fragments were combined with 0.14 μg of pSE008 and transformed into *BY1548* by electroporation.

For an alternative assembly approach, the 17 overlapping DNA fragments were pre-assembled in vitro via Gibson assembly using Gibson Assembly Master Mix following manufacture's protocol (New England Biolabs, Ipswich, MA, USA) into 5 sub-assemblies of the promoter-gene-terminator cassettes each consisting of 3 or 4 of the original 17 fragments. Equal molar of each DNA fragment (0.24 pmol) was combined for each of these 5 sub-assemblies in a total of 5 μL, to which 5 μL of the 2× Gibson assembly mix was added, and incubated at 50 °C for 1 h. After the completion of the Gibson assembly reaction, each sub-assembly was PCR amplified using primers that anneal to the 5′ and 3′ ends of each assembled DNA fragment, using the Phusion DNA polymerase (New England Biolabs), 0.4 μM of each PCR primer, and PCR condition as follows: 98 °C, 10 s; 35 cycles of 98 °C for 1 s, 57 °C for 5 s, 72 °C for 60 s; and 72 °C for 2 min.

To transform the 5 sub-assemblies of DNA fragments into the yeast host strain *BP1458Cas9*, equal molar amounts of each of the 5 fragments totaling 1 μg were combined with 140 ng of plasmid pSE008 and transformed into 50 μL of competent cells by electroporation as described above, and selected on SC-histidine plates for 4 days at 30 °C. This transformation event yielded 8 colonies, which were grown for 4 days at 30 °C. After 4 days, 7 of 8 clones had more intense orange colors relative to wild type yeast host strain.

Integration of combinatorial gene expression cassettes for the β-carotene pathway at the *HO* locus

PXI319 from the first carotenoid cassette assembly and integration step contains both the Cas9 expression plasmid pSE005 (pRS426.*PDC1*-Cas9) and the sgRNA expression vector pSE008 (pRS423.*SNR52*-sgRNA.CAN1.Y). To remove the sgRNA vector while retaining the Cas9 vector, the cells of *PXI213* were restreaked to single colonies on an SC-uracil plate and incubated at 30 °C for 3 days. The presence of the pSE005 plasmid (*URA3* marker) and the loss of the pSE008 plasmid were confirmed by restreaking on an SC-uracil plate and by PCR amplification using primers that hybridize to the *HIS3* locus to confirm the absence of the *HIS3* gene. The presence of the plasmid pSE005 (*URA3* marker) was confirmed by restreaking on an SC-uracil plate. This Cas9 expressing strain was named *PXI213* and used as host strain for the next round of sgRNA/Cas9 mediated integration. The assembly of the combinatorial DNA fragment library at the *HO* locus of *PXI213* was carried out with a total of 12 DNA fragments (Table S2, Supplementary Files). The *PDC1* promoter fragment (PCR #24) was synthesized by PCR amplification using genomic DNA from strain *PXI213* that contains the L1 linker-PDC1 promoter sequence in the CAN1.Y integration cassette, using a 5′ primer that includes 50 bp of 5′ *HO* gene sequence (SE173), and a 3′ primer (P2069) that contains 50 bp overlapping sequences with *crtY* gene. The *RPS5* promoter fragment (PCR #25) was amplified from gBlock DNA (GB30) using SE173 as the 5′ primer that contains 50 bp of 5′ *HO* gene sequence, and P2109 as the 3′ primer that includes 50 bp overlap with the *crtY* gene. *PGI1* promoter was PCR amplified (PCR#20) from GB31(gBlock DNA) using primers P2082 and P2110, which generated the same linker sequences as TDH3 promoter, with L2 linker at the 5′ end, and a 50 bp overlap with *crtI* gene at the 3′ end. Similarly, *ENO2* promoter (PCR# 21) was amplified using GB32 (gBlock DNA) as template, and P2084 and P2111 as 5′ and 3′ primers to generate the L3 linker at the 5′ end, and 50 bp overlap with *crtB* at the 3′ end. Each of the three genes was amplified using PCR primers at 5′ and 3′ ends of the respective ORFs. The *TKL1t* sequence was amplified using PCR primers (P2093, SE174) with gBlock DNA GB020 (Table S4, Supplementary Files) as template. To perform the combinatorial assembly with two possible promoter elements per gene, these 12 DNA fragments were combined in equal molar concentrations to a total of 1 μg, and transformed into *PXI213* together with pSE045 by electroporation.

Several hundred colonies were obtained from the transformation after incubating the SC-histidine plate for 3 days at 30 °C, of which 28 transformants were selected randomly to be evaluated. From these clones, 22 transformants

that had equal or higher levels of orange color formation based on visual inspection were screened (6 other clones with similar orange color as control were not screened) for cassette integration at the HO-G locus by colony PCR using the amplification primer pair SE175/SE090 (Table S3, Supplementary Files) that hybridize to 5′ and 3′ of the HO-G integration site. 16 of the 22 clones showed a PCR amplification product that is consistent with an expected insert size of 8 kb; while the other 6 clones gave 300 bp PCR products indicating the presence of the wild type *HO* gene with no insertion (Fig. 4b). The production of β-carotene was confirmed by HPLC analysis [6].

Results

Selection and testing of sgRNAs using template dsDNAs for ORF deletion

To test the efficiency of sgRNA/Cas9 directed chromosomal integration, we first used double stranded oligonucleotides (dsDNA) as template DNA to perform the homology directed repair of the DNA double strand breaks generated by the Cas9 enzyme at several sgRNA targets in the genome. We chose three sgRNA targets in the yeast genome, including the CAN1.Y site in the *CAN1* gene (encoding arginine permease) [7], and HO.F and HO.G target sites in the *HO* endonuclease gene. In each case, 90 bp dsDNA containing 45 bp from the 5′- and 3′- of DNA sequences flanking the entire ORF (open reading frame) was used as template DNA for the repair of the dsDNA break generated by sgRNA/Cas9. Plasmid pSE005 (pRS426-PDC1-Cas9) was transformed into the haploid yeast strain *BP1548* (*MATa ura3Δ::loxP his3Δ*), to create *BP1458Cas9* strain for the constitutive expression of Cas9 under the *PDC1* promoter (Table 1). To perform the *CAN1* ORF deletion, strain *BP1458Cas9* was co-transformed with plasmid pSE008 (pRS423-*SNR52*-sgRNA.CAN1.Y) and the 90 bp dsDNA (SE80/SE81) flanking the *CAN1* gene. 44 % of the transformed clones on SC-histidine plate were shown to be canavanine resistant by restreaking on Minimal agar plates containing canavanine. Colony PCR and DNA sequencing of canavanine resistant (Can^R) transformants showed that 8 out of 8 randomly picked Can^R transformants have complete deletion of the *CAN1* gene with the expected sequences from the homology-directed recombination of the 90 bp dsDNA. In the absence of the Cas9 expression vector, co-transformation of CAN1-sgRNA vector and the HR donor DNA fragment (90 mer dsOligo) into the *BP1458* host strain resulted in $>10^3$ transformants containing the sgRNA vector while no canavanine resistant clone was detected. This indicates that the presence of Cas9 is required for the dsDNA break mediated disruption of the *CAN1* gene without using canavanine resistance as a positive selection marker, although it is known that homologous recombination is already quite efficient in *S. cerevisiae*.

For the deletion of the *HO* gene via sgRNA-guided Cas9, we selected and compared two sgRNA sequences (HO.F and HO.G) using a DuPont proprietary Browser (described in "Materials and Methods" section). To probe the ORF deletion efficiencies in diploid industrial strains, we performed the *HO* deletion in *PXI318*, a derivative of the diploid strain *YE1358* isolated by William Rutherford (DuPont Pioneer). A 90 bp annealed dsDNA (SE107/SE108) flanking the *HO* gene was used as homologous recombination template DNA. Plasmid pSE045 was co-transformed into the diploid yeast strain *PXI318* together with the SE107/SE108 dsDNA template, and selected for tryptophan prototrophy. Eight colonies from each sgRNA-targeted transformation were analyzed by PCR for the deletion of the *HO* gene, which showed that the sgRNA-HO.F generated biallelic ORF deletion in 25 % of the transformants, with no monoallelic deletion identified. In contrast, 100 % (8 out of 8 clones screened) of the sgRNA-HO.G transformants had complete deletion of both alleles of the *HO* gene (Table 2).

Design of the β-carotene biosynthetic pathway cassette for integration into yeast genome via sgRNA-guided Cas9

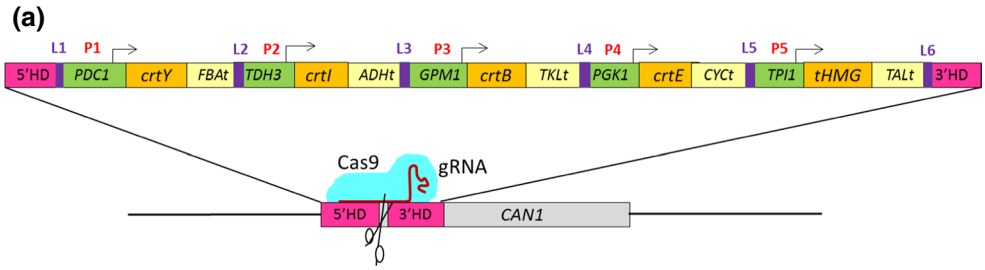
The biosynthetic pathway for β-carotene was used as proof of principle for the one-step assembly of a multi-gene pathway at a single locus in the *S. cerevisiae* genome. In plants and certain fungi, β-carotene is made from the sterol biosynthetic pathway precursor isopentenyl diphosphate (IPP), which is synthesized via the mevalonate pathway from acetyl-CoA. IPP is converted to farnesyl diphosphate (FPP) by FPP synthase, and FPP is subsequently converted to β-carotene via the carotenoid pathway, or to squalene via the sterol pathway. To engineer the yeast *S. cerevisiae* into a β-carotene producing strain, the expression of four enzymes is required to redirect flux from FPP and IPP into the β-carotene biosynthetic pathway. These include geranylgeranyl diphosphate (GGPP) synthase (*crtE*), phytoene desaturase (*crtB*), neurosporene desaturase-lycopene synthase (*crtI*), and lycopene cyclase (*crtY*) [31]. A rate-limiting step in the mevalonate pathway is the conversion of 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) to mevalonate by the action of HMG-CoA reductase. To remove product inhibition of HMG-CoA reductase (encoded by *HMG1*), the soluble, non-membrane bound catalytic domain of Hmg1 (tHmg1) was overexpressed in *S. cerevisiae,* resulting in a higher flux from FPP to squalene, an intermediate of the ergosterol pathway [34]. *tHMG1* expression has been included in β-carotene biosynthetic

Table 1 Strains and plasmids

Name	Genotype	Description	Source
S. cerevisiae strains			
CEN.PK113	*MATa MAL2-8c SUC2*	Prototrophic haploid lab strain	EUROSCARF
BP1548	*MATa ura3Δ::loxP his3Δ*	*CEN.PK113, ura⁻ his⁻* auxotroph	A. Kruckeberg, B. Paul
YE1358	*MATa MATα*	Wild type prototrophic diploid *S. cerevisiae* strain	DuPont Pioneer
PXI104	*MATa MATα ura3Δ::KanMX ura3Δ::KanMX*	*YE1358, ura⁻* auxotroph	This study
PXI318	*MATa MATα ura3Δ ura3Δ trp1Δtrp1Δ*	*PXI104, ura⁻ trp⁻* auxotroph	Z. Xue
PXI318Cas9	*MATa MATα ura3Δ ura3Δ trp1Δtrp1Δ pSE005*	Expression of Cas9 under PDC1 promoter in pSE005 plasmid	
BP1548Cas9	*MATa ura3Δ::loxP his3Δ, pSE005*	Expression of Cas9 under PDC1 promoter in pSE005 plasmid	This study
PXI319	*MATa ura3Δ::loxP his3Δ, can1::PDC1-crtY-FBAt.TDH3-crtI-ADHt.GPM-crtB-TKLt.PGK-crtE-CYCt.TPI-tHMG1-TALt, pSE005, pSE008*	Integration of β-carotene pathway cassette into CAN1 locus, clone R1–22	This study
PXI213	*MATa ura3Δ::loxP his3Δ, can1::PDC1-crtY-FBAt.TDH3-crtI-ADHt.GPM-crtB-TKLt.PGKp-crtE-CYCt.TPI-tHMG1-TALt,* pSE005	β-Carotene producing strain with expression of Cas9 under PDC1 promoter in pSE005 plasmid	This study
Plasmids			
pYPS102	*crtY*, *crtZ*, *crtW* expression	*Y. lipolitica* expression vector	P. Sharpe
pYCRTEBI	*crtE*, *crtB*, *crtI* expression	*Y. lipolitica* expression vector	P. Sharpe
pRS426-PDC1-PmeI-SfiI	pRS426-*PDC1-ILV5t*	*S. cerevisiae* expression vector	W. Suh
pSE005	pRS426-*PDC1-Cas9-ILV5t*	*S. cerevisiae* expression vector for Cas9	This study
pSE008	pRS423-*SNR52*-sgRNA.CAN1.Y	*S. cerevisiae* expression vector for sgRNA.CAN1.Y	Based on Dicarlo et al.
pSE028	pRS424-*SNR52*-sgRNA.HO.F	*S. cerevisiae* expression vector for sgRNA.HO.F	This study
pSE029	pRS424-*SNR52*-sgRNA.HO.G	*S. cerevisiae* expression vector for sgRNA.HO.G	This study
pSE045	pRS423-*SNR52*-sgRNA.HO.G	S. cerevisiae expression vector for sgRNA.HO.G	This study

Table 2 sgRNA/Cas9 targeting sequences used in this study, and the efficiencies of these sgRNAs in gene targeting for ORF deletion

Target gene	Guide RNA	Target sequence	Plasmid	Template dsDNA	Targeting efficiency (%)
CAN1	CAN1.Y	gatacgttctctatggaggatgg	pSE008 (*HIS3*)	SE80/SE81	44
HO	HO.F	tcttacgttatgtgcgcagatgg	pSE028 (*TRP1*)	SE107/SE108	25
HO	HO.G	ccagagtgtataaaatgtggcgg	pSE029 (*TRP1*)	SE107/SE108	100
HO	HO.G	ccagagtgtataaaatgtggcgg	pSE045 (*HIS3*)	SE107/SE108	100



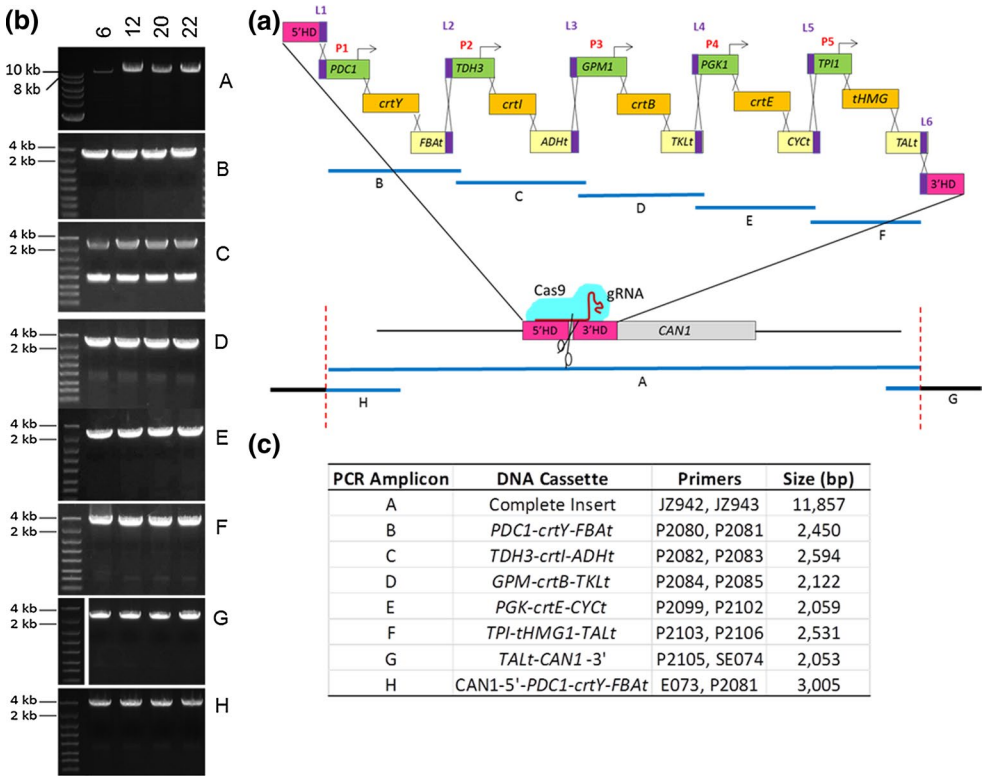
(b)

Linker	Sequence (5'-3')	Restriction Site
L1	ggccctgcaggcctcatagctgtttcctgtgtgaaattgttatccgctca	SfiI
L2	gcggccgccgactcactatagggcgaattggagctccaccgcggtggc	NotI
L3	ggcgcgccctccctttagtgagggttaattgcgcgcttggcgtaatcat	AscI
L4	gcccgggcctggaaagcgggcagtgagcgcaacgcaattaagtgagttag	SrfI
L5	cctgcaggcattaggcaccccaggctttacactttatgcttccggctcgt	SbfI
L6	ccttaacgtgagtttcgttccactgagcgtcagaccgtagggccctgcaggcc	SfiI

Fig. 1 **a** Design of the one-step assembly of the β-carotene biosynthetic pathway in the CAN1.Y locus via sgRNA/Cas9 mediated chromosome integration. The β-carotene biosynthetic pathway is shown as a fully assembled and integrated 11.8 kb DNA fragment composed of five gene cassettes for the expression of *crtY*, *crtI*, *crtB*, *crtE*, *tHMG* genes (*orange*) under yeast promoters (P1–P5, green) *PDC1*, *TDH3*, *GPM1*, *PGK1*, and *TPI1*, with respective terminator sequences (*FBAt*, *ADHt*, *TKLt*, *CYCt*, *TALt*, *yellow*). L1-L6 (*violet*): unique linker sequences between adjacent gene expression cassettes and between 5′ and 3′ terminal gene cassette and the homology domains (5′HD and 3′ HD). **b** L1–L6 linker sequences

pathway engineering in *S. cerevisiae* strains, which led to improved β-carotene production while maintaining an appropriate level of ergosterol to support normal growth of the yeast cells [25, 26].

We designed a combined expression cassette for the *tHMG1* gene together with *crtE*, *crtB*, *crtI* and *crtY* genes from *Pantoea stewartii* DC413 which were codon optimized for expression in the oleaginous yeast *Y. lipolytica* [31]. Expression of these 5 genes was under control of 5 glycolytic yeast promoters: *TPI1, PGK1, GPM1, TDH3* and *PDC1,* respectively (Fig. 1). The two medium level promoters *TPI1* and *PGK1* are utilized for expression of the two upstream genes *tHMG1* and *crtE*, while the three strong promoters *GPM1, TDH3* and *PDC1* were selected for expression of the three downstream genes *crtB*, *crtI* and *crtY*, to maximize the pathway flux towards β-carotene from FPP. The yeast promoters were selected based on transcription profile data in the Yeast Genome Database (http://yeastgenome.org/). Several unique 50 bp linker sequences (L1–L6, Fig. 1) were introduced at the intergenic regions between each of the individual gene cassettes, as well as before the first promoter (*PDC1*) and after the last terminator (*TAL1t*). Each 50 bp linker sequence contains a distinctive 8-bp restriction enzyme recognition site (Table S2, Supplementary Files). The linkers can be used as unique priming sites for PCR amplification using linker-specific primers, as well as sites for restriction mapping. These linker sequences are included as part of the PCR amplification primers that hybridize to the 5′ end of the promoter and 3′ end of the terminator of each gene cassette. To integrate the five gene expression cassettes into the *CAN1* locus in *S. cerevisiae* genome, two homology sequences (5′HD/GB10 and 3′HD/GB11, Fig. 1 and Table S4, Supplementary Files) were synthesized as gBlock DNAs (IDT), each containing 200 bp flanking DNA sequences from the CAN.Y sgRNA site. The 5′ *CAN1*-flanking DNA fragment GB10 also contains the 50 bp linker L1 which overlaps with the *PDC1* promoter PCR fragment, while the 3′-*CAN1*-flanking DNA fragment GB11 contains the 50 bp linker L6, overlapping with the *TAL1* terminator PCR DNA fragments (Fig. 1). Therefore, there are a total of 17 overlapping DNA fragments, with 5 groups of promoter-gene-terminator



PCR Amplicon	DNA Cassette	Primers	Size (bp)
A	Complete Insert	JZ942, JZ943	11,857
B	*PDC1-crtY-FBAt*	P2080, P2081	2,450
C	*TDH3-crtI-ADHt*	P2082, P2083	2,594
D	*GPM-crtB-TKLt*	P2084, P2085	2,122
E	*PGK-crtE-CYCt*	P2099, P2102	2,059
F	*TPI-tHMG1-TALt*	P2103, P2106	2,531
G	*TALt-CAN1* -3'	P2105, SE074	2,053
H	CAN1-5'-*PDC1-crtY-FBAt*	E073, P2081	3,005

Fig. 2 Simultaneous in vivo assembly and chromosomal integration of the 17 overlapping DNA fragments into the CAN1.Y Cas9 target site. **a** Depiction of the 17 overlapping DNA fragments. Each DNA fragment contains 50 bp overlapping sequences with the adjacent DNA fragment(s). **b** Agarose gel analysis for the confirmation of the assembled cassette in the *CAN1* locus by PCR amplification of the complete integration cassette (Amplicon *A*), five individual gene cassettes (Amplicons *B*, *C*, *D*, *E*, *F*), 5′ integration junction (Amplicon *H*), and 3′ integration junction (Amplicon *G*). **c** Table for the PCR amplicons, with descriptions for the DNA cassettes, PCR primer pairs, and PCR product sizes

assemblies, and 2 homologous recombination domains for integration of the combinatorial cassette into the *CAN1* locus.

Direct assembly of 17 overlapping DNA fragments in *S. cerevisiae* genome

To assemble the multiple overlapping DNA fragments efficiently in vivo, it is necessary to utilize Cas9 target sites in *S. cerevisiae* genome that are known to be highly efficient in generating double strand breaks via the sgRNA-targeted cleavage by the Cas9/sgRNA complex. For this study, we used the sgRNA-CAN1.Y site located 207 bp downstream of ATG start codon in the *CAN1* gene, and the sgRNA-HO.G site located 100 bp upstream of the TAA stop codon in the *HO* gene. Based on the design of the combined gene expression cassette (Fig. 1), the 17 DNA fragments with 50 bp overlapping sequences were generated by PCR amplification of the carotenoid pathway bearing plasmids (pYPS102 and pYCRTEBI) and *BY1548* genomic DNA as templates, or gBlock DNA fragments (IDT, Coralville, Iowa). Each of the amplified fragments has 50 bp of sequence that is homologous to the adjacent fragments upstream and downstream. The overlapping DNA fragments would be assembled in vivo via homologous recombination with its neighboring fragment, into the CAN1.Y sgRNA target locus to generate the entire five gene cassette integrated into the genome (Fig. 2a). First, the Cas9 expression plasmid pSE005 (pRS426.*PDC1*-Cas9) was transformed into *BP1548* to generate the host strain *BP1548Cas9*. This allows the constitutive expression of Cas9 before the introduction of the sgRNA vector and donor DNA fragments. Second, the 17 overlapping DNA fragments (Table S1, Supplementary Files) were transformed into the strain *BP1548Cas9* together with plasmid pSE008 (pRS423.*SNR52*-sgRNA.CAN1.Y) and the transformants were plated onto SC-uracil& histidine selective media. A total of 24 colonies were obtained from the transformation, which were repatched on SC-arginine-lysine media containing 60 μg/mL L-canavanine. Four of the 24 transformants showed positive growth on the canavanine plate, indicating that the *CAN1* gene was disrupted. PCR analysis of the 4 Can^R transformants, using primers JZ942 and JZ943 that hybridize to the *CAN1* locus at 5′ and 3′ of the sgRNA-CAN1.Y site, respectively, produced bands of the expected size for the integrated region (11.8 kb, Fig. 2b (A)). Each of the 5 gene expression cassettes was confirmed to be present based on PCR reactions using primers specific to each of the linkers (L1–L6) in the integration cassette (Fig. 2b (B–F)). In addition, the 5′ and 3′ integration junctions of the combined cassette in the CAN1.Y site were confirmed by PCR amplification using primer SE073 located upstream in the *CAN1* locus, and reverse primer P2081 located in the L1 linker (Fig. 2b (G); and forward primer P2105 in the *TAL1* terminator and the *CAN1* locus reverse primer SE074 (Fig. 2b (H)). Therefore, all four

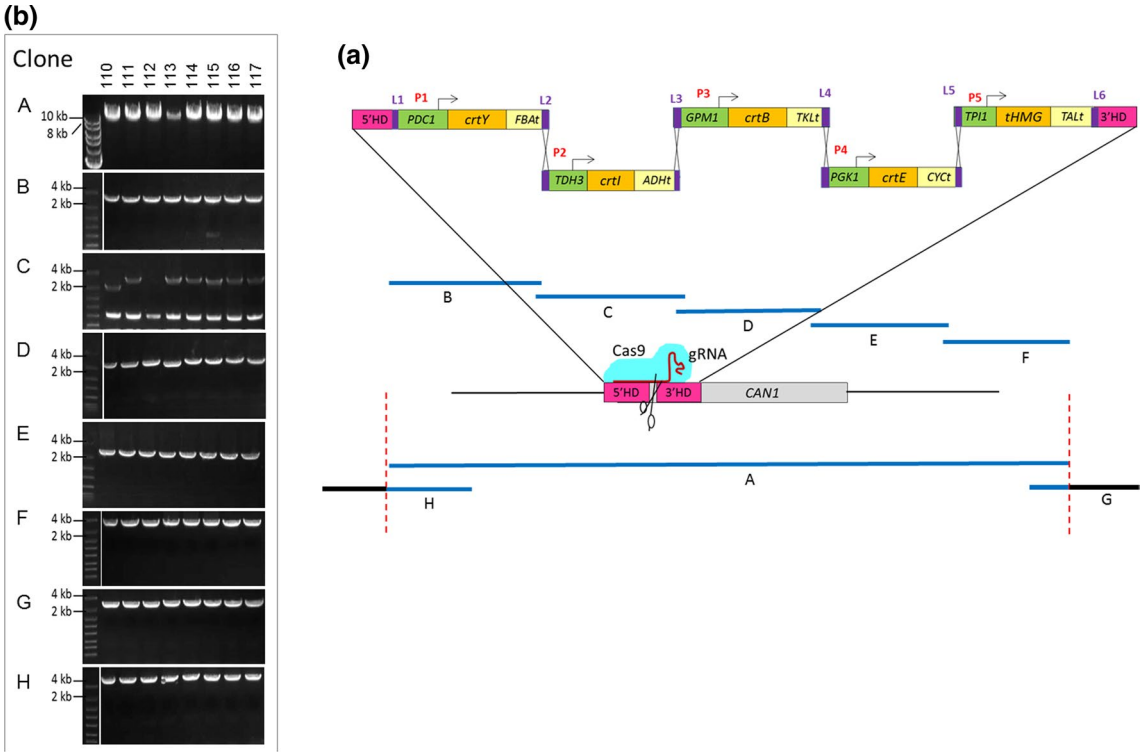


Fig. 3 Construction of five β-carotene biosynthetic gene expression cassettes via Gibson DNA assembly and simultaneous in vivo assembly and chromosomal integration of the 5 piece overlapping DNA fragments into the CAN1.Y Cas9 target site. **a** Depiction of the 5 overlapping DNA fragments pre-assembled from 17 overlapping DNA fragments. Each of the five pre-assembled DNA cassettes contains 50 bp overlapping sequences with the adjacent DNA cassette or the chromosome DNA sequence at the CAN.Y site. **b** Agarose gel analysis for the confirmation of the assembled cassette in the *CAN1* locus. Amplicons *A* to *F* are described in Fig. 2

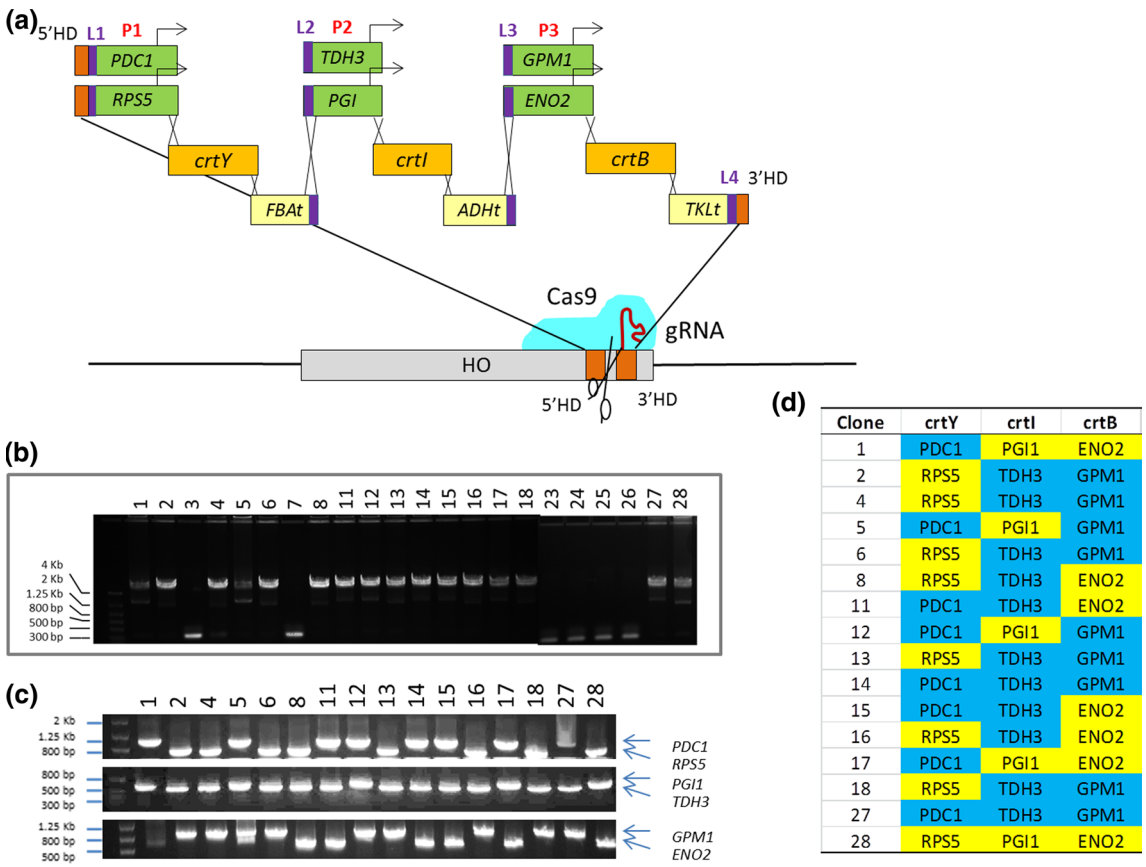
transformants were confirmed to contain the integrated β-carotene pathway genes and their corresponding promoters and terminators as designed in Fig. 1. These four transformants also showed orange color formation after 4 day incubation on YPD plates (Fig. 5a, *BP1548*, R1–6, R1–12, R1–20, and R1–22).

As an alternative pathway assembly strategy, the 17 DNA fragments were pre-assembled into 5 separate gene cassettes via Gibson assembly (Fig. 3). The five cassettes assembled in vitro are 5′HD(*CAN1*)-L1-*PDC1-crtY-FBAt*-L2, L2-*TDH3-crtI-ADHt*-L3, L3-*GPM1-crtB-TLKt*-L4, L4-*PGK1-crtE-CYCt*-L5, and L5-*TPI1-tHMG-TALt*-L6-3′HD(*CAN1*). For the pre-assembled fragment *CAN1*-5′HD-*PDC1-crtY-FBAt* at the 5′ end of the pathway cassette, the Gibson assembly reaction contained PCR fragments 1, 2, 3, and 16 to include the 5′ *CAN1* homology sequences (*CAN1*-5′HD, Table S1, Supplementary Files); while the last gene cassette at the 3′ end (*TPI-tHMG1-TALt-CAN1*-3′HD) contained PCR fragments 13, 14, 15 and 17 to include the 3′ *CAN1* homology sequence (*CAN1*-3′HD). The five pre-assembled DNA fragments were transformed with plasmid pSE008 into strain *BP1548Cas9* for recombinational integration at the sgRNA-CAN1.Y site. Seven of eight transformants on the SC-histidine plate were confirmed for the insertion of the 5-gene cassette, and 6 of these 7 confirmed transformants developed orange colored colonies on YPD plates after 4 days.

Generation of a combinatorial library and screening for improved β-carotene biosynthesis

We continued to use the β-carotene biosynthetic pathway as a model to demonstrate combinatorial gene expression library construction via sgRNA/Cas9. The β-carotene producing strain *PXI213* (clone #R1–22 (*PXI319*) from the 17-piece DNA assembly with pSE008 plasmid removed) contains all the necessary enzymes for the pathway, but is not yet optimized for β-carotene production.

We aimed to improve the pathway flux downstream from GGPP, the first isoprenoid pathway intermediate committed to the β-carotene pathway, to β-carotene. The genes (*crtB*, *crtI* and *crtY*) encoding the last three enzymes leading to β-carotene were integrated as a second copy into the genome. This integration was performed by Cas9-mediated integration at the



(d)

Clone	crtY	crtI	crtB
1	PDC1	PGI1	ENO2
2	RPS5	TDH3	GPM1
4	RPS5	TDH3	GPM1
5	PDC1	PGI1	GPM1
6	RPS5	TDH3	GPM1
8	RPS5	TDH3	ENO2
11	PDC1	TDH3	ENO2
12	PDC1	PGI1	GPM1
13	RPS5	TDH3	GPM1
14	PDC1	TDH3	GPM1
15	PDC1	TDH3	ENO2
16	RPS5	TDH3	ENO2
17	PDC1	PGI1	ENO2
18	RPS5	TDH3	GPM1
27	PDC1	TDH3	GPM1
28	RPS5	PGI1	ENO2

Fig. 4 The assembly and integration of the *crtY*, *crtI, crtB* combinatorial gene expression cassettes at the HO.G locus by RNA-guided Cas9. **a** DNA fragments for the combinatorial assembly and integration. Two sets of promoters including strong promoter (*PDC1*, *TDH3*, and *GPM1*) and medium strength promoters (*RPS5*, *PGI1*, *ENO2*) are selected for the expression of the three genes, *crtY*, *crtI* and *crtB*. Several unique 50 bp linkers are incorporated in the cassette construction, including L1, L2, L3, and L4. L1 and L4 linkers were fused with 50 bp 5′ and 3′ homology domains (5′HD and 3′HD), respectively, as part of the first promoter pair (PDC1 and RPS5) and the *TKL1* terminator PCR product. **b** PCR analysis of the promoter elements in individual strains with the integrated cassette. The entire cassette was PCR amplified, and used as template for subsequence PCR amplification of individual gene cassettes using promoter specific PCR primers. **d** 16 of 28 picked strains were analyzed for the distribution of the promoters in the combinatorial cassette. The strong promoters are shaded in *blue*, and the medium strength promoters are shaded in *yellow*

sgRNA-HO.G locus of *PXI213* strain. Instead of incorporating a single transcription unit for each of the three genes, we chose to construct a combinatorial library directly in the genome.

The assembly of the combinatorial DNA fragment library at the *HO* locus of *PXI213* was carried out as follows (Fig. 4). A total of 12 DNA fragments (Table S2, Supplementary Files) were synthesized as components for the overlap assembly, of which PCR fragments #2–#8 were previously used as components for the 17 piece DNA assembly into *CAN1* locus as described above. DNA fragments comprising two different promoters were synthesized with the same 5′ and 3′ linkers for the expression of *crtB*, while two additional pairs of promoter DNAs were synthesized for *crtI* and *crtY*, respectively, thus providing a combination of high and medium level promoters for each gene (Fig. 4). The following 6 promoters were chosen based on transcription profile data in Yeast Genome database (http://yeastgenome.org): *PDC1* and *RPS5* promoters for *crtY*, *TDH3* and *PGI1* promoters for *crtI*, and *GPM1* and *ENO2* promoters for *crtB* (Fig. 4). The combinatorial assembly of the six promoters with the three respective genes would generate 8 combinations of expression levels for *crtB*, *crtI* and *crtY* genes; resulting in transformants that can be readily screened for their levels of β-carotene production. Taken together, each pair of two neighboring DNA fragments contains the same 50 bp linker sequence (Fig. 4a) for homologous recombination in vivo. The last DNA fragment in the combinatorial assembly sequence is the *TKL1* terminator (*TKL1t*) fragment, which remains unchanged regardless of the choice of the recombination partner DNA fragment. Combining the two promoter DNA elements per gene, three gene fragments and three terminator elements gives a total of 12 DNA fragments each containing 50 bp

overlapping sequences with their respective homologous recombination partners. The integration of the combinatorial assembly at the HO.G site results in deletion of 23 bp and 2 bp at 5′ and 3′ of the PAM site, respectively, with the concomitant elimination of the PAM sequence.

Strain *PXI319* from the previous carotenoid cassette assembly and integration step contains both the Cas9 expression plasmid pSE005 (pRS426.*PDC1*-Cas9) and the sgRNA expression vector pSE008 (pRS423.*SNR52*-sgRNA.CAN1.Y). Upon removal of the pSE008 vector, this Cas9 expressing strain was named *PXI213* and used as host strain for the next round of sgRNA/Cas9 mediated integration of second copies of the three downstream pathway genes: *crtY*, *crtI* and *crtB*. The combined DNA fragments for the combinatorial assembly (Table 3) were co-transformed into strain *PXI213* together with the sgRNA expression vector pSE045 (pRS423.*SNR52*-sgRNA.HO.G), and selected on the SC-histidine selective media for clones that contain the transformed sgRNA vector. A total of 16 clones from 28 randomly selected transformants contained the *crtY*, *crtI* and *crtB* gene cassettes, with each gene fused to one or the other of the two possible promoters (Fig. 4c). Seven of the eight possible combinations of the 3 gene cassettes were present in the group of 16 clones (Fig. 4d). These 16 clones were restreaked on YPD plates and incubated at 30 °C for 4 days, together with the host strain *BP1458* as a non-carotenogenic control strain and the strain *PXI213* from the first round of β-carotene pathway integration as a reference strain (Fig. 5). All 16 clones appeared to have greater orange color formation compared to the two parent strains. The production of β-carotene was confirmed by HPLC analysis, which showed that most of these clones had up to fourfold increases in β-carotene production compared to the first generation strain *PXI213* (see table in Fig. 5).

Discussion

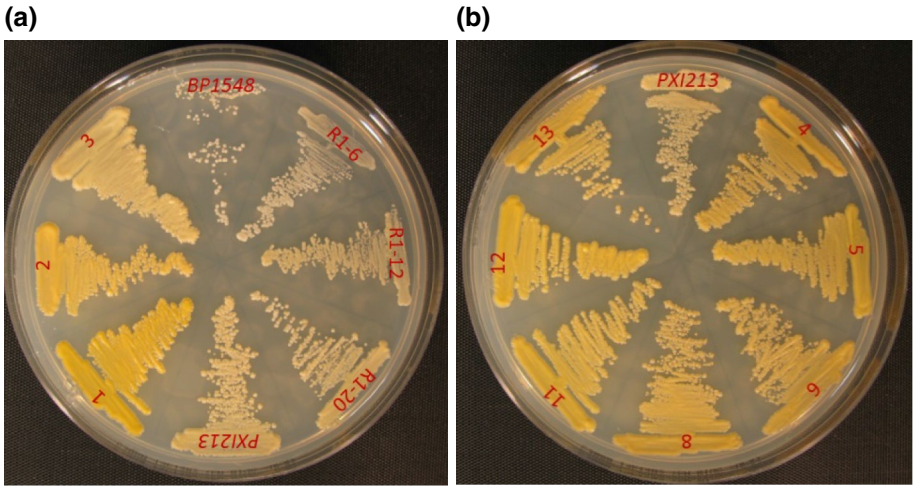
The RNA-guided Cas9 system has enabled precise and rapid genome editing in many organisms. The current study together with several recent publications demonstrated the versatile utility of RNA-guided Cas9 in establishing a platform for accelerated gene assembly and metabolic pathway engineering in *S. cerevisiae*. As double strand DNA breaks generated by RNA-guided Cas9 in the genome are toxic to the cell, it provides a strong positive selection for clones with gene editing or insertion/deletion at the sgRNA target site. Consequently, multiple overlapping DNA fragment assembly directly in the genome becomes feasible, allowing the integration of in vivo assembled DNA fragments at any desired genomic loci. To achieve high productivity of an engineered metabolic pathway, the expression levels of the pathway enzymes need to be optimized for better pathway flux and minimal accumulation of intermediates and byproducts, as well as reduced protein burden. Traditionally, the enzyme expression levels of an engineered pathway are individually optimized; this stepwise approach usually requires multiple rounds of re-optimization due to potentially unbalanced fluxes in a multi-step pathway. We have demonstrated that multiple contiguous gene expression cassettes can be assembled in a combinatorial fashion directly into the yeast genome to generate a library of pathway expression strains, therefore, bypassing the multiple steps of plasmid-based screening of combinatorial pathways, or of serial cassette integrations into the genome. Although we were not able to identify a single best promoter combination for the gene expression optimization of the lower carotenoid pathway, the results showed that having all three strong promoters (PDC1, TDH1, GMP1) did not lead to higher levels of β-carotene production. In addition, the β-carotene levels produced by the engineered strains in this study are lower compared to some other studies. This may be due to the lack of additional mutations that can further improve the flux towards the mevalonate and carotenoid pathways, as well as the use of *crt* genes that were codon optimized for *Y. lipolytica* instead of *S. cerevisiae*. Improved codon optimization may lead to higher flux in the engineered pathway. On the other hand, the use of *Y. lipolytica* codons for the Cas9 gene was likely appropriate, as we compensated for the lack of codon optimization using the strong FBA promoter for the Cas9 expression.

We identified the HO.G Cas9 target site as a highly efficient RNA-guided Cas9 cleavage site based screening of two potential Cas9 target sites identified using an in the HO.G site with the 90 bp double stranded recombination donor DNA spanning the 5′ and 3′ flanking sequences outside of the *HO* gene resulted in 100 % biallelic ORF deletion via Cas9/sgRNA mediated homologous recombination in the diploid yeast strain *PXI318Cas9*. This demonstrates that homologous recombination-directed DNA repair can occur efficiently at sequences up to 1.6 kb away from the Cas9 cleavage site in both alleles of the diploid yeast strain. For strain construction in which certain genes need to be eliminated, such complete ORF deletions can be advantageous. This is a first example for biallelic deletion of complete ORFs carried out in a single transformation for a diploid yeast strain. Due to the efficient Cas9 cleavage at the CAN1.Y [7] and HO.G target sgRNA sequences, these two sites were used for the integration of β-carotene biosynthetic pathway genes.

We demonstrated for the first time that up to 17 overlapping DNA fragments can be directly assembled in the yeast genome via sgRNA/Cas9-mediated chromosomal double strand DNA breaks and homologous recombination. This

Table 3 Comparison of some recent studies on sgRNA/Cas9 mediated gene editing and genome engineering in *S. cerevisiae*

Method	sgRNA targets	No. of multiplex targeting loci	No. of DNA parts per locus	Integration cassette	5′ and 3′ homology arm length (each) (bp)	Integration efficiency (%)	Ploidity	References
Single knockout	*CAN1*	1	1	90 mer dsOligo and *KanMX* cassette	45	99	Haploid	Dicarlo et al. [7]
Single and Multiple knockout	*YJL064 W, YPL062 W, BTS1, ROX1, ERG9*	5	1	90 mer dsOligo	45	50–100	Haploid	Jakočiūnas et al. [15]
Multiple knockout	*URA3, TRP1, LEU2, HIS3*	1	1	90 mer dsOligo	45	15–60	Polyploid	Zhang et al. [36]
Multiple knockout	*ITR1, PDR12, MCH1, MCH2, MCH5, AQY1*	6	1	6 genes, each in pre-assembled cassette	≥250	65	Polyploid	Mans et al. [23]
ORF Deletion	*CAN1, HO*	1	1	90 mer dsOligo	45	25–100	Haploid and diploid	This study
CRISPRm	Multiple targets	up to 3	1	140 mer dsOligo	50	19–80	Haploid and diploid	Ryan et al. [29]
CRISPRm	*URA3*	1	3	P_{TEF1}/Nat^R/*TEF1t*	50	85	Diploid	Ryan et al. [29]
HI-CRISPR	*CAN1*, *ADE2*, *LYP1, ATF2*	3	1	100 mer dsOligo	50	27–100	Haploid	Bao et al. [2]
CasEMBLR	*ADE2, HIS3, URA3, ARO10, PDC5*	3	5	5 DNA fragments per gene cassette	50	31–97	Haploid	Jakočiūnas et al. [16]
CrEdit	Intergenic loci *XII-5, XI-2* and *X-3*	3	1	3 genes, each in pre-assembled cassettes	500	84	Haploid	Ronda et al. [28]
Multiplexing/Gene Assembly	*GAL80, HO, ARO1*	3	2	11 genes pre-assembled in 3 cassettes	500	4	Haploid	Horwitz et al. [14]
Multiplexing/Gene Assembly	*ACS1, ACS2*	2	6 (*ACS1*), 1 (*ACS2*)	6 genes, each in pre-assembled cassette	50	100	Haploid	Mans et al. [23]
Gene Assembly	*PHO13, ALD6*	1	2	3 gene cassettes in 2 fragments each	60	25–50	Haploid	Tsai et al. [33]
Multipart Gene Assembly	*CAN1*	1	5	5 genes, each in pre-assembled cassettes	50	75	Haploid	This study
Multipart Gene Assembly	*HO*	1	9	3 genes (individual promoter/gene/terminator fragments)	50	50	Haploid	This study
Multipart Gene Assembly	*CAN1*	1	17	5 genes (individual promoter/gene/terminator fragments)	50	17	Haploid	This study



(d)

Strain	β-Carotene (mg/L)
BP1548	0
PXI213	0.04
Clone 1	0.18
Clone 2	0.16
Clone 4	0.16
Clone 5	0.21
Clone 6	0.15
Clone 8	0.14
Clone 11	0.2
Clone 12	0.21
Clone 13	0.14
Clone 14	0.14
Clone 15	0.23
Clone 16	0.13
Clone 17	0.13
Clone 18	0.15
Clone 27	0.11
Clone 28	0.2

Fig. 5 Yeast cells of host strain *BP1548*, first and second round chromosomal integration of the β-carotene biosynthesis cassettes streaked on YPD plate. **a** WT (*BP1548*), and clones R1–6, R1–12, R1–20 and R1–22 from the 17 piece DNA assembly and 1st round integration; and clones 1, 2, 3 from the combinatorial assembly and 2nd round chromosomal integration of *crtY*, *crtI*, *crtB* expression cassettes. Clone R1–22 (with pSE008 removed) is renamed *PXI213*. **b** *PXI213*, and clones 4, 5, 6, 8, 11, 12, 13 from second round integration. **c** *PXI213*, and clones 14, 15, 16, 17, 18, 27 and 28 from second round integration

is by far the largest number of DNA fragments assembled directly in yeast genome instead of a circular plasmid such as 2μ based vectors. Such an approach allows the construction of complete multi-enzyme metabolic pathways directly from individual DNA fragments encoding promoters, ORFs and terminators in a single step without the co-integration of a selectable marker. This will greatly facilitate the assembly of complex metabolic pathways to be stably expressed from the yeast genome, thus enabling the construction of laboratory and industrial yeast strains without extensive rounds of integration and marker recycling steps. Both the efficiency of sgRNA/Cas9 targeting (e.g. 44 % for *CAN1* gene ORF deletion) and multi-fragment assembly (17 % for the assembly of 17 fragments in a single locus) reported in this study can be further improved by more efficiently targeted sgRNA/Cas9 complex, and increasing the in vivo concentrations of HR donor DNA fragments. Combined with multiplex targeting demonstrated in several other studies, the simultaneous assembly of even more DNA fragments can be envisioned. As indicated by Gibson et al. [11], it may be possible to assemble a large number of overlapping DNA sequences in yeast (>25), albeit the probability of the entry of all the different DNA fragments into the nucleus of a single cell is reduced significantly with the increased complexity.

Alternatively, it is possible to assemble a metabolic pathway using overlapping DNA fragments containing various promoters, genes and terminators, and several pairs of 5′ and 3′ homology sequences to direct the integration of the same pathway cassette or different subsets of the metabolic pathway to multiple loci in the yeast genome. This will result in multi- copy integration of the pathway genes, which may be necessary for high-level expression of the engineered pathways. To demonstrate the potential

application of multiple DNA fragment assembly, we generated a small combinatorial DNA library for the three downstream genes of the β-carotene biosynthetic pathway (*crtB*, *crtI* and *crtY*) directly integrated into the yeast genome to create a diversity library of strains. This enabled the screening of combinatorial libraries in stable chromosomally integrated strains for the simultaneous testing of many hypotheses including the screening of mutant libraries of enzymes and transcription factors aimed at rapid improvements of product titers. The diversity library generation can be applied to the combinatorial expression of multiple CRISPR-associated catalytically inactive dCas9 proteins [12] to select for desired levels of multiplex gene regulation. Using this approach, it is also possible to carry out DNA shuffling while directly integrating the various shuffled and assembled genes, gene fragments, promoters, etc., in the genome. From our experience of metabolic pathway engineering, the two promoter-per-genes strategy utilized in the current study is likely a practical approach for range finding in metabolic pathway engineering.

In summary, we established an expanded toolbox for the single step sgRNA/Cas9 mediated assembly and marker-free chromosomal integration of multiple combinatorial gene expression cassettes and libraries. Combined with Cas9 directed genome editing, multiplexing ORF deletion and insertion, the approaches developed here will enable the rapid strain engineering in the future directed at scouting of a variety of biosynthetic pathways leading to new targets and commercial products derived from metabolic pathway engineering.

Acknowledgments We would like to thank the following DuPont colleagues for their assistance with the work described here. We thank Xiaochun Fan, Brian Paul, William Rutherford, Pamela Sharpe, Melissa Bosak, Wonchul Suh, Bhaska Bhadra and Zhixiong Xue for sharing yeast strains, plasmids and protocols. We thank John Kozlowski and Beth Ann Fenner for performing HPLC analysis of β-carotene. Arle Kruckeberg, Raymond Seung-Pyo Hong, Mick Ward, and Zhixiong Xue are appreciated for editing the manuscript. We also thank Ethel Jackson, Ryan Frisch, and Xiaochun Fan for valuable discussions. In addition, we would like to thank Wayne Alphin and Eva Nagel for their help with imaging.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests in this work.

References

1. Akada RKT, Kaneko S, Toyonaga D, Ito S, Kakihara Y, Hoshida H, Morimura S, Kondo A, Kida K (2006) PCR-mediated seamless gene deletion and marker recycling in *Saccharomyces cerevisiae*. Yeast 23:399–405
2. Bao Z, Xiao H, Liang J, Zhang L, Xiong X, Sun N, Si T, Zhao H (2015) Homology-Integrated CRISPR−Cas (HI-CRISPR) system for one-step multigene disruption in *Saccharomyces cerevisiae*. ACS Synth Biol 4:585–594
3. Barrangou RC-MA, Stahl B, Chavichvily I, Damange F, Romero DA, Boyaval P, Fremaux C, Horvath P (2013) Genomic impact of CRISPR immunization against bacteriophages. Biochem Soc Trans 41:1383–1391
4. Borodina I, Nielsen J (2014) Advances in metabolic engineering of yeast *Saccharomyces cerevisiae* for production of chemicals. Biotechnol J 9:609–620
5. Cong LZF (2015) Genome engineering using CRISPR-Cas9 system. Methods Mol Biol 1239:197–217
6. Cunningham F Jr, Gantt E (2005) A study in scarlet: enzymes of ketocarotenoid biosynthesis in the flowers of Adonis aestivalis. Plant J 41:478–492
7. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucl Acids Res 41:4336–4343
8. Doudna JA (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science 346(6213):1258096
9. Feng Z, Zhang B, Ding W, Liu X, Yang DL, Wei P, Cao F, Zhu S, Zhang F, Mao Y, Zhu JK (2013) Efficient genome editing in plants using a CRISPR/Cas system. Cell Res 23:1229–1232
10. Friedland AE, Tzur YB, Esvelt KM, Colaiacovo MP, Church GM, Calarco JA (2013) Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. Nat Methods 10:741–743
11. Gibson DG, Benders GA, Axelrod KC, Zaveri J, Algire MA, Moodie M, Montague MG, Venter JC, Smith HO, Hutchison CA 3rd (2008) One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. Proc Natl Acad Sci USA 105:20404–20409
12. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154:442–451
13. Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. Science 327:167–170
14. Horwitz AA, Walter JM, Schubert MG, Kung SH, Hawkins K, Platt DM, Hernday AD, Mahatdejkul-Meadows T, Szeto W, Chandran SS, Newman JD (2015) Efficient multiplexed integration of synergistic alleles and metabolic pathways in yeasts via CRISPR-Cas. Cell Systems. doi:10.1016/j.cels.2015.02.001.24
15. Jakočiūnas T, Bonde I, Herrgard M, Harrison SJ, Kristensen M, Pedersen LE, Jensen MK, Keasling JD (2015) Multiplex metabolic pathway engineering using CRISPR/Cas9 in *Saccharomyces cerevisiae*. Metab Eng 28:213–222
16. Jakočiūnas T, Zhang J, Arsovska D, Rodriguez A, Jendresen CB, Skjødt ML, Nielsen AT, Borodina I, Jensen MK, Keasling JD (2015) CasEMBLR: Cas9-facilitated multiloci genomic integration of in vivo Assembled DNA Parts in *Saccharomyces cerevisiae*. ACS Synth Biol 4(11):1226–1234
17. Jensen MK, Keasling JD (2014) Recent applications of synthetic biology tools for yeast metabolic engineering. FEMS Yeast Res. doi:10.1111/1567-1364.12185
18. Jones JA, Vernacchio VR, Lachance DM, Lebovich M, Fu L, Shirke AN, Schultz VL, Cress B, Linhardt RJ, Koffas MA (2015) ePathOptimize: a combinatorial approach for transcriptional balancing of metabolic pathways. Sci Rep. 5:11301. doi:10.1038/srep11301
19. Kuijpers NG, Chroumpi S, Vos T, Solis-Escalante D, Bosman L, Pronk JT, Daran JM, Daran-Lapujade P (2013) One-step assembly and targeted integration of multigene constructs assisted by the I-SceI meganuclease in *Saccharomyces cerevisiae*. FEMS Yeast Res 13:769–781
20. Leber C, Da Silva NA (2014) Engineering of *Saccharomyces cerevisiae* for the synthesis of short chain fatty acids. Biotechnol Bioeng 111:347–358

21. Lee ME, DeLoache WC, Cervantes B, Dueber JE (2015) A highly characterized yeast toolkit for modular, multipart assembly. ACS Synth Biol doi:10.1021/sb500366v
22. Liu R, Chen L, Jiang Y, Zhou Z, Zou G (2015) Efficient genome editing in filamentous fungus *Trichoderma reesei* using the CRISPR/Cas9 system. Cell Discov 1:15007
23. Mans R, van Rossum HM, Wijsman M, Backx A, Kuijpers NG, van den Broek M, Daran-Lapujade P, Pronk JT, van Maris AJ, Daran JM (2015) CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. FEMS Yeast Res 15(2):fov004
24. Peccoud J, Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S (2011) A modular cloning system for standardized assembly of multigene constructs. PLoS One 6:e16765
25. Polakowski T, Stahl U, Lang C (1998) Overexpression of a cytosolic hydroxymethylglutaryl-CoA reductase leads to squalene accumulation in yeast. Appl Microbiol Biotechnol 49:66–71
26. Qian Li ZS, Li Jing, Zhang Yansheng (2013) Enhancing beta-carotene production in *Saccharomyces cerevisiae* by metabolic engineering. FEMS Microbiol Lett 345:94–101
27. Horton RM (1995) PCR-mediated recombination and mutagenesis. SOEing together tailor-made genes. Mol Biotechnol 3(2):93–99
28. Ronda C, Maury J, Jakociunas T, Jacobsen SA, Germann SM, Harrison SJ, Borodina I, Keasling JD, Jensen MK, Nielsen AT (2015) CrEdit: CRISPR mediated multi-loci gene integration in *Saccharomyces cerevisiae*. Microb Cell Fact 14:97
29. Ryan OW, Skerker JM, Maurer MJ, Li X, Tsai JC, Poddar S, Lee ME, DeLoache W, Dueber JE, Arkin AP, Cate JH (2014) Selection of chromosomal DNA libraries using a multiplex CRISPR system. Elife 3:e03703
30. Shao Z, Zhao HUA, Zhao H (2009) DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucl Acids Res 37:e16
31. Sharpe P, Ye R, Zhu, Q (2007) US Patent 8846374, Carotenoid production in a recombinant oleagenous yeast
32. Sun J, Alper HS (2015) Metabolic engineering of strains: from industrial-scale to lab-scale chemical production. J Ind Microbiol Biotechnol 42:423–436
33. Tsai CS, Kong II, Lesmana A, Million G, Zhang GC, Kim SR, Jin YS (2015) Rapid and marker-free refactoring of xylose-fermenting yeast strains with Cas9/CRISPR. Biotechnol Bioeng. doi:10.1002/bit.25632
34. Xie W, Liu M, Lv X, Lu W, Gu J, Yu H (2014) Construction of a controllable beta-carotene biosynthetic pathway by decentralized assembly strategy in *Saccharomyces cerevisiae*. Biotechnol Bioeng 111:125–133
35. Yan Y, Kohli A, Koffas MA (2005) Biosynthesis of natural flavanones in *Saccharomyces cerevisiae*. Appl Environ Microbiol 71(9):5610–5613
36. Zhang GC, Kong II, Kim H, Liu JJ, Cate JHD, Jin YS (2014) Construction of a quadruple auxotrophic mutant of an industrial polyploid *Saccharomyces cerevisiae* Strain by Using RNA-Guided Cas9 Nuclease. App Env Microbiol 80(24):7694–7701