Synthetic and Systems Biotechnology 6 (2021) 63–68

Contents lists available at ScienceDirect

Synthetic and Systems Biotechnology

journal homepage: www.sciencedirect.com/journal/synthetic-and-systems-biotechnology

Screening neutral sites for metabolic engineering of methylotrophic yeast *Ogataea polymorpha*

Wei Yu [a,d], Jiaoqi Gao [a,b,c], Xiaoxin Zhai [a,b,c], Yongjin J. Zhou [a,b,c,*]

[a] Division of Biotechnology, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, 457 Zhongshan Road, Dalian, 116023, PR China
[b] CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, 116023, PR China
[c] Dalian Key Laboratory of Energy Biotechnology, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, 116023, PR China
[d] University of Chinese Academy of Sciences, Beijing, 100049, PR China

ARTICLE INFO

Keywords:
Neutral sites
Ogataea polymorpha
Metabolic engineering
CRISPR-Cas9
Fatty alcohol biosynthesis

ABSTRACT

Methylotrophic yeast *Ogataea polymorpha* is capable to utilize multiple carbon feedstocks especially methanol as sole carbon source and energy, making it an ideal host for bio-manufacturing. However, the lack of gene integration sites limits its systems metabolic engineering, in particular construction of genome-integrated pathway. We here screened the genomic neutral sites for gene integration without affecting cellular fitness, by genomic integration of an enhanced green fluorescent protein (*eGFP*) gene via CRISPR-Cas9 technique. After profiling the growth and fluorescent intensity in various media, 17 genome positions were finally identified as potential neutral sites. Finally, integration of fatty alcohol synthetic pathway genes into neutral sites NS2 and NS3, enabled the production of 4.5 mg/L fatty alcohols, indicating that these neutral sites can be used for streamline metabolic engineering in *O. polymorpha*. We can anticipate that the neutral sites screening method described here can be easily adopted to other eukaryotes.

Introduction

The methylotrophic yeast *Ogataea polymorpha* (formerly named as *Hansenula polymorpha*) has a broad substrate spectrum including glucose, xylose and methanol, which makes it a potential host for bio-manufacturing with relieved competition with food industry [1–3]. In addition, *O. polymorpha* is highly tolerant against harsh industrial conditions, such as high temperature up to 50 °C, low pH and high osmotic pressure [4–6]. Altogether, these characters show the high potential of *O. polymorpha* as a promising host toward industrial processes.

Construction of microbial cell factories involves rewiring native metabolic pathways and reconstruction of heterologous pathways, which is always realized by overexpression of multiple genes. Although the episomal plasmid has been widely used for gene expression, its segregational and structural instability may cause plasmid loss [7,8]. Furthermore, the limited DNA accommodation capacity of a plasmid (usually less than 20 k base pairs) makes it challenging for expression of multiple genes in extensive metabolic engineering. Alternatively, the genomic integration is considered as a more feasible approach for construction of stable microbial cell factories [9,10]. In this case, the integration loci with no interruption of cell fitness is very essential. Though, the loci of normally used auxotrophic markers *LEU2, URA3 and HIS4* can be used for genomic integration, the limited auxotroph loci makes it difficult in integration of multiple genes and it may bring cell interfering when integration of genes into auxotroph loci. Thus, non-coding DNA sequences is more feasible as neutral site for gene integration, since it may avoid disrupting the cellular fitness [11]. Unlike the model yeast *Saccharomyces cerevisiae* [10,12,13], there is lacking genetic tools and elements including neutral sites for the non-conventional yeast *O. polymorpha.*

We here characterized a variety of potential neutral sites in the genome of *O. polymorpha* based on the genome sequence of NCYC 495 leu1.1 [14] by using a green fluorescent protein gene (*eGFP*) as a reporter. We also evaluated the growth profile of genomic integrated strains in the synthetic media with glucose, xylose or methanol as sole carbon source, respectively. Finally, a fatty alcohol biosynthetic pathway was constructed by integrating its three genes into two neutral sites, which confirmed that these neutral sites can be employed for

Peer review under responsibility of KeAi Communications Co., Ltd.
* Corresponding author. Division of Biotechnology, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, 457 Zhongshan Road, Dalian, 116023, PR China.
E-mail address: zhouyongjin@dicp.ac.cn (Y.J. Zhou).

https://doi.org/10.1016/j.synbio.2021.03.001
Received 25 January 2021; Received in revised form 19 February 2021; Accepted 10 March 2021

metabolic engineering of *O. polymorpha*.

Materials and methods

Plasmids, strains and cell cultivation

Stains and plasmids used in this study are listed and described in Supplemental Table S1.

O. polymorpha strains for preparation of competent cells were cultivated in YPD medium consisting 10 g/L yeast extract, 20 g/L tryptone and 20 g/L glucose. Recombinant strains were selected in the synthetic droplet (SD) medium containing 6.7 g/L yeast nitrogen base (YNB) and 20 g/L glucose. The *URA3* maker was removed and selected against on SD + URA+5-FOA plates, which contained 6.7 g/L YNB, 20 g/L glucose, 20 mg/L uracil, and 1 g/L 5-fluoroorotic acid. All yeast solid media were added with 20 g/L agar. All yeast strains were cultivated in 37 °C. *Escherichia coli* DH5α used for plasmid transformation were cultivated at 37 °C in LB medium (10 g/L peptone, 5 g/L yeast extract, 10 g/L NaCl). 100 μg/mL ampicillin was added for plasmid maintenance.

Yeast shake flask fermentations were carried out at 37 °C in Delft minimal medium (pH5.6) containing 2.5 g/L $(NH_4)_2SO_4$, 14.4 g/L KH_2PO_4, 0.5 g/L $MgSO_4{\cdot}7H_2O$, 20 g/L glucose, 2 mL/L trace metal and 1 mL/L vitamin solutions [15] supplemented with 60 mg/L uracil if needed. For growth profiling, Delft minimum medium containing 20 g/L glucose (Delft-Glucose), 20 g/L xylose (Delft-Xylose) and 10 g/L methanol (Delft-Methanol) was used. Delft-Glucose medium was used for fatty alcohol production.

Genetic manipulation

Super Fidelity DNA polymerase, Taq DNA polymerase and ClonExpress MultiS One Step Cloning Kit were purchased from Vazyme Biotech, Nanjing, China. DNA gel purification and plasmid extraction kits were purchased from Omega Bio-Tek, US. All oligonucleotides (Supplementary Table 1) were synthesized by Sangon Biotech, Shanghai, China. All chemicals were purchased from Sangon Biotech unless otherwise stated.

NS2, NS3, NS5, NS6, NS11, NS12, NS14, NS16, NS18, NS19, 2NS2, 3N2, 4NS5, 4NS11, 5NS6, 5NS12, 6NS9 7NS9 gRNA expression plasmids were constructed by separating our previous constructed plasmid pHp39 [16] into two parts: backbone part and gRNA parts. Backbone and gRNA parts were amplified with primer pair BB-F/BB-R, NSxgRNA-F/HpgRNA-R, respectively, by using pHp1 as a template. Then the two parts were gel purified and assembled using ClonExpress MultiS One Step Cloning Kit. The assembled plasmid was transformed into *E. coli* DH5α competent cell for propagation.

Gene integration in *O. polymorpha* was performed by using a CRISPR-Cas9 genome engineering technology established in our laboratory [16]. Guide RNA plasmid targeting 20 bp of candidate neutral sites in the genome, which was chosen based on the CHOPCHOP online software [17], was constructed by using pHp39 as backbone. Donor DNA, consisting upstream homologous arm (HA), promoter, *eGFP* gene, terminator and downstream HA, was generated by using fusion PCR. Taking NS2-eGFP as an example, 1000 bp upstream HA, 700 bp *GAP* promoter and 1000 bp downstream HA were PCR amplified from *O. polymorpha* NCYC 495 genome DNA by using primer pair NS2up-F/NS2up-R, P_{GAP}-F/P_{GAP}-R, and NS2dw-F/NS2dw-R, respectively. T_{TDH2} and *eGFP* gene was amplified from *S. cerevisiae* S288C genome and plasmid peGFP using primer pair T_{TDH2}-F/T_{TDH2}-R, and eGFP-F/eGFP-R, respectively. The donor DNA was gel purified and transformed with gRNA expression plasmid into *O. polymorpha* competent cells and selected in SD plates. 10 clones were selected and verified by colony PCR. Clones with correct integration were cultivated overnight in YPD liquid medium and then plated on SD + URA+5-FOA plates for removing gRNA expression plasmid.

Gene deletion was conducted by CRISPR-Cas9 genome editing tool as described in our previous study [16]. In detail, the guide RNA expression plasmid, containing 20bp guide RNA targeting this gene's ORF, was constructed by integration of guide RNA sequence to the backbone of pHp39. Donor DNA was constructed by fusion PCR of 1000 bp upstream and 1000 bp downstream homologous arm. Then the guide RNA plasmid and donor DNA were transformed into *O. ploymorpha* via electroporation. The transformants were cultivated on SD media and screened by colony-PCR [18]. To remove gRNA plasmids, transformants were plated on SD + URA+5-FOA plates.

Transformation of O. polymorpha

Transformation of *O. polymorpha* was performed by using a high efficient electro-transformation method [19], with small modifications. 10 mL of *O. polymorpha* culture with OD_{600} of 0.8–1.0 were harvested by centrifugation and resuspended in 4 mL of 50 mM potassium phosphate buffer (pH7.5) containing 25 mM dithiotreitol. After incubation at 37 °C for 15 min, cells were sequentially washed with 10 mL and 5 mL of electroporation buffer STM (270 mM sucrose, 10 mM Tris-HCl pH7.5 and 1 mM $MgCl_2$), and then suspended in 200 μL STM. 60 μL cell suspension was mixed with gRNA expression plasmid and donor DNA, and tapped to the bottom of prechilled 2 mm electroporation cuvette following the electric field pulse (7.5 kV/cm, 50 μF, 129 Ω, resulting in pulse length of 4 ms). The cell/DNA mixture was then mixed with 1 mL YPD medium and then incubated at 37 °C for 1 h without agitation, followed by harvesting and washing with 1 mL sterilized H_2O. Finally, the cells was resuspended in 100 μL H_2O and spread on SD medium. The plates were incubated at 37 °C for 2–3 days.

Fluorescence measurements

For determination of eGFP fluorescent intensity of engineered NS strains, cells were harvested and centrifugated at 13000×*g* for 3 min. The cell pellet was washed with 1 mL H_2O, and diluted to OD_{600} = 0.2–0.8. Then the cell suspension was subject to the Tecan SPARK microplate reader (Tecan, Switzerland) in a 96-well black-walled clear-bottom plate for fluorescence and absorbance determination. The excitation and emission wavelengths were set to 485 nm and 528 nm, respectively. Culture density was also measured as the absorbance at 600 nm (OD_{600}). All fluorescence measurements were normalized by OD_{600}.

Fatty alcohol synthetic pathway construction and evaluation

The fatty alcohol synthetic pathway was constructed by using a modular pathway engineering strategy [20]. The *CAR* and *NpgA* genes, which were codon optimized based on codon preference of *O. polymorpha*, were obtained by PCR amplification from pHp63 [16] by using primer pairs CAR-F/CAR-R and npgA-F/npgA-R, respectively. The *ADH5* module was amplified from pHp65 using primer pair ADH5-F/ADH5-R. The fatty alcohols were quantified as previously described [21]. Briefly, cell pellets were collected from 5 mL cell culture and then freeze-dried for 48 h. Fatty alcohol was extracted by chloroform: methanol solution (2:1) containing pentadecanol as internal standards. Then fatty alcohols were extracted by using microwave digestion/extraction system. The samples were vortexed vigorously and placed in the microwave reaction vessel containing 10 mL Mili-Q water then sealed with a TFM screw cap. 1 mL NaCl solution (0.73%, w/v) was added and then the samples were vortexed vigorously. The samples were centrifuged at 1000×g for 10 min for phase separation and then the organic phase was transferred into an extraction tube. The extracted fractions were dried by using a centrifugal concentrator. 100 μL BSTFA (bis(trimethylsilyl)trifluoroacetamide) was added into the vial for silylation (80 °C, 30 min) and the silylated fatty alcohol was dissolved in 400 μL hexane. GC-MS analysis was as follows: initial temperature of 45 °C held for 2.5 min; then ramp to 220 °C at a rate of 20 °C/min and held for 2 min; ramp to 300 °C at a rate of 20 °C/min and held for 5min.

The temperature of inlet, mass transfer line and ion source were kept at 250, 300 and 230 °C, respectively. The flow rate of the carrier gas (helium) was 1.0 mL/min. MS data were acquired at full-scan mode (50–650 *m/z*) and quantified with Xcalibur software.

Results

Screening candidate neutral sites

Genome sequence of *O. polymorpha* NCYC 495 leu1.1 (Accession No. NW_017264698.1) from National Center for Biotechnology Information (NCBI) was used as reference for screening the candidate neutral sites. Here, the interval regions between two genes were selected for further investigation, since gene integration into any genomic sequence encoding genes or hypothetic genes may interfere cellular fitness. In eukaryotes, the promoter and terminator are essential for functional gene expression and regulation. Thus, the length of interval region as potential neutral sites should be longer than the total length of promoter and terminator. According to the flanking genes' transcription direction (Fig. 1A), the interval regions were classified into 3 categories: HH (the head-to-head localization of two adjacent open reading frames (ORFs), which means that this interval region contains two promoters); HT or TH (head-to-tail or tail-to-head localization of two adjacent ORFs, which means that this interval region contains one promoter and one terminator); TT (tail-to tail localization of two ORFs, which means this region contain two terminators). Normally, the length of a promoter and a terminator is approximately 600 bp and 400 bp, respectively. Therefore, interval regions longer than 1500 bp (HT, TH and HH) or 1200bp (TT) were selected as candidate neutral sites. With these criteria, a total of 36 candidate neutral sites (18 sites in chromosome VII, and 3 longest sites in each chromosome I-VI, respectively) were found for further investigation (Fig. 1B), with the 18 correct sites marked with red (results below). The detailed information of all potential neutral sites were listed in Supplemental Table S2.

For convenient genomic integration, we used an optimized CRISPR-Cas9 platform and the host *O. polymorpha* strain y34 with enhanced homologous recombination [16]. This optimized platform enabled 10%–40% integrating efficiencies in targeting each neutral site with *eGFP* gene (Table 1), which can ensure the accurate integration of heterologous gene into these candidate sites. Those sites did not get correct transformant or positive fluorescence were ruled out. A total of 18 neutral sites located in chromosome I-VII got varied green fluorescence intensity with integrated *eGFP* gene (Fig. 2B). In detail, strain NS3, NS18 and NS19 had relatively high fluorescent intensity, while NS3 was the highest one, which suggested that specific chromosome structure at different sites might affect gene expression. This gene expression variation may be helpful for balancing the metabolic pathways by fine tuning the expression of corresponding genes.

Profiling the growth of NS strains

As 18 candidate neutral sites were obtained, the next step was to verify whether they were neutral. In other word, it should be verified that whether gene integration into these candidate sites influence cell growth. We here profiled the growth of neutral site integrated strains and the parent strain y34 in synthetic media with glucose (20 g/L), xylose (20 g/L) or methanol (10 g/L) as a carbon source. All NS strains exhibited similar growth profile as y34 in all media (Fig. 2), which indicated that *eGFP* gene integration into these 18 candidate sites did not affect the yeast growth with carbon sources of glucose, xylose or methanol. It should be mentioned that there is a little varied cell growth among all the NS strains in methanol containing media (Fig. 2E), which might be attributed to varied methanol volatilization.

Dynamics of fluorescence, representing the concentration of eGFP, showed a similar profiles among the NS strains. In Delft-Glucose, the fluorescent intensity reached maximum in the exponential phase (36 h growth), then dropped to half at the stationary phase (60 h growth), and finally vanished at the late stationary phase (96 h growth), indicating the gradual degradation of eGFP by intracellular protease. The similar decline of fluorescent intensity was observed in xylose and methanol medium, though not so sharp as that in glucose medium. The fluorescent intensities were highest in glucose compared to that in methanol and xylose, which can be explained that *GAP* promoter (P_{GAP}, the promoter of glyceraldehyde-3-phosphate dehydrogenase gene) had the highest transcription activity in glucose. It can be expected that carbon source responsive promoter should be helpful for driving the gene expression in xylose or methanol media. Interesting, the 2NS2 integrated strain showed almost no fluorescence in methanol medium with some

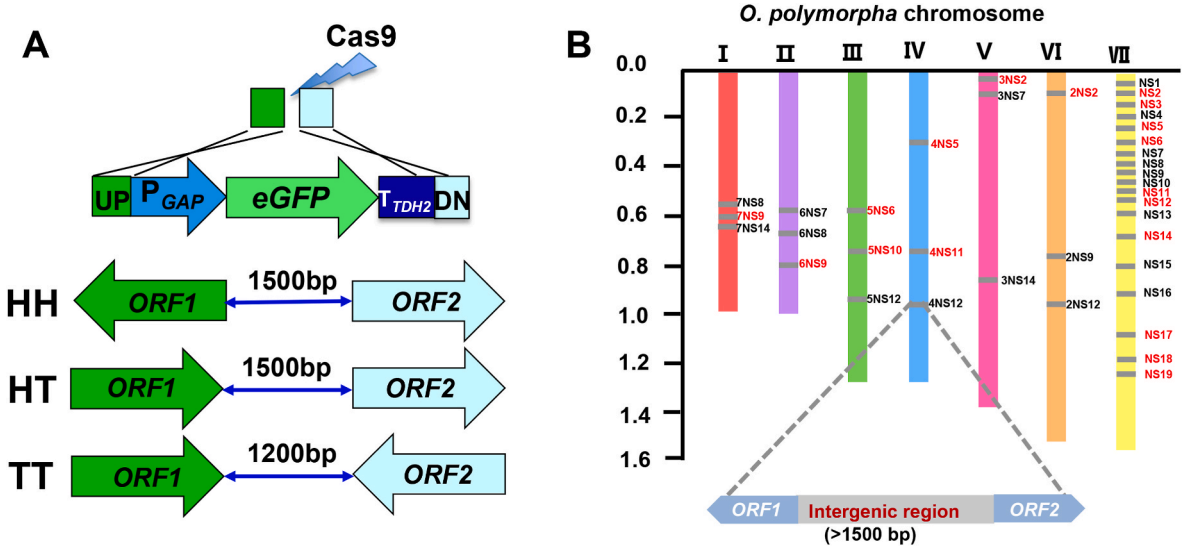

Fig. 1. The criteria for screening candidate neutral sites in *O. polymorpha* genome. (A) The scheme of integration of *eGFP* gene into candidate sites through CRISPR-Cas9 system. There are 3 categories of interval regions: HH (the head-to-head localization of two adjacent ORFs, which means that this interval region contains two promoters); HT (head-to-tail localization of two adjacent ORFs, which means that this interval region contains one promoter and one terminator); TT (tail-to tail localization of two ORFs, which means this interval region contains two terminators). (B) The localization of candidate neutral sites on *O. polymorpha* genome (The 18 sites that showed fluorescent signal were marked with red).

Table 1
Integration efficiency of *eGFP* cassette into 18 candidate sites. The efficiency was defined as the number of correct strains with green fluorescent intensity/10.

candidate neutral site	chromosome	position	guide RNA sequence	flanking gene direction	length (bp)	integration efficiency
NS2	VII	131978–133389	AATGGTGACATACTTTTATA	TT	1411	10%
NS3	VII	178019–180335	CAAGCGGCTGGAAAACAAGA	TT	2316	10%
NS5	VII	362598–364150	CCGACGATGATGGAAACTTT	HT	1552	10%
NS6	VII	376830–378761	CCCTCAACTTGTTATTTTAC	HH	1931	10%
NS11	VII	523032–524734	GACAGCAGTAGATGCAGCTG	HH	1702	40%
NS12	VII	535041–537172	AGTATATTTGTGAGTGACTC	HH	2131	30%
NS14	VII	700610–702392	GTTCGTCGCAAACAGCCGAT	HT	1782	30%
NS16	VII	996634–997983	ATGTAAACATTAAAACACGA	TT	1350	30%
NS18	VII	1089677–1090906	GTGGGTATATCGTACTATCC	TT	1230	30%
NS19	VII	1149379–1150807	TCGTGCTTGTCTCTGACGAA	TT	1428	40%
2NS2[a]	VI	316781–319322	TTTTTCCTAGACCCCTTGAA	TT	2541	10%
3NS2	V	11624–15232	TTGCAGAATCACTGCGCTAG	HH	3608	40%
4NS5	IV	308571–309919	GGAGCTGGCCTAGACTCTCT	TT	1348	10%
4NS11	IV	848607–864654	GTCAGAAAATATCAAAGTAG	TT	16947	10%
5NS6	III	679834–682584	AAAATGGACAGTTACAGAAA	TT	2750	20%
5NS10	III	981637–984032	CAAATGCTCGACGTGGACGA	TT	2395	40%
6NS9	II	796537–798545	GAACAAGATAGACGACCTAG	TT	2008	30%
7NS9	I	639609–642532	ATACCTTAACGAGGAACAAT	TT	2923	10%

[a] Represents the false sites as having no fluorescence in Delft-Methanol media.

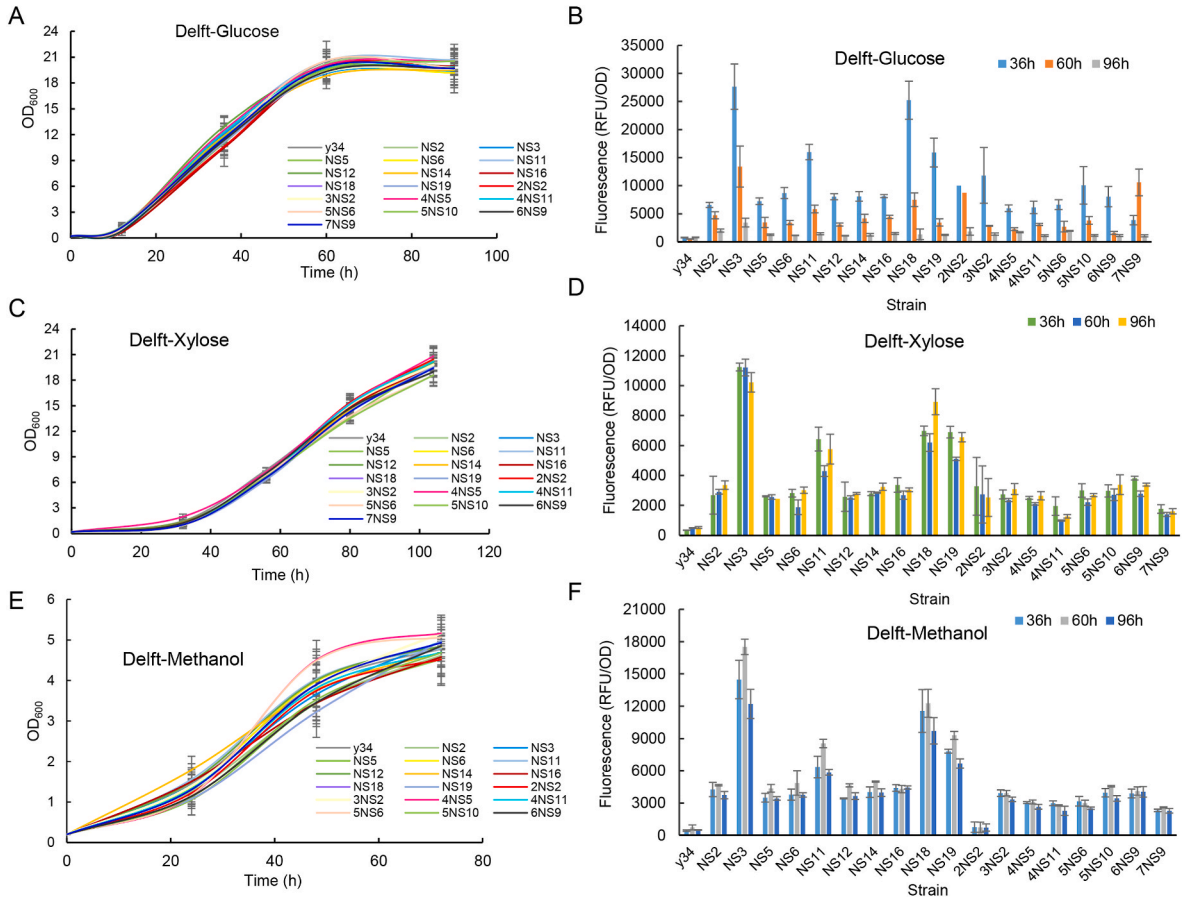

Fig. 2. Profiling the growth and fluorescent intensity of 18 NS strains and parent strain y34 (control) in different media. (A) The growth curve of corresponding strains in Delft-Glucose (20 g/L) media. (B) The fluorescent intensity of corresponding strains in Delft-Glucose media. (C) The growth curve of corresponding strains in Delft-Xylose (20 g/L) media. (D) The fluorescent intensity of corresponding strains in Delft-Xylose media. (E) The growth curve of corresponding strains in Delft-Methanol (10 g/L) media. (F) The fluorescent intensity of corresponding strains in Delft-Methanol media.

unknown reasons, which thus was excluded as a neutral site. The growth and fluorescence profiling of NS strains with parental y34 strain, revealed that 17 candidate sites can be recognized as neutral sites.

Metabolic engineering by using neutral sites

We finally applied the neutral sites for genome integration of a fatty alcohol synthetic pathway. Fatty alcohols, which has been widely used in lubricant, surfactant, personal care and pharmaceuticals [21], can be biosynthesized by the fatty acid reduction pathway [21]. The fatty

alcohol biosynthetic pathway (Fig. 3A) is composed of a carbolic acid reductase Car from *Mus musculus*, 4′-phosphopantetheinyl transferase NpgA from *Aspergillus nidulans* and aldehyde dehydrogenase Adh5 from *S. cerevisiae* [21]. The hexadecenal dehydrogenase gene *HFD1* and fatty acyl-CoA synthetase gene *FAA1* were deleted for blocking the reverse reactions (Fig. 3A), which has been showed to be essential for accumulation of the precursors fatty acids and fatty aldehydes [21]. It was observed that *FAA1* deletion caused severely decreased transformation and integration efficiency of *O. polymorpha* (Supplemental Fig. S1). We therefore used Δ*hfd1* strain as a host for integration of *CAR*, and *NpgA* + *ADH5* genes into the NS2 and NS3 neutral sites, respectively (Fig. 3B). Finally, the *FAA1* gene was deleted upon the integration of fatty alcohol biosynthetic pathway. The engineered strain FOH04 enabled the production of 4.5 mg/L fatty alcohol in Delft-Glucose medium, with the hexadecanol (C16:0) and (9Z)-octadecen-1-ol (C18:1) as the main products (Fig. 3B). And the control strain y34 produced no fatty alcohols (Supplemental Fig. S2). The success of construction of fatty alcohol synthetic pathway demonstrated the feasibility of the neutral sites screened in this study for metabolic engineering of *O. polymorpha* and expanded its genetic toolbox for synthetic biology of methylotrophic yeasts.

Discussion

O. polymorpha, a methylotrophic yeast that was previously employed for production of recombinant proteins [22], has drawn increasing attention as a cell factory for methanol biotransformation recently [4, 23–25]. Though CRISPR-Cas9 genome editing tools have been established for genetic manipulation in *O. polymorpha* [16,26–28], the lack of genome integration sites may hinder extensive metabolic rewiring for biosynthesis of chemicals from methanol. This study therefore adopted a systematic procedure to screen neutral sites in *O. polymorpha* and verified its applicability in pathway construction.

Though auxotropic marker loci have been used for gene integration, which however can't be recognized as neutral loci, because insertion genes into these site might perturb cell fitness. For example, we found that *LEU2* mutation significantly affects cell fitness (Supplemental Fig. S3). Therefore, *LEU2* loci cannot be used for integration of *eGFP* or other gene cassettes. Other selection markers may also affect cell fitness more or less.

The multiple copy of rDNA sites are another choice for genome integration, which might cause the integration of multiple genes at unpredictable manner and result in instability due to recombination [29]. Therefore, relatively large number of genomic interval regions are the best choice as neutral sites, which can avoid disrupting the transcription of functional genes when integrating gene cassettes. There are a limited number of neutral sites identified in prokaryotes such as Cyanobacterium *Synechocystis* PCC 6803 by genome and transcriptome sequencing [11,30], since prokaryotic genes are usually clustered with few base pairs in interval region that is not suitable for gene integration [31]. In contrast, there are more and longer interval regions in eukaryotes, providing convenient screening of sufficient neutral sites. We here showed that characterizing the interval regions of 1500 bp or 1200 bp enabled the identification 17 neutral sites by using CRISPR-Cas9 techniques. 17 sites are usually enough for metabolic engineering and synthetic biology in *O. polymorpha*. A previous study showed that 750 bp interval regions could be functional neutral sites in *S. cerevisiae* [10]. It can be anticipated that more neutral sites could be identified if we search the shorter interval regions. Anyway, no retardation on cellular growth suggested these neutral sites can be used for integration of multiple genes in *O. polymorpha*.

The variation of fluorescent intensity among NS strains integrated with *eGFP* gene may be caused by the variation of chromosome structure. This phenomenon has been also observed in *S. cerevisiae*, where integration of *lacZ* gene into different neutral sites resulted a 2.5-fold variation of β-galactosidase activity [10]. At the same time, each neutral site can be selected and combined with different promoters to fine-tune gene expression. We proved that genome integration sites influence expression level in *O. polymorpha*, which was also observed in *Escherichia coli* [32]. This kind of gene expression variation can be used for fine tuning the expression of corresponding genes in balancing metabolic pathways. Genomic integration of a fatty alcohol synthetic pathway genes into NS2 and NS3 sites, resulted in an engineered strain producing 4.5 mg/L fatty alcohol, demonstrating that neutral sites described here can be used in metabolic engineering of *O. polymorpha*.

In summary, we systematically characterized the interval regions by using the *eGFP* cassette, which enabled the identification of 17 neutral sites in *O. polymorpha*. These neutral sites should be helpful for construction of stable microbial cell factory for bio-manufacturing, and the screen procedure described here can be easily implemented in other eukaryotes.

CRediT authorship contribution statement

Wei Yu: Methodology, Investigation, Writing – original draft. **Jiaoqi Gao:** Methodology, Resources, Writing – review & editing. **Xiaoxin Zhai:** Methodology, Resources. **Yongjin J. Zhou:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

This work has been included in patent applications (Application number: 202010628649.3 and 202010626783.X) by Dalian Institute of

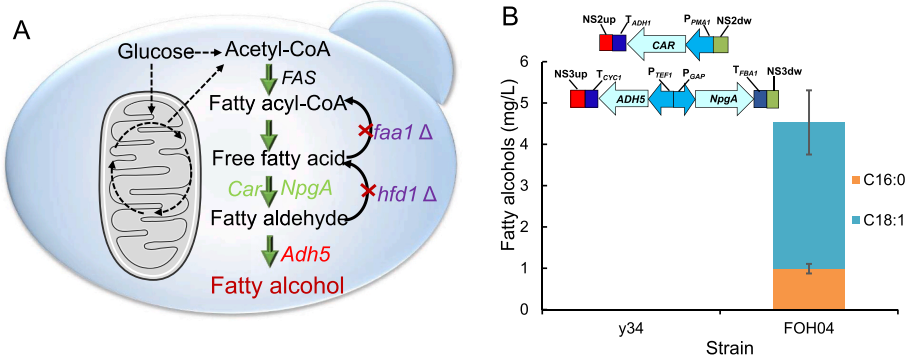

Fig. 3. Construction of fatty alcohol biosynthetic pathway by using neutral sites. A) The designed fatty alcohol biosynthetic pathway. B) The production of fatty alcohols in engineered *O. polymorpha* strain. CAR, the gene encoding carboxylic acid reductase from *M. musculus*; *NpgA*, the gene encoding 4′-phosphopantetheinyl transferase from *A. nidulans*; *ADH5*: the gene encoding alcohol dehydrogenase 5 from *S. cerevisiae*.

Chemical Physics, CAS.

Acknowledgements

This study was financially supported by National Natural Science Foundation of China (21922812 and 21808216), 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 Chemicals Physics, CAS.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.synbio.2021.03.001.

References

[1] Kurylenko O, Ruchala J, Hryniv O, Abbas C, Dmytruk K, Sibirny A. Metabolic engineering and classical selection of the methylotrophic thermotolerant yeast *Hansenula polymorpha* for improvement of high-temperature xylose alcoholic fermentation. Microb Cell Factories 2014;13:122–31.

[2] Ravin N, Eldarov M, Kadnikov V, Beletsky A, Schneider J, Mardanova E, et al. Genome sequence and analysis of methylotrophic yeast *Hansenula polymorpha* DL1. Microb Cell Factories 2013;14:837–56.

[3] Gao J, Zhou YJ. Advances in methanol bio-transformation. Syn Biol J 2020;1:158–73.

[4] Kurylenko OO, Ruchala J, Vasylyshyn RV, Stasyk OV, Dmytruk OV, Dmytruk KV, et al. Peroxisomes and peroxisomal transketolase and transaldolase enzymes are essential for xylose alcoholic fermentation by the methylotrophic thermotolerant yeast, *Ogataea* (*Hansenula*) *polymorpha*. Biotechnol Biofuels 2018;11:197.

[5] Gong Z, Nielsen J, Zhou YJ. Engineering robustness of microbial cell factories. Biotechnol J 2017;12.

[6] Kata I, Semkiv MV, Ruchala J, Dmytruk KV, Sibirny AA. Overexpression of the genes PDC1 and ADH1 activates glycerol conversion to ethanol in the thermotolerant yeast *Ogataea* (*Hansenula*) *polymorpha*. Yeast 2016;33:471–8.

[7] Peeters B, de Boer J, Bron S, Venema G. Structural plasmid instability in *Bacillus subtilis*: effect of direct and inverted repeats. Mol Gen Genet 1988;212:450–8.

[8] Kiewiet R, Kok J, Seegers J, Venema G, Bron S. The mode of replication is a major factor in segregational pasmid instability in *Lactococcus lactis*. Appl Environ Microbiol 1993;59:358–64.

[9] Chen N, Wang J, Zhao Y, Deng Y. Metabolic engineering of *Saccharomyces cerevisiae* for efficient production of glucaric acid at high titer. Microb Cell Factories 2018;17:67.

[10] Mikkelsen MD, Buron LD, Salomonsen B, Olsen CE, Hansen BG, Mortensen UH, Halkier BA. Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform. Metab Eng 2012;14:104–11.

[11] Ng AH, Berla BM, Pakrasi HB. Fine-tuning of photoautotrophic protein production by combining promoters and neutral sites in the Cyanobacterium *Synechocystis* sp. Strain PCC 6803. Appl Environ Microbiol 2015;81:6857–63.

[12] Chen R, Yang S, Zhang L, Zhou YJ. Advanced strategies for production of natural products in yeast. iScience 2020;23:100879.

[13] Yang S, Cao X, Yu W, Li S, Zhou YJ. Efficient targeted mutation of genomic essential genes in yeast *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol 2020;104:3037–47.

[14] Riley R, Haridas S, Wolfe K, Lopes M, Hittinger C, Goker M, Salamov A, Wisecaver J, Long T, Calvey C, Aerts A. Comparative genomics of biotechnologically important yeasts. Proc Natl Acad Sci USA 2016;113:9882–7.

[15] Verduyn C, Postma E, Scheffers A, Dijken J. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. Yeast 1992;8:501–17.

[16] Gao J, Gao N, Zhai X, Zhou YJ. Recombination machinery engineering for precise genome editing in methylotrophic yeast *Ogataea polymorpha*. iScience 2021;24:102168.

[17] Labun K, Montague TG, Krause M, Torres Cleuren YN, Tjeldnes H, Valen E. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. Nucleic Acids Res 2019;47:W171–4.

[18] Looke M, Kristjuhan K, Kristjuhan A. Extraction of genomic DNA from yeasts for PCR-based applications. Biotechniques 2011;50:325–8.

[19] Faber KN, Haima P, Harder W, Veenhuis M, Geert AB. Highly-efficient electrotransformation of the yeast *Hansenula polymorpha*. Curr Genet 1994;25:305–10.

[20] Zhou YJ, Gao W, Rong Q, Jin G, Chu H, Liu W, Yang W, Zhu Z, Li G, Zhu G, et al. Modular pathway engineering of diterpenoid synthases and the mevalonic acid pathway for miltiradiene production. J Am Chem Soc 2012;134:3234–41.

[21] Zhou YJ, Buijs NA, Zhu Z, Qin J, Siewers V, Nielsen J. Production of fatty acid-derived oleochemicals and biofuels by synthetic yeast cell factories. Nat Commun 2016;7:11709.

[22] Manfrao-Netto JHC, Gomes AMV, Parachin NS. Advances in using *Hansenula polymorpha* as chassis for recombinant protein production. Front Bioeng Biotechnol 2019;7:94.

[23] Clomburg JM, Crumbley AM, Gonzalez R. Industrial biomanufacturing: the future of chemical production. Science 2017;355.

[24] Zhou YJ, Kerkhoven EJ, Nielsen J. Barriers and opportunities in bio-based production of hydrocarbons. Nat Energy 2018;3:925–35.

[25] Duan X, Gao J, Zhou YJ. Advances in engineering methylotrophic yeast for biosynthesis of valuable chemicals from methanol. Chin Chem Lett 2018;29:681–6.

[26] Wang L, Deng A, Zhang Y, Liu S, Liang Y, Bai H, Cui D, Qiu Q, Shang X, Yang Z, et al. Efficient CRISPR-Cas9 mediated multiplex genome editing in yeasts. Biotechnol Biofuels 2018;11:277.

[27] Juergens H, Varela JA, de Vries ARG, Perli T, Gast VJM, Gyurchev NY, Rajkumar AS, Mans R, Pronk JT, Morrissey JP, Daran JMG. Genome editing in *Kluyveromyces* and *Ogataea* yeasts using a broad-host-range Cas9/gRNA co-expression plasmid. FEMS Yeast Res 2018;18.

[28] Cai P, Gao J, Zhou Y. CRISPR-mediated genome editing in non-conventional yeasts for biotechnological applications. Microb Cell Factories 2019;18:63.

[29] Wang L, Deng A, Zhang Y, Liu S, Liang Y, Bai H, Cui D, Qiu Q, Shang X, Yang Z, et al. Efficient CRISPR-Cas9 mediated multiplex genome editing in yeasts. Biotechnol Biofuels 2018;11:277.

[30] Wang M, Luan G, Lu X. Systematic identification of a neutral site on chromosome of *Synechococcus* sp. PCC7002, a promising photosynthetic chassis strain. J Biotechnol 2019;295:37–40.

[31] Yu W, Li X, Lu J, Xie G. Citrulline production by lactic acid bacteria in Chinese rice wine. J Inst Brew 2018;124:85–90.

[32] Goormans AR, Snoeck N, Decadt H, Vermeulen K, Peters G, Coussement P, et al. Comprehensive study on *Escherichia coli* genomic expression: does position really matter? Metab Eng 2020;62:10–9.