Electronic Supplementary Information

Construction and optimization of microbial cell factories for sustainable production of bioactive dammarenediol-II glucosides

Zong-Feng Hu[1], An-Di Gu[1], Lan Liang, Ting Gong, Yan Li, Jing-Jing Chen, Tian-Jiao Chen, Jin-Ling Yang[*] and Ping Zhu[*]

State Key Laboratory of Bioactive Substance and Function of Natural Medicines & NHC Key Laboratory of Biosynthesis of Natural Products, Institute of Materia Medica, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100050, China.

[*] Corresponding authors.

E-mail addresses: yangjl@imm.ac.cn (J. L. Yang), zhuping@imm.ac.cn (P. Zhu).

[1] These authors contributed equally to this work.

Table of contents

1 Supplementary materials

1.1 Plasmid and module construction

Construction of genome integration module Ⅰ: *δ1-1-P_{TEF1}-synDS-GFP-T_{CYC1}-P_{PGK1}-tHMG1-T_{ADH1}* (Fig. S1)

DS was synthesized according to the codon bias of *Saccharomyces cerevisiae* and fused with *GFP* by overlap extension PCR (OE-PCR) to generate pESC-HIS-DS-GFP. The *synDS-GFP* gene was amplified with primers DS-TEF1-F and DS-CYC1-R using pESC-HIS-DS-GFP as the template, *TEF1* promoter (P_{TEF1}) and *CYC1* terminator (T_{CYC1}) were amplified from the genomic DNA of *S. cerevisiae* INVSc1 using primer sets A-TEF1-Delta1-F/TEF1-DS-R and CYC1-DS-F/CYC1-PGK1-R. The purified fragments *TEF1*, *synDS-GFP*, *CYC1* were joined through OE-PCR using primers A-TEF1-Delta1-F and CYC1-PGK1-R to generate P_{TEF1}-*synDS-GFP*-T_{CYC1}, and cloned into pEASY-Blunt-Zero, resulting in pEASY-DS-GFP. Homologous recombination region of *δ* site was amplified from the genomic DNA of *S. cerevisiae* INVSc1 using primers Delta1-2F/Delta2-1R, and cloned into pEASY-Blunt-Simple, resulting in pEASY-INδ. Fragment *δ1-1* was amplified from plasmid pEASY-INδ using primers Delta1-2F/Delta1-TEF1-2R. Fragments *δ1-1* and P_{TEF1}-*synDS-GFP*-T_{CYC1} were joined through OE-PCR using primers Delta1-2F and CYC1-PGK1-R to generate *δ1-1*-P_{TEF1}-*synDS-GFP*- T_{CYC1}.

Fragment P_{PGK1}-*tHMG1*-T_{ADH1} was amplified as follows. The tHMG1 gene, *PGK1* promoter (P_{PGK1}) and *ADH1* terminator (T_{ADH1}) were amplified from the chromosome DNA of *S. cerevisiae* YPH499 using primer sets PGK1-CYC1-F/PGK1-tHMG1-R, tHMG1-PGK1-F/tHMG1-ADH1-R, ADH1-tHMG1-F/ADH1-R. The purified fragments *PGK1*, *ADH1*, *tHMG1* were joined through OE-PCR using primers PGK1-CYC1-F and ADH1-TDH3-R to generate P_{PGK1}-*tHMG1*-T_{ADH1}, and cloned into pEASY-Blunt-Zero, resulting in pEASY-tHMG1.

The purified fragments *δ1-1-*P_{TEF1}-*synDS-GFP*-T_{CYC1} and P_{PGK1}-*tHMG1*-T_{ADH1} were used as the template and joined through OE-PCR using primers Delta1-2F/ADH1-TDH3-R to generate the 6386 bp of genome integration module *δ1-1*-

P_{TEF1}-*synDS-GFP*-T_{CYC1}-P_{PGK1}-*tHMG1*-T_{ADH1}, and cloned into pEASY-Blunt-Zero, resulting in pEASY-3, then verified by DNA sequencing.

Construction of genome integration module Ⅱ: *overlap-*P_{TDH3}*-synPgUGT74AE2-*T_{ADH2}*-HIS-δ1-2* (Fig. S1)

To generate an overlap of approximately 400 to 500 bp, a fragment consisting of a sequence of the last 450 nucleotides was amplified from plasmid pEASY-3 using primers B-400F/ADH1-TDH3-R. *TDH3* promoter (P_{TDH3}) and *ADH2* terminator (T_{ADH2}) were amplified from the chromosome DNA of *S. cerevisiae* YPH499 using primer sets TDH3-ADH1-F/TDH3-UGT74AE2-R and ADH2-UGT74AE2-F/ADH2-HIS-R. The codon-optimized *synPgUGT74AE2* was amplified from the plasmid pUC57-PgUGT74AE2 with primers UGT74AE2-TDH3-F/UGT74AE2-ADH2-R. *TDH3*, *ADH2*, *synPgUGT74AE2* were fused by OE-PCR using primers TDH3-ADH1-F and ADH2-HIS-R, and cloned into pEASY-Blunt-Zero, resulting in pEASY-PgUGT74AE2. The purified 450 bp fragments and *TDH3p-synPgUGT74AE2-ADH2t* were joined through OE-PCR using primers B400-F and ADH2-HIS-R to generate *overlap-TDH3p-synPgUGT74AE2-ADH2t*.

The auxotrophic marker gene HIS was amplified using plasmid pESC-HIS as the template with primers HIS-ADH2-F and HIS-Delta2-R. Fragment *δ1-2* was amplified from plasmid pEASY-INδ using primers Delta2-HIS-1F/Delta2-1R. These two fragments was purified and joined through OE-PCR using primers HIS-ADH2-F and Delta2-1R to generate *HIS-δ1-2*.

The purified fragments *overlap-*P_{TDH3}*-synPgUGT74AE2-*T_{ADH2} and *HIS-δ1-2* were used as the template and joined through OE-PCR using primers B400-F/Delta2-1R to generate the 4571 bp of genome integration module *overlap-TDH3p-synPgUGT74AE2-ADH2t-HIS-δ1-2*, and cloned into pEASY-Blunt-Zero, resulting in pEASY-1, then verified by DNA sequencing.

Construction of genome integration module Ⅲ: *overlap-*P_{TDH3}*-synUGTPg1-*T_{ADH2}*-HIS-δ1-2* (Fig. S1)

To generate an overlap of approximately 400 to 500 bp, a fragment consisting of a sequence of the last 450 nucleotides was amplified from plasmid pEASY-3 using primers B-400F/ADH1-TDH3-R. *TDH3* promoter (P_{TDH3}) and *ADH2* terminator (T_{ADH2}) were amplified from the chromosome DNA of *S. cerevisiae* YPH499 using

primer sets TDH3-ADH1-F/TDH3-UGTPg1-R and ADH2-UGTPg1-F/ADH2-HIS-R. The codon-optimized *UGTPg1* was amplified from the plasmid pUC57-UGTPg1 with primers UGTPg1-TDH3-F/UGTPg1-ADH2-R. *TDH3*, *synUGTPg1*, *ADH2* were fused by OE-PCR using primers TDH3-ADH1-F and ADH2-HIS-R, and cloned into pEASY-Blunt-Zero, resulting in pEASY-UGTPg1. The purified 450 bp fragments and P_{TDH3}-*synUGTPg1*-T_{ADH2} were joined through OE-PCR using primers B400-F and ADH2-HIS-R to generate *overlap*-P_{TDH3}-*synUGTPg1*-T_{ADH2}.

The purified fragments *overlap*-P_{TDH3}-*synUGTPg1*-T_{ADH2} and *HIS*-*δ1-2* were used as the template and joined through OE-PCR using primers B400-F/Delta2-1R to generate the 4646 bp of genome integration module *overlap*-P_{TDH3}-*synUGTPg1*-T_{ADH2}-*HIS*-*δ1-2* and cloned into pEASY-Blunt-Zero, resulting in pEASY-2, then verified by DNA sequencing.

Construction of genome integration moduleⅣ: *δ4-1*-P_{TDH3}-*IDI1*-T_{TPI1}-P_{PGK1}-*ERG20*-T_{ADH1} (Fig. S1)

The *δ4-1* sequence was amplified from the genomic DNA of *S. cerevisiae* YPH499 using primers TY4-F1 and TY4-R2, and cloned into pEASY-Blunt-Simple, resulting in pEASY-TY4. PCR amplification of the 282 bp of the *δ4* sequence of TY4 transposon was performed with primers TY4-F1 and TY4-1-TDH3-R from pEASY-TY4. The *IDI1* gene, *TDH3* promoter (P_{TDH3}) and *TPI1* terminator (T_{TPI1}) were amplified from pEASY-TY4 using primer sets TDH3-TY4-1-F/TDH3-IDI1-R, IDI1-TDH3-F/IDI1-TPI1-R and TPI1-IDI1-F/TPI1-PGK1-R, and fused by OE-PCR with primers TDH3-TY4-1-F/TPI1-PGK1-R, and cloned into pEASY-Blunt-Zero, resulting in pEASY-IDI1. The purified fragments *δ4-1* and P_{TDH3}-*IDI1*-T_{TPI1} were joined through OE-PCR using primers TY4-F1 and TPI1-PGK1-R to generate *δ4-1*-P_{TDH3}-*IDI1*-T_{TPI1}.

The *ERG20* gene, *PGK1* promoter (P_{PGK1}) and *ADH1* terminator (T_{ADH1}) were amplified from the chromosome DNA of *S. cerevisiae* YPH499 using primer sets PGK1-TPI1-F/PGK1-ERG20-R, ERG20-PGK1-F/ERG20-ADH1-R and ADH1-ERG20-F/ADH1-TEF1-R. The purified fragments *PGK1*, *ADH1*, *ERG20* were joined through OE-PCR using primers PGK1-TPI1-F and ADH1-TEF1-R to generate P_{PGK1}-*ERG20*-T_{ADH1}, and cloned into pEASY-Blunt-Zero, resulting in pEASY-ERG20.

The purified fragments *δ4-1*-P_{TDH3}-*IDI1*-T_{TPI1} and P_{PGK1}-*ERG20*-T_{ADH1} were used as the template and joined through OE-PCR using primers TY4-F1/ADH1-TEF1-R to

generate the 4282 bp of genome integration module *δ4-1-*P_{TDH3}*-IDI1-*T_{TPI1}-P_{PGK1}-*ERG20-*T_{ADH1}, and cloned into pEASY-Blunt, resulting in pEASY-S28, then verified by DNA sequencing.

Construction of genome integration module Ⅵ: *P_{PGK1}-ERG1-T_{ADH1}-P_{TEF1}-ERG7$^-$-T_{CYC1}-LEU-δ4-2* (Fig. S1)

PCR amplification of the *ERG1* gene, *PGK1* promoter (P_{PGK1}) and *ADH1* terminator (T_{ADH1}) was performed from the chromosome DNA of *S. cerevisiae* YPH499 with primer sets ERG1-PGK1-F/ERG1-ADH1-R, PGK1-CYC1-F/PGK1-ERG1-R and ADH1-ERG1-F/ADH1-TEF1-R. The purified fragments *PGK1*, *ADH1*, *ERG1* were joined through OE-PCR using primers PGK1-CYC1-F and ADH1-TEF1-R to generate P_{PGK1}-*ERG1-*T_{ADH1}, and cloned into pEASY-Blunt-Zero, resulting in pEASY-ERG1.

The *ERG7* antisense gene was amplified with primers ERG7-TEF1-F/ERG7-CYC1-R using plasmid pESC-URA-ERG7$^-$ as the template. *TEF1* promoter (P_{TEF1}) and *CYC1* terminator (T_{CYC1}) were amplified from the chromosome DNA of *S. cerevisiae* YPH499 using primer sets TEF1-ADH1-F /TEF1-ERG7-R and CYC1-ERG7-F/CYC1-LEU-R. The purified fragments *TEF1*, *CYC1*, *ERG7*$^-$ were joined through OE-PCR using primers TEF1-ADH1-F and CYC1-LEU-R to generate P_{TEF1}-*ERG7*$^-$-T_{CYC1}, and cloned into pEASY-Blunt-Zero, resulting in pEASY-ERG7$^-$.

PCR amplification of the 234 bp of the *δ4* sequence was performed with primers TY4-2-LEU-F and TY4-2R using YPH499 chromosome DNA as the template. PCR amplification of the auxotrophic marker *LEU* was performed with primers LEU-CYC1-F and LEU-TY4-2-R using plasmid pESC-LEU (Invitrogen, USA) as the template. The purified fragments *LEU*, *δ4-2* were joined through OE-PCR using primers LEU-CYC1-F and TY4-R2 to generate fragment *LEU-δ4-2*.

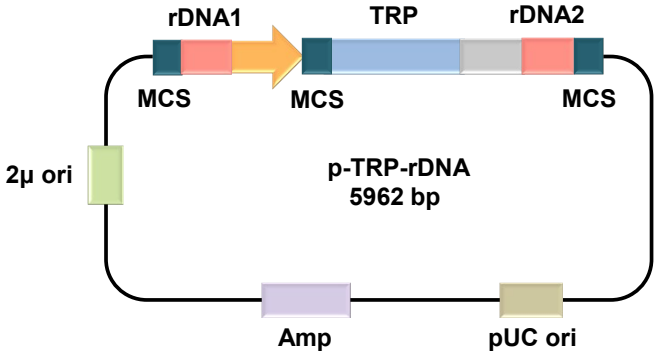
The purified fragments P_{PGK1}-*ERG1-*T_{ADH1}, P_{TEF1}-*ERG7*$^-$-T_{CYC1} and *LEU-δ4-2* were used as the template and joined through OE-PCR using primers PGK1-CYC1-F and TY4-2R to generate 6925 bp of genome integration module P_{PGK1}-*ERG1-*T_{ADH1}-P_{TEF1}-*ERG7*$^-$-T_{CYC1}-*LEU-δ4*-2, and cloned into pEASY-Blunt, resulting in pEASY-S1319, then verified by DNA sequencing.

Construction of genome integration module Ⅴ: *overlap-P_{TEF1}-ERG9-T_{CYC1}-overlap* (Fig. S1)

To generate an overlap of approximately 300 to 500 bp, a fragment consisting of a sequence of the last 515 nucleotides was amplified from plasmid pEASY-S28 using primers S28-400F/TEF1-ERG9-R. Another fragment consisting of a sequence of the first 550 nucleotides was amplified from plasmid pEASY-S1319 using primers PGK1-CYC1-F/S1319-400R. PCR amplification of the *ERG9* gene, *TEF1* promoter (P_{TEF1}) and *CYC1* terminator (T_{CYC1}) was performed with primer sets TEF1-ADH1-F/TEF1-ERG9-R, ERG9-TEF1-F/ERG9-CYC1-R and CYC1-ERG9-F/CYC1-PGK1-R from the chromosome DNA of *S. cerevisiae* YPH499. Then 100 ng each of the purified fragments *TEF1*, *CYC1*, *ERG9* were joined through OE-PCR using primers TEF1-ADH1-F and CYC1-PGK1-R to generate P_{TEF1}-*ERG9*-T_{CYC1}, and cloned into pEASY-Blunt-Zero, resulting in pEASY-ERG9.

The purified fragments mentioned above were used as the template and joined through OE-PCR using primers S28-400F and S1319-400R to give the genome integration module *overlap*-P_{TEF1}-*ERG9*-T_{CYC1}-*overlap*, and cloned into pEASY-Blunt, resulting in pEASY-S813, then verified by DNA sequencing.

Construction of plasmid p-TRP-rDNA:



The *rDNA* sequence was amplified from the genomic DNA of *S. cerevisiae* YPH499 using primers rDNA1-MQWD-F and rDNA2-MQWD-R, resulting in pEASY-rDNA. PCR amplification of the 766 bp of the *rDNA* sequence was performed with primers GJ-rDNA1-2μ-F and rDNA1-MQWD-R from pEASY-rDNA, and PCR amplification of the 628 bp of the *rDNA* sequence was performed with primers rDNA2-TRP-F and GJ-rDNA2-PUC-R from pEASY-rDNA. The auxotrophic marker *TRP* was amplified with primers TRP-MQWD-F and TRP-rDNA2-R using plasmid pESC-TRP (Invitrogen, USA) as the template. The plasmid scaffold was amplified with primers PUC-GJ-rDNA2-F and 2μ-GJ-rDNA1-R using plasmid pESC-TRP (Invitrogen, USA) as the template. The four purified fragments mentioned above

were joined using the In-Fusion Cloning Kit (Clontech), resulting in p-TRP-rDNA, then verified by DNA sequencing.

Construction of genome integration module Ⅶ: *rDNA1-P_{TEF1}-BiP-T_{CYC1}-TRP-rDNA2* (Fig. S1)

The *BiP* gene, *TEF1* promoter (P_{TEF1}) and *CYC1* terminator (T_{CYC1}) were amplified from the chromosome DNA of *S. cerevisiae* YPH499 using primer sets GJ-F/TEF1-BIP-R, BIP-TEF1-F/BIP-CYC1-R and CYC1-BIP-F/GJ-R. The purified fragments *BiP*, *TEF1*, *CYC1* were joined through OE-PCR using primers GJ-F and GJ-R to generate P_{TEF1}-*BiP*-T_{CYC1}, and cloned into pEASY-Blunt-Zero, resulting in pEASY-BiP. Fragment P_{TEF1}-*BiP*-T_{CYC1} was amplified from pEASY-BiP using primers GJ-F and GJ-R, and cloned into plasmid p-TRP-rDNA at *Sal* I and *Xho* I sites, resulting in p-TRP-BiP, then verified by DNA sequencing. DNA integration module Ⅶ was obtained from plasmid p-TRP-BiP digested with *Bam*H I and *Sac* I.

Construction of genome integration module Ⅷ: *rDNA1-P_{TEF1}-HAC1-T_{CYC1}-TRP -rDNA2* (Fig. S1)

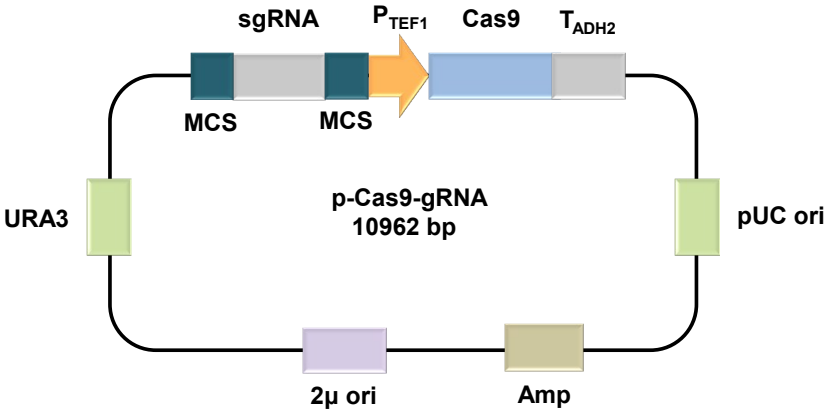
The *HAC1* gene, *TEF1* promoter (P_{TEF1}) and *CYC1* terminator (T_{CYC1}) were amplified from the chromosome DNA of *S. cerevisiae* YPH499 using primer sets GJ-F/TEF1-HAC1-R, HAC1-TEF1-F/HAC1-CYC1-R and CYC1-HAC1-F/GJ-R. The purified fragments *HAC1*, *TEF1*, *CYC1* were joined through OE-PCR using primers GJ-F and GJ-R to generate P_{TEF1}-*HAC1*-T_{CYC1}, and cloned into pEASY-Blunt-Zero, resulting in pEASY-HAC1. Fragment P_{TEF1}-*HAC1*-T_{CYC1} was amplified from pEASY-HAC1 using primers GJ-F and GJ-R, and cloned into plasmid p-TRP-rDNA at *Sal* I and *Xho* I sites, resulting in p-TRP-HAC1, then verified by DNA sequencing. DNA integration module Ⅷ was obtained from plasmid p-TRP-HAC1 digested with *Bam*H I and *Sac* I.

Construction of genome integration module Ⅸ: *rDNA1-P_{TEF1}-PDI1-T_{CYC1}-TRP -rDNA2* (Fig. S1)

The *PDI1* gene, *TEF1* promoter (P_{TEF1}) and *CYC1* terminator (T_{CYC1}) were amplified from the genome DNA of *S. cerevisiae* YPH499 using primer sets GJ-F/TEF1-PDI-R, PDI1-TEF1-F/HAC1-CYC1-R and CYC1-PDI1-F/GJ-R. The purified fragments *PDI1*, *TEF1*, *CYC1* were joined through OE-PCR using primers GJ-F and GJ-R to generate P_{TEF1}-*PDI1*-T_{CYC1}, and cloned into pEASY-Blunt-Zero,

resulting in pEASY-PDI1. Fragment P_{TEF1}-*PDI1*-T_{CYC1} was amplified from pEASY-PDI1 using primers GJ-F and GJ-R, and cloned into plasmid p-TRP-rDNA at *Sal* I and *Xho* I sites, resulting in p-TRP-PDI1, then verified by DNA sequencing. DNA integration module Ⅸ was obtained from plasmid p-TRP-PDI1 digested with *Bam*H I and *Sac* I.

Construction of p-Cas9-gRNA:



For the construction of Cas9 and gRNA co-expressing plasmid, the 4272 bp of human codon-optimized *cas9* was amplified with primers Cas9-TEF1-F and Cas9-ADH2-R from plasmid FM-1, and *TEF1* promoter (P_{TEF1}) and *ADH2* terminator (T_{ADH2}) were amplified from the chromosome DNA of *S. cerevisiae* YPH499 using primer sets TEF1-SUP4t-MSC-F/TEF1-Cas9-R and ADH2-Cas9-F/ADH2-pESC-R. The *δ* site specific gRNA expressed cassette with *SNR52* promoter and *SUP4* terminator was amplified from plasmid pUC57-gRNA using primers SNR52p-MSC-pESC-F and SUP4t-MSC-TEF1-R. The purified fragments *gRNA*, *TEF1*, *Cas9*, *ADH2* were joined through OE-PCR using primers SNR52p-MSC-pESC-F and ADH2-pESC-R to generate *gRNA*-P_{TEF1}-*Cas9*-T_{ADH2}. Plasmid backbone sequence was amplified from pESC-URA using primers pESC-ADH2-F/pESC-SNR52P-MCS-R. Then fragments *gRNA*-P_{TEF1}-*Cas9*-T_{ADH2} and plasmid backbone were joined using the In-Fusion Cloning Kit (Clontech), resulting in Cas9 and gRNA co-expressing plasmid p-Cas9-gRNA, then verified by DNA sequencing.

1.2 Gene deletion cassette construction

Gene *HXK2* deletion cassette was constructed as follows:

The genome DNA of *S. cerevisiae* YPH499 was used as template to obtain the

upstream and downstream sequences of gene *HXK2*, *SER3* and *SOR1*, respectively. The 414 bp *HXK2* upstream sequence was amplified by primers HXK2-1F and HXK2-1R, and 386 bp *HXK2* downstream sequence was amplified by primers HXK2-2F and HXK2-2R. The primers contained about 20 bp homologous sequences on their respective 5′ ends to create the overlap regions around the *Kan*MX cassette of the pUC6 plasmid. The *Kan*MX cassette was amplified using primers Kan-F and Kan-R. The three purified fragments were joined through OE-PCR using primers HXK2-1F and HXK2-2R, and cloned into pEASY-Blunt-Simple, resulting in plasmid pEASY-HXK2, then verified by DNA sequencing.

Gene *SER3* deletion cassette was constructed as follows:

The 390 bp *SER3* upstream sequence was amplified by primers SER3-1F and SER3-1R, and 388 bp *SER3* downstream sequence was amplified by primers SER3-2F and SER3-2R. The two purified fragments and *Kan*MX cassette were joined through OE-PCR using primers SER3-1F and SER3-2R, and cloned into pEASY-Blunt-Simple, resulting in plasmid pEASY-SER3, then verified by DNA sequencing.

Gene *SOR1* deletion cassette was constructed as follows:

The 427 bp *SOR1* upstream sequence was amplified by primers SOR1-1F and SOR1-1R, and 334 bp *SOR1* downstream sequence was amplified by primers SOR1-2F and SOR1-2R. The two purified fragments and *Kan*MX cassette were joined through OE-PCR using primers SOR1-1F and SOR1-2R, and cloned into pEASY-Blunt-Simple, resulting in plasmid pEASY-SOR3, then verified by DNA sequencing.

2 Supplementary figures

Fig. S1 The design and construction of integration modules for *S. cerevisiae* transformation

The DNA integration modules used in this study were designed to be integrated into one of three chromosomal loci in the yeast genome. Blue blocks indicate gene expression cassettes with a promoter, DNA coding sequence, and a terminator. Red blocks indicate the upstream and downstream sequences of chromosomal loci. Gray blocks indicate about 500 bp DNA homologous sequences of the adjacent modules. The genes synthesized according to the codon bias of yeast are labeled in red body. The plasmids constructed in this study are listed in Table S1 and primers used in construction of DNA integration modules are listed in Table S2.

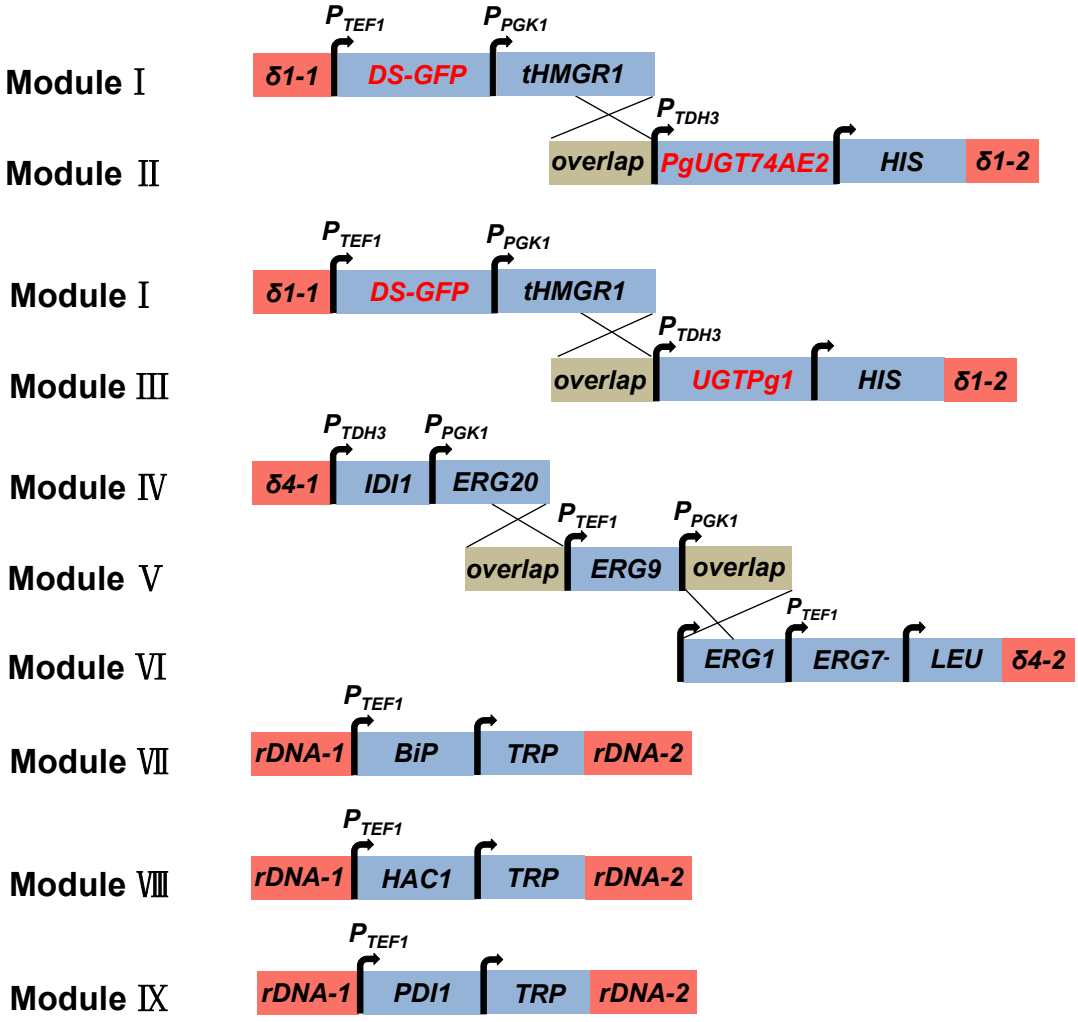


Fig. S2 SDS-PAGE of the cell extracts of *E. coli* recombinants

1: soluble proteins of pET32a-UGTPg1-transformed *E.coli* BL21 (DE3). 2: soluble proteins of pET32a-PgUGT74AE2-transformed *E.coli* BL21 (DE3). 3: soluble proteins of pET32a-transformed *E.coli* BL21 (DE3).

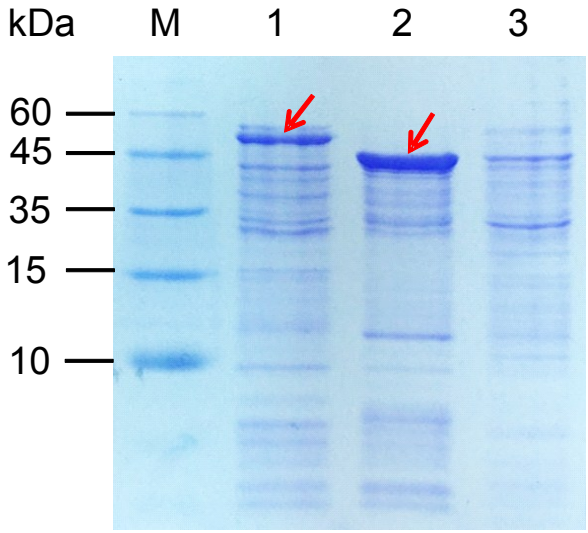
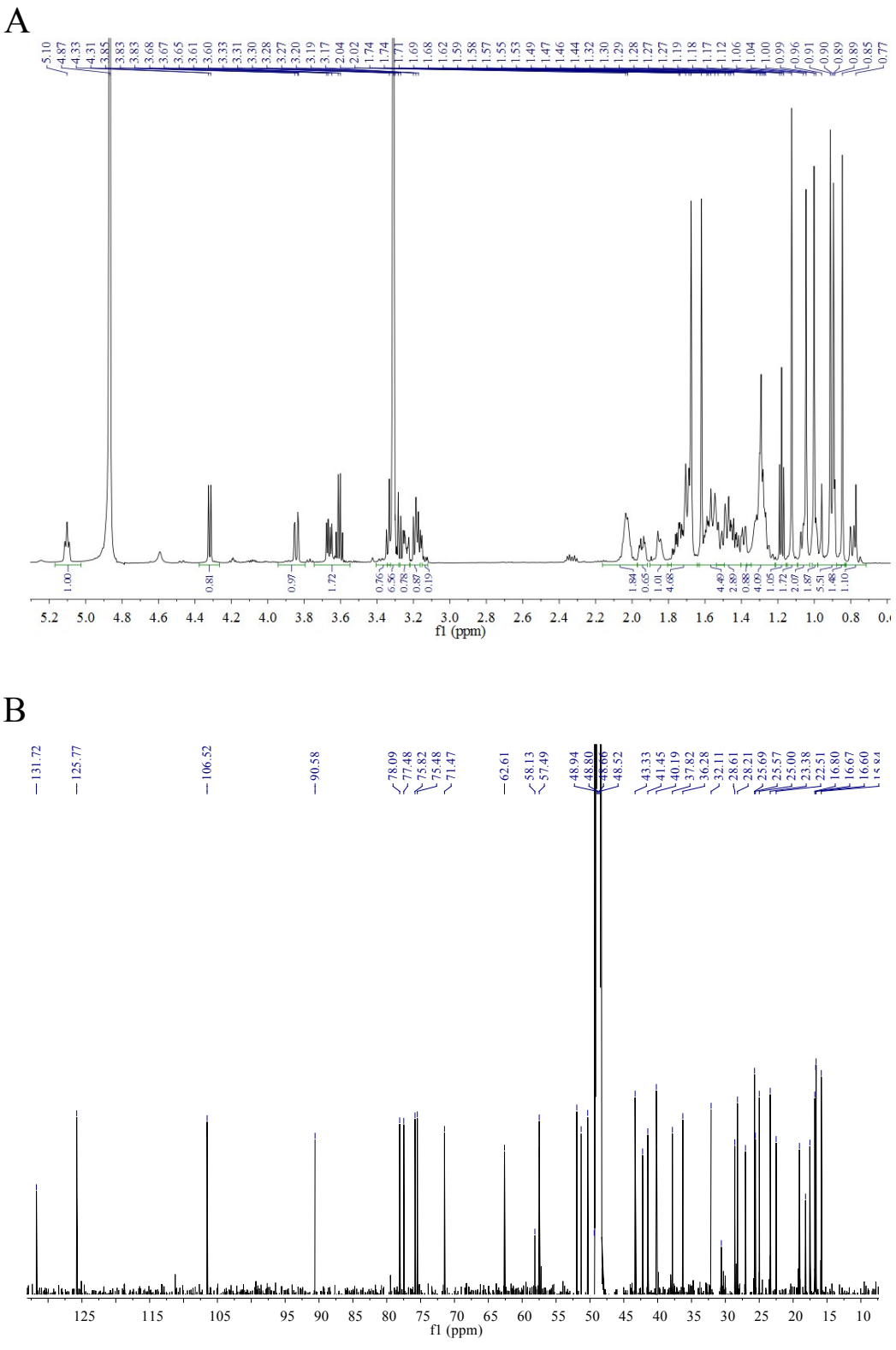
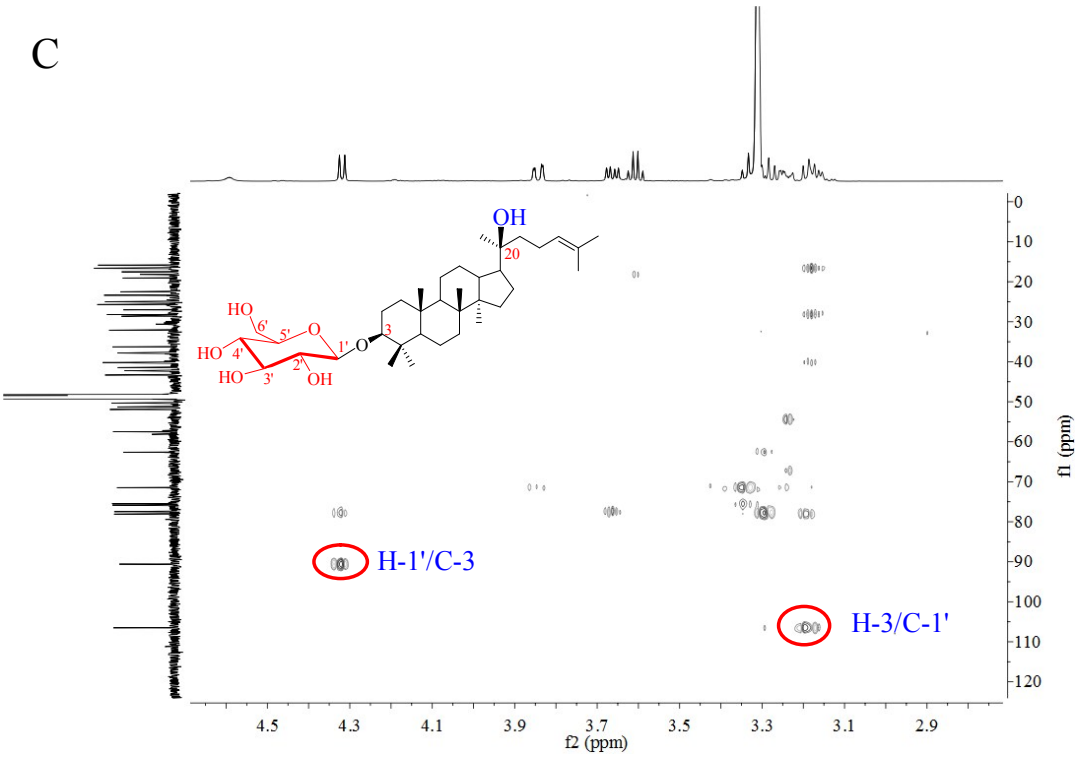


Fig. S3 The ^{1}H NMR (A), ^{13}C NMR (B), HMBC (C) and HSQC (D) spectra of 3β-*O*-Glc-DM produced by *in vitro* reaction in methanol-d_4



C


OH
20
HO
6'
5'
O
1'
HO
4'
2'
OH
HO
3'
3
H-1'/C-3
H-3/C-1'
f2 (ppm)
f1 (ppm)

D

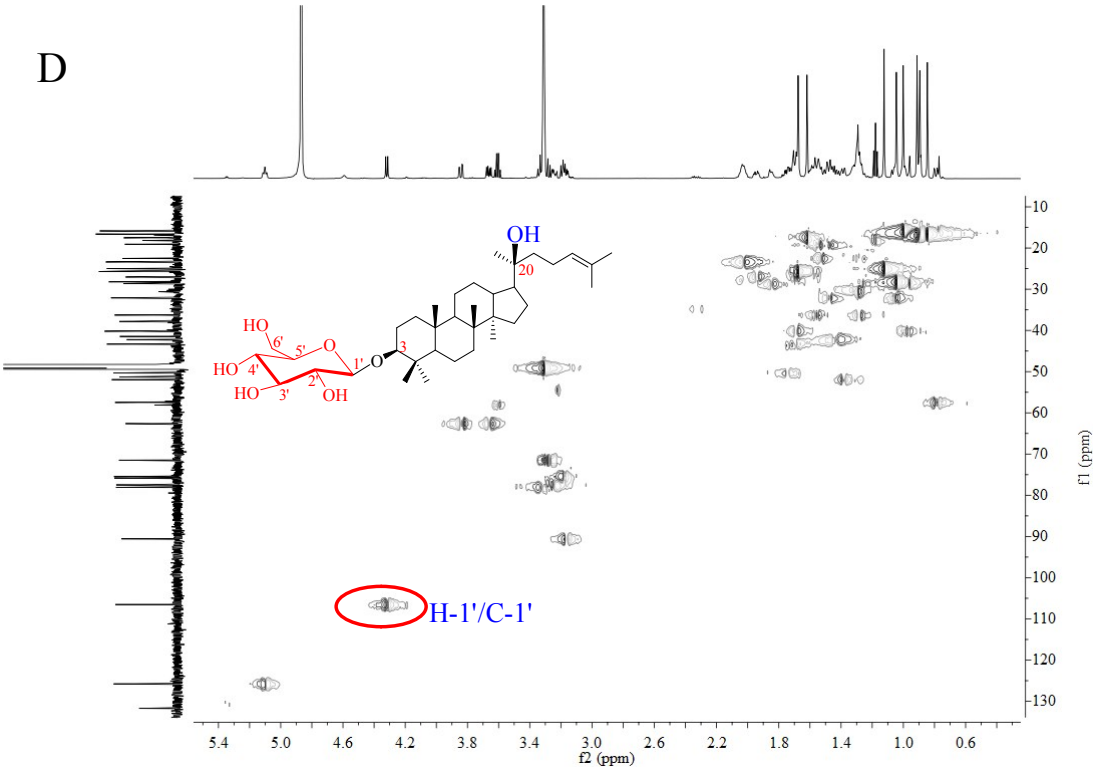
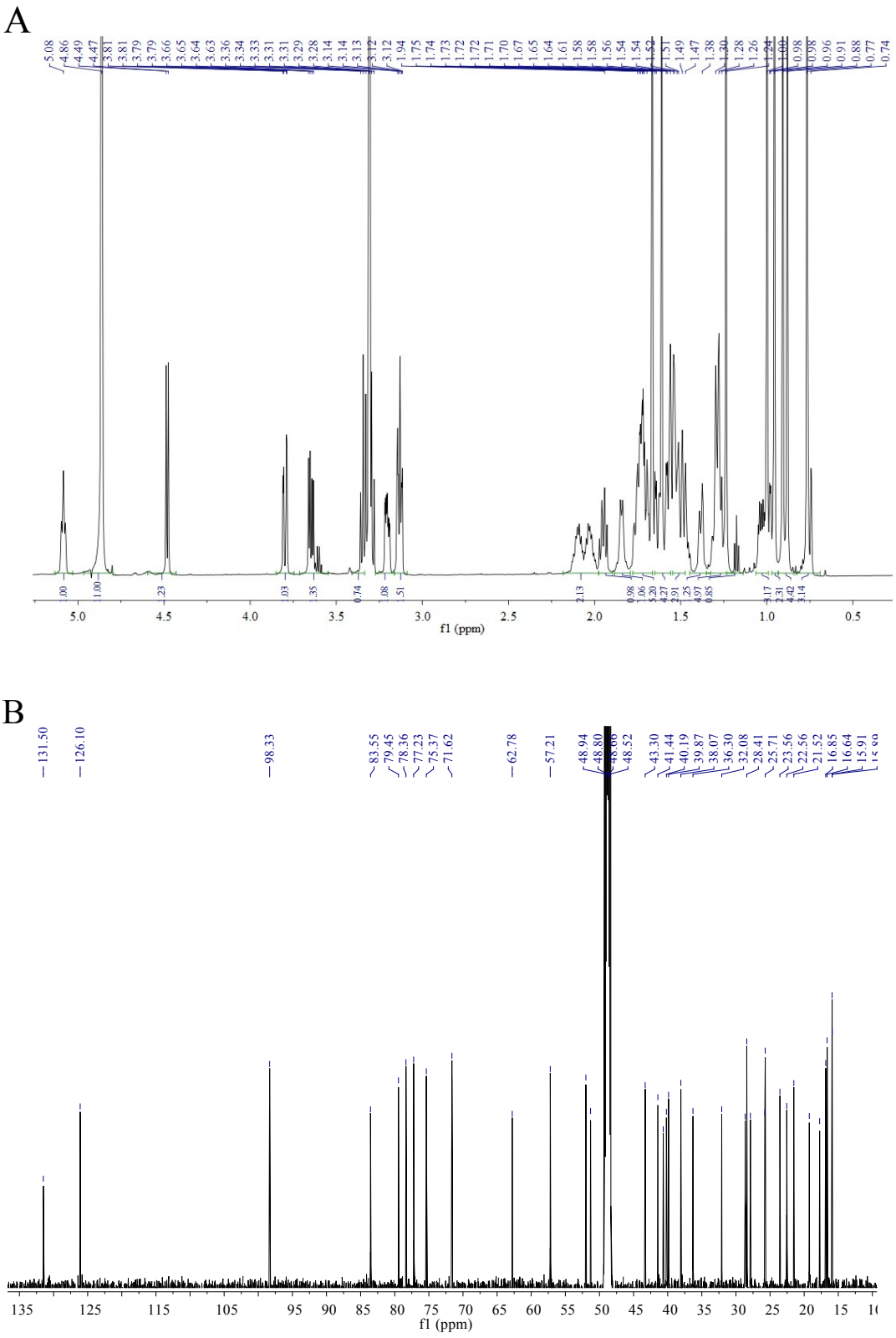
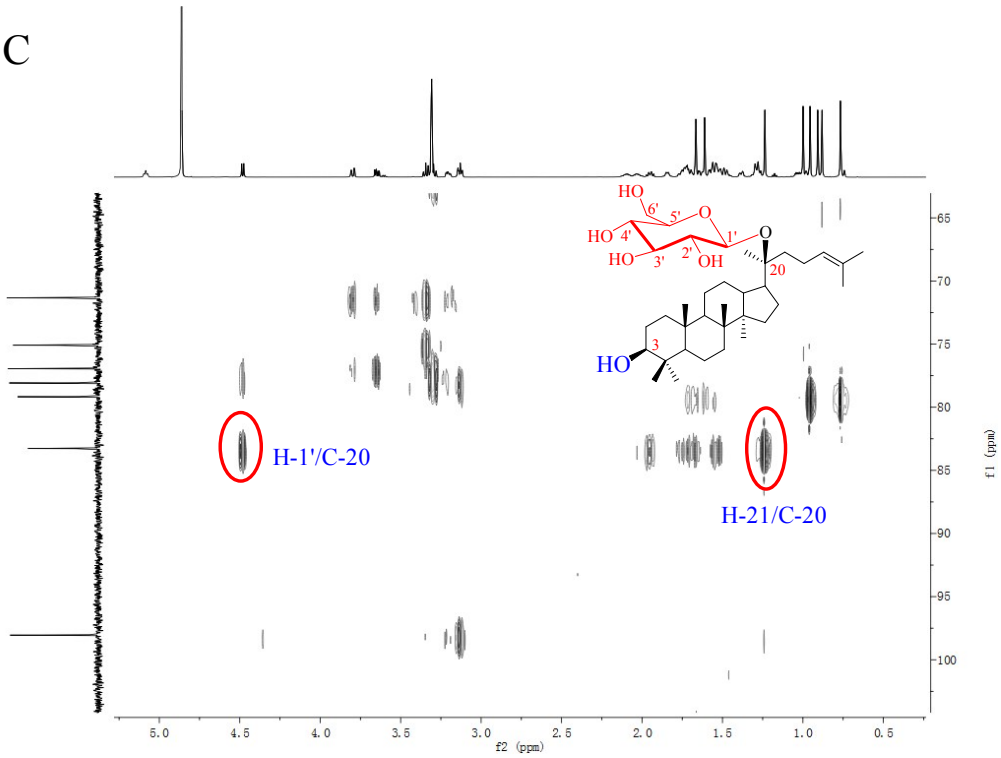

OH
20
HO
6'
5'
O
1'
HO
4'
2'
OH
HO
3'
3
H-1'/C-1'
f2 (ppm)
f1 (ppm)

Fig. S4 The ^{1}H NMR (A), ^{13}C NMR (B), HMBC (C) and HSQC (D) spectra of 20*S*-*O*-Glc-DM produced by *in vitro* reaction in methanol-d_4



C


HO
6'
5'
O
HO
4'
1'
O
HO
3'
2'
OH
20
HO
3
H-1'/C-20
H-21/C-20
f1 (ppm)
f2 (ppm)

D

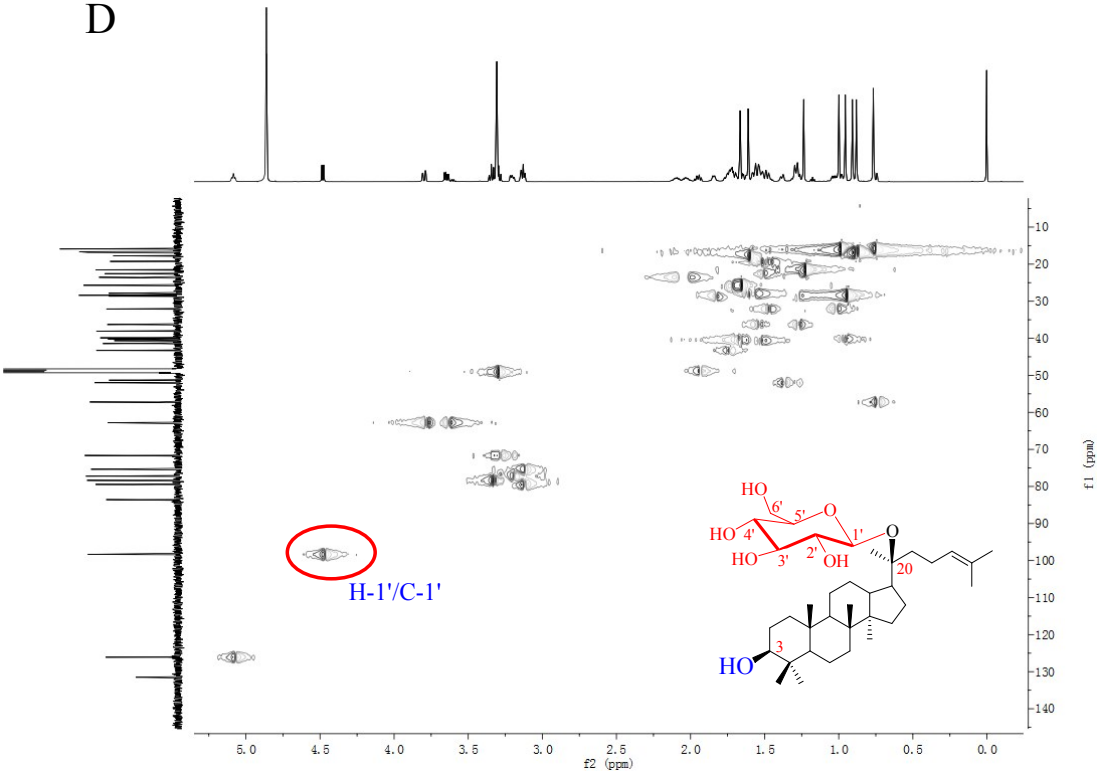
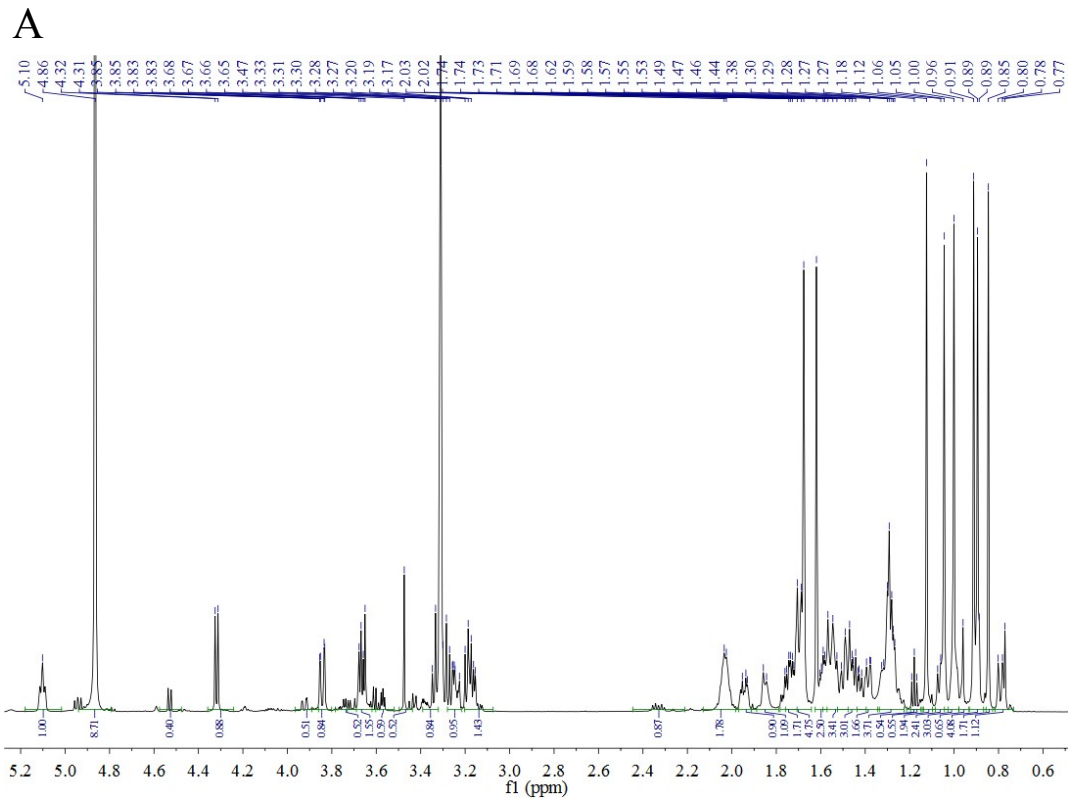
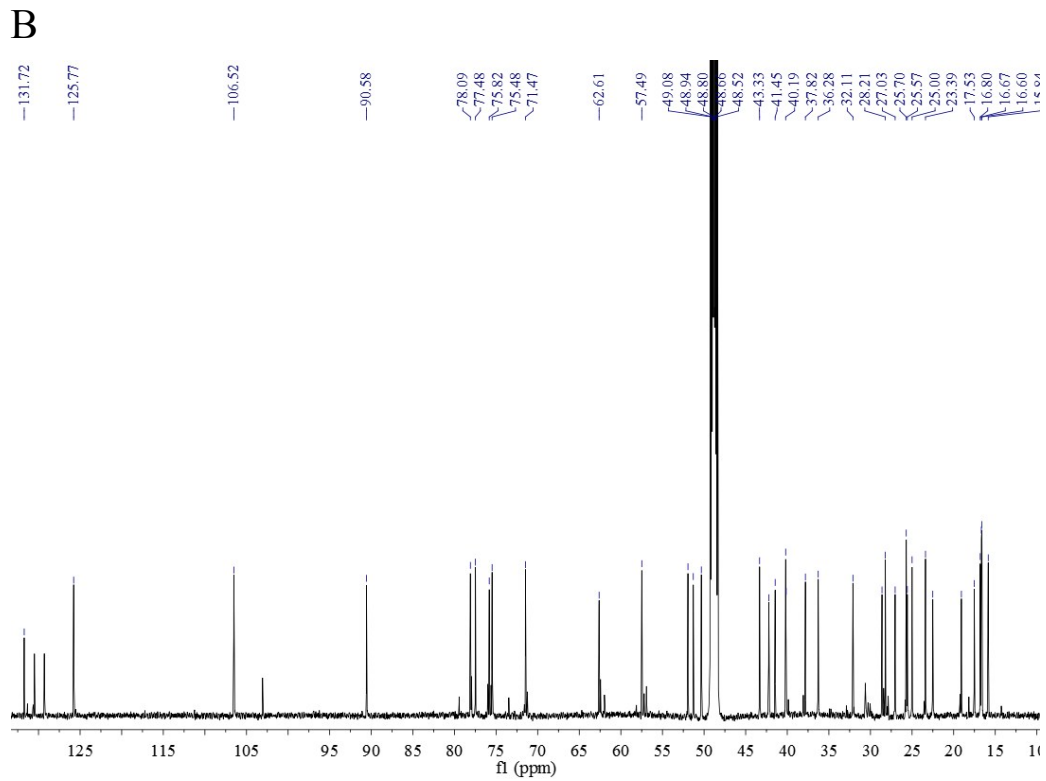
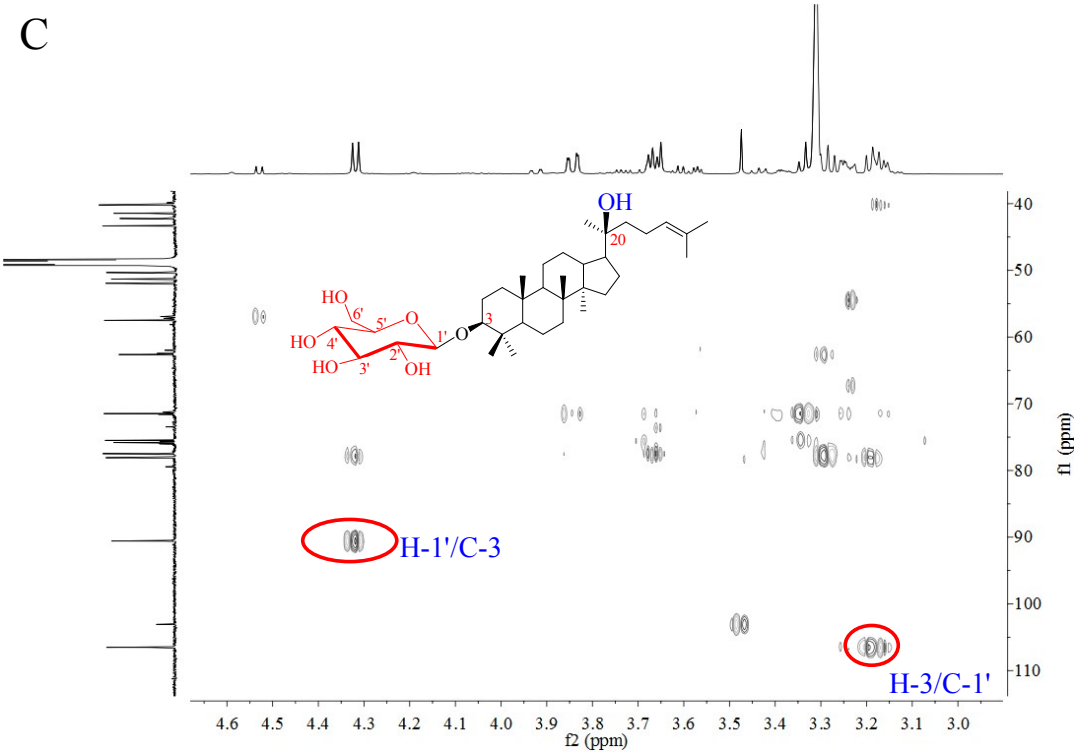

HO
6'
5'
O
HO
4'
1'
O
HO
3'
2'
OH
20
HO
3
H-1'/C-1'
f1 (ppm)
f2 (ppm)

Fig. S5 The ^{1}H NMR (A), ^{13}C NMR (B), HMBC (C) and HSQC (D) spectra of 3β-*O*-Glc-DM produced by engineered yeast in methanol-d_4





C


OH
20
HO
6'
5'
O
HO
4'
1'
O
3
HO
3'
2'
OH
H-1'/C-3
H-3/C-1'
f1 (ppm)
f2 (ppm)

D

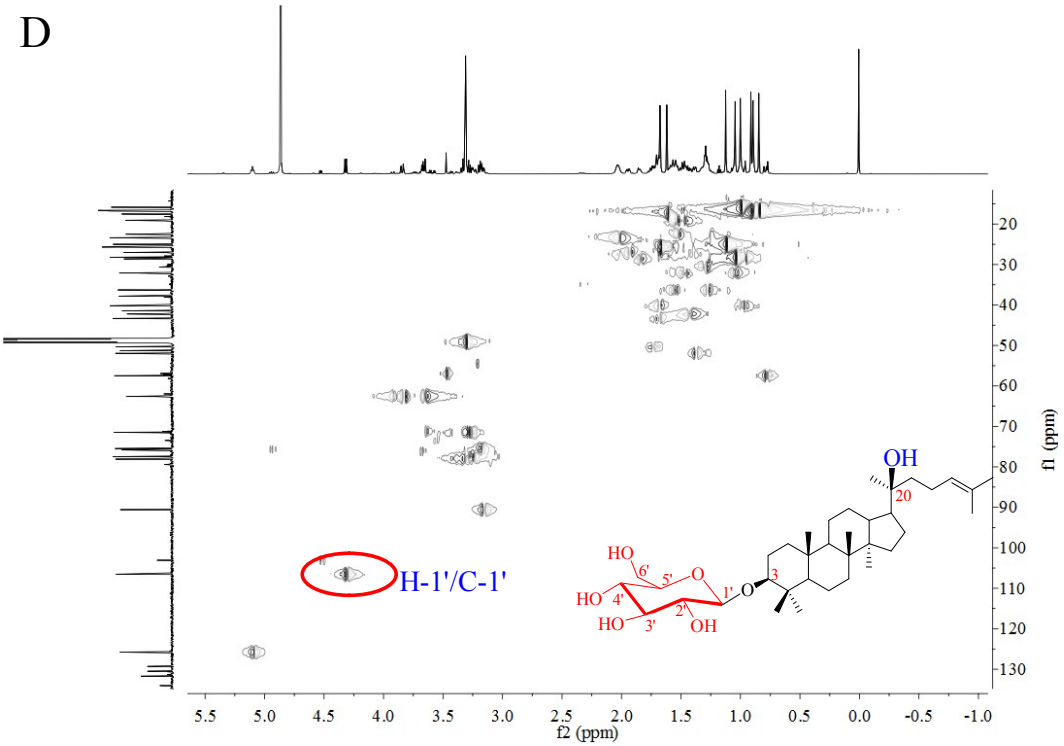
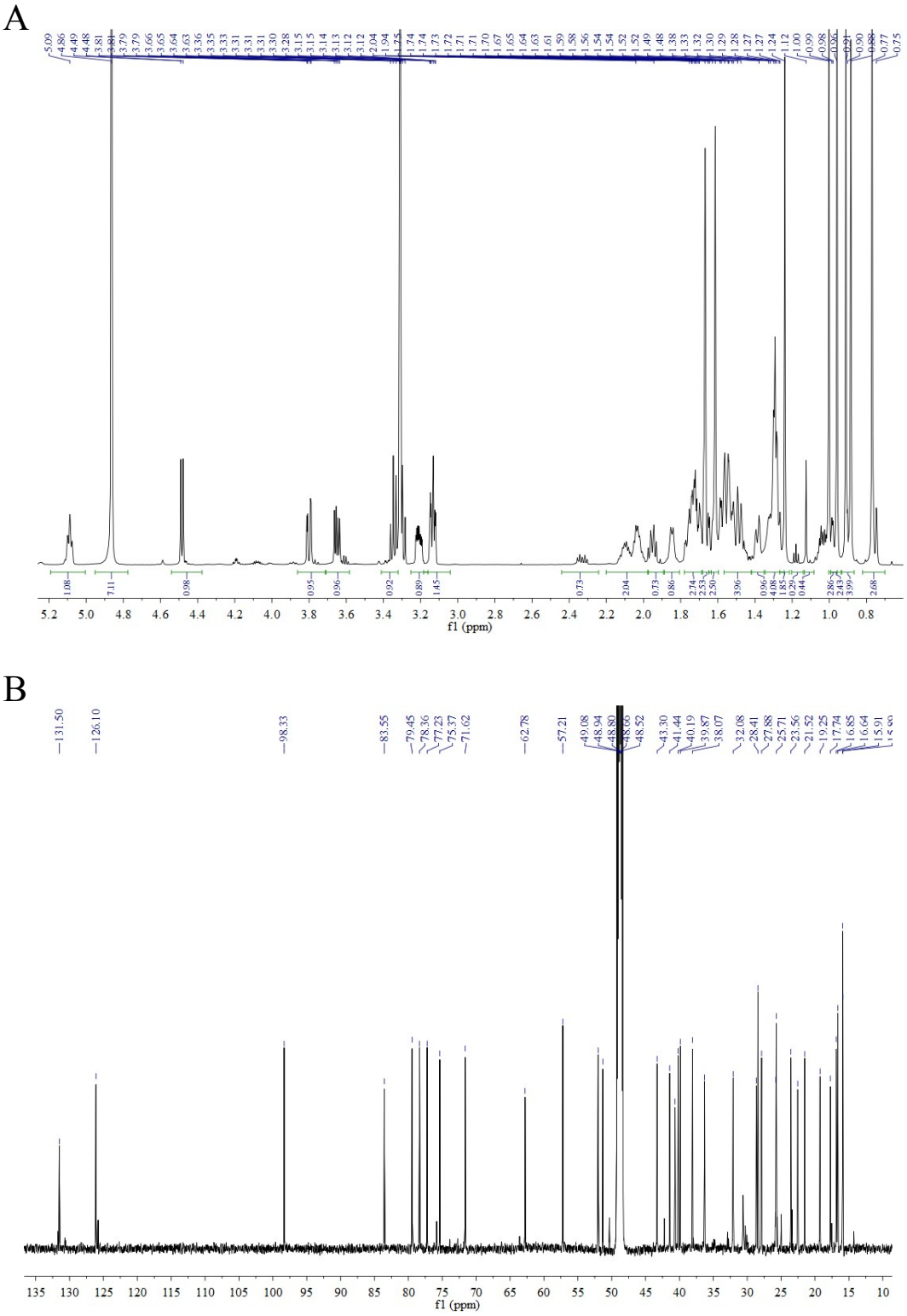
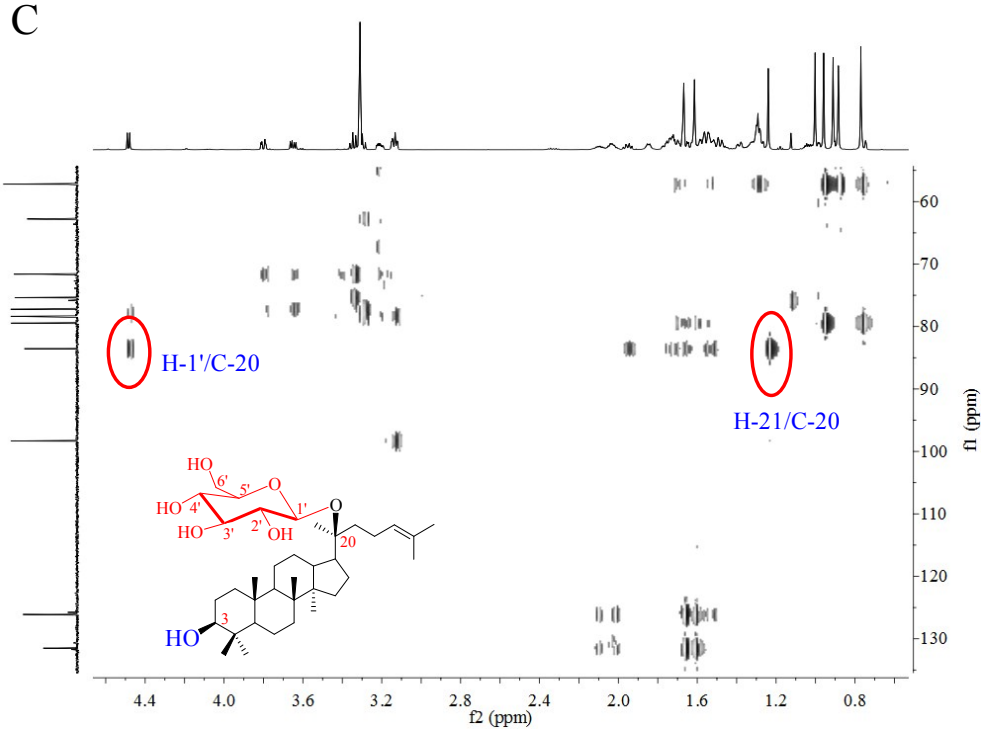

H-1'/C-1'
OH
20
HO
6'
5'
O
HO
4'
1'
O
3
HO
3'
2'
OH
f1 (ppm)
f2 (ppm)

Fig. S6 The ^{1}H NMR (A), ^{13}C NMR (B), HMBC (C) and HSQC (D) spectra of 20*S*-*O*-Glc-DM produced by engineered yeast in methanol-d_4



C


H-1'/C-20
H-21/C-20
HO
6'
5'
O
1'
HO
4'
HO
3'
2'
OH
20
HO
3
f1 (ppm)
f2 (ppm)

D

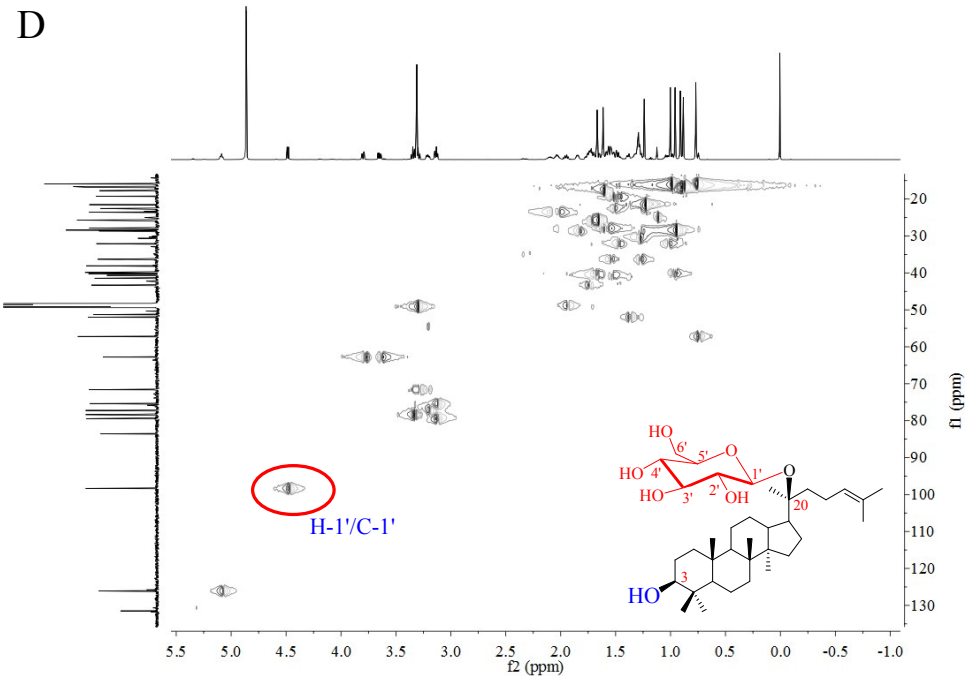
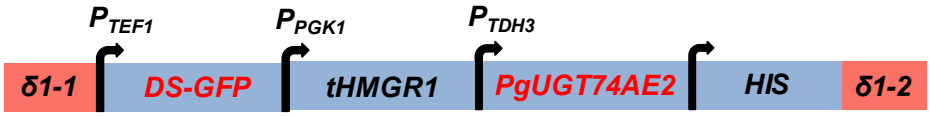

H-1'/C-1'
HO
6'
5'
O
1'
HO
4'
HO
3'
2'
OH
20
HO
3
f1 (ppm)
f2 (ppm)

Fig. S7 Screening of yeast strains Y1C and Y2C

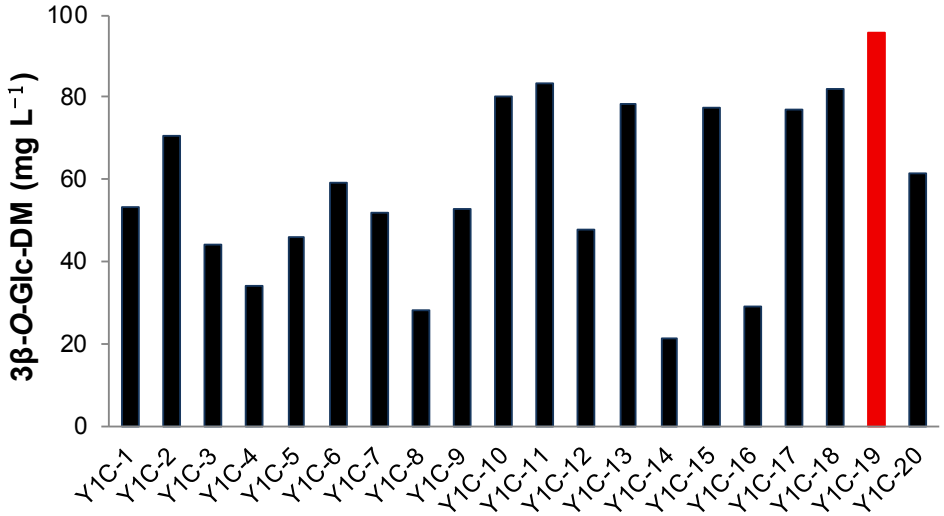
Comparisons of 3β-*O*-Glc-DM and 20*S*-*O*-Glc-DM titers of different yeast transformants after being cultured in YPD for 3 days. (A) Genotype of Y1C. (B) 3β-*O*-Glc-DM titers of the transformants Y1C-1 to Y1C-20. (C) Genotype of Y2C. (D) 20*S*-*O*-Glc-DM titers of the transformants Y2C-1 to Y2C-20. The highest titers are labeled in red.

A Y1C:

Y-△HXK2 + CRISPR/Cas9 + Module Ⅰ + Module Ⅱ

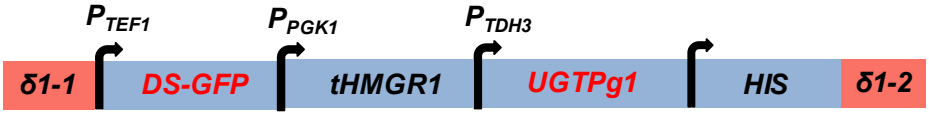


B



C Y2C:

Y-△HXK2 + CRISPR/Cas9 + Module Ⅰ + Module Ⅲ



D

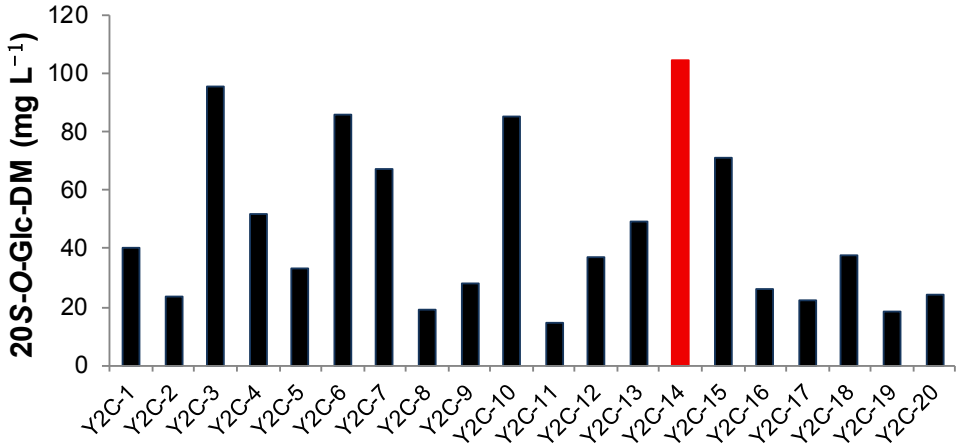


Fig. S8 Screening of yeast strains Y1CS and Y2CS

Comparisons of 3β-*O*-Glc-DM and 20*S*-*O*-Glc-DM titers of different yeast transformants after being cultured in YPD for 3 days. (A) Genotype of Y1CS. (B) 3β-*O*-Glc-DM titers of the transformants Y1CS-1 to Y1CS-20. (C) Genotype of Y2CS. (D) 20*S*-*O*-Glc-DM titers of the transformants Y2CS-1 to Y2CS-20. The highest titers are labeled in red and the titers of the controls Y1C-19 and Y2C-14 (see Fig. S7) are labeled in blue.

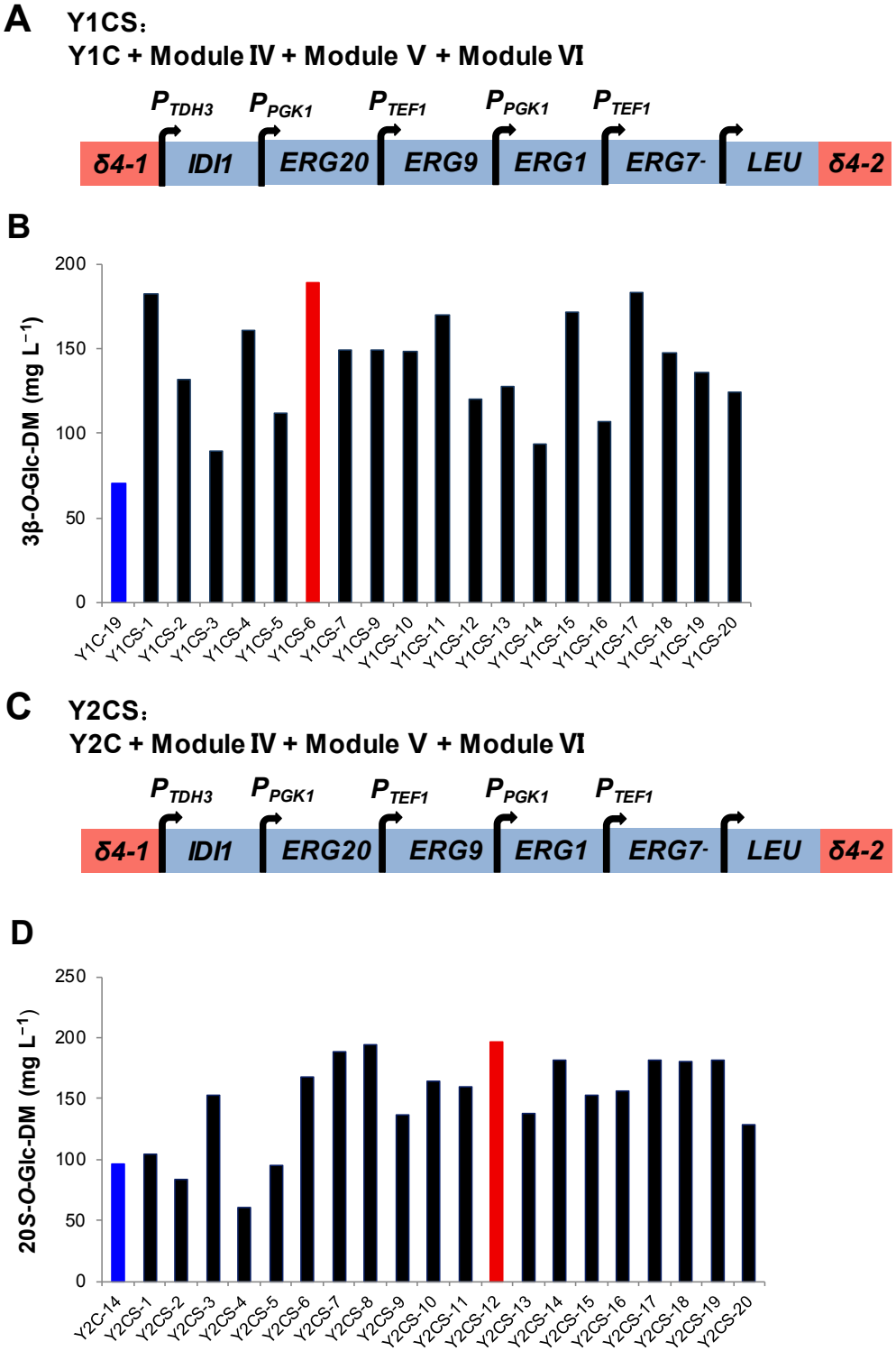
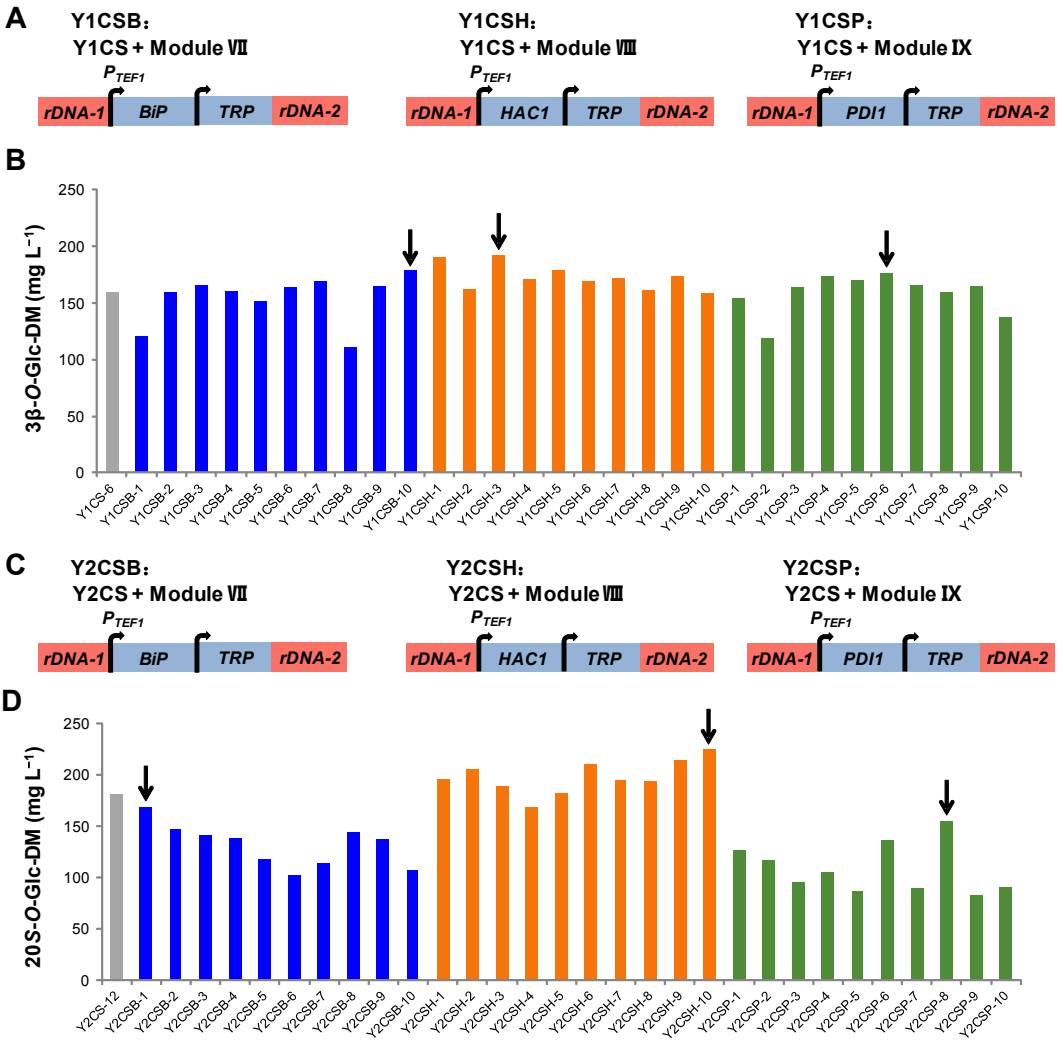


Fig. S9 Screening of yeast strains Y1CSB, Y1CSH, Y1CSP and Y2CSB, Y2CSH, Y2CSP

Comparisons of 3β-*O*-Glc-DM and 20*S*-*O*-Glc-DM titers of different yeast transformants after being cultured in YPD for 3 days. (A) Genotype of Y1CSB, Y1CSH, Y1CSP. (B) 3β-*O*-Glc-DM titers of the transformants Y1CSB-1 to Y1CB-10 are labeled in blue; 3β-*O*-Glc-DM titers of the transformants Y1CSH-1 to Y1CSH-10 are labeled in orange; 3β-*O*-Glc-DM titers of the transformants Y1CSP-1 to Y1CSP-10 are labeled in green and 3β-*O*-Glc-DM titer of the control Y1CS-6 (see Fig. S8) is labeled in gray. (C) Genotype of Y2CSB, Y2CSH, Y2CSP. (D) 20*S*-*O*-Glc-DM titers of the transformants Y2CSB, Y2CSH and Y2CSP. 20*S*-*O*-Glc-DM titers of the transformants Y2CSB-1 to Y2CB-10 are labeled in blue; 20*S*-*O*-Glc-DM titers of the transformants Y2CSH-1 to Y2CSH-10 are labeled in orange; 20*S*-*O*-Glc-DM titers of the transformants Y2CSP-1 to Y2CSP-10 are labeled in green and 20*S*-*O*-Glc-DM titer of the control Y2CS-12 (see Fig. S8) is labeled in gray. The highest titers are pointed out by black arrows.


A
Y1CSB:
Y1CS + Module Ⅶ
Y1CSH:
Y1CS + Module Ⅷ
Y1CSP:
Y1CS + Module Ⅸ
P TEF1
rDNA-1
BiP
TRP
rDNA-2
HAC1
PDI1
B
3β-O-Glc-DM (mg L⁻¹)
250
200
150
100
50
0
Y1CS-6
Y1CSB-1
Y1CSB-2
Y1CSB-3
Y1CSB-4
Y1CSB-5
Y1CSB-6
Y1CSB-7
Y1CSB-8
Y1CSB-9
Y1CSB-10
Y1CSH-1
Y1CSH-2
Y1CSH-3
Y1CSH-4
Y1CSH-5
Y1CSH-6
Y1CSH-7
Y1CSH-8
Y1CSH-9
Y1CSH-10
Y1CSP-1
Y1CSP-2
Y1CSP-3
Y1CSP-4
Y1CSP-5
Y1CSP-6
Y1CSP-7
Y1CSP-8
Y1CSP-9
Y1CSP-10
C
Y2CSB:
Y2CS + Module Ⅶ
Y2CSH:
Y2CS + Module Ⅷ
Y2CSP:
Y2CS + Module Ⅸ
D
20S-O-Glc-DM (mg L⁻¹)
Y2CS-12
Y2CSB-1
Y2CSB-2
Y2CSB-3
Y2CSB-4
Y2CSB-5
Y2CSB-6
Y2CSB-7
Y2CSB-8
Y2CSB-9
Y2CSB-10
Y2CSH-1
Y2CSH-2
Y2CSH-3
Y2CSH-4
Y2CSH-5
Y2CSH-6
Y2CSH-7
Y2CSH-8
Y2CSH-9
Y2CSH-10
Y2CSP-1
Y2CSP-2
Y2CSP-3
Y2CSP-4
Y2CSP-5
Y2CSP-6
Y2CSP-7
Y2CSP-8
Y2CSP-9
Y2CSP-10

3 Supplementary tables

Table S1 Plasmids used in this study

Name	Description	Source
pEASY-Blunt	Cloning vector for blunt ligation, Amp^r, Kan^r	TransGen Biotech
pEASY-Blunt Zero	Cloning vector for blunt ligation, Amp^r, Kan^r	TransGen Biotech
pEASY-Blunt Simple	Cloning vector for blunt ligation, Amp^r, Kan^r	TransGen Biotech
pESC-HIS-DS-GFP	2*μ*, *pGAL1-DS-GFP*, *HIS3*	Liang et al., 2017
pESC-URA	2*μ*, *URA3*	(Invitrogen, USA)
pESC-TRP	2*μ*, *TRP1*	(Invitrogen, USA)
pESC-LEU	2*μ*, *LEU2*	(Invitrogen, USA)
pESC-URA-ERG7	2*μ*, *ERG7* antisense fragment, *URA3*	Wang et al., 2015b
pUC57-PgUGT74AE2	Cloning codon-optimized *PgUGT74AE2* into pUC57	Tianjin University
pUC57-UGTPg1	Cloning codon-optimized *UGTPg1* into pUC57	Tianjin University
pET-UGTPg1	Cloning *UGTPg1* gene into pET-32a	This study
pET-PgUGT74AE2	Cloning *PgUGT74AE2* gene into pET-32a	This study
FM-1	Cas9 expressing plasmid	Zhang et al., 2016
pEASY-HXK2	Cloning *HXK2* gene deletion cassette into pEASY-Blunt Simple	This study
pEASY-SER3	Cloning *SER3* gene deletion cassette into pEASY-Blunt Simple	This study
pEASY-SOR1	Cloning *SOR1* gene deletion cassette into pEASY-Blunt Simple	This study
pEASY-INδ	Cloning *δ1* site into pEASY-Blunt Simple	This study
pEASY-TY4	Cloning *δ4* site into pEASY-Blunt Simple	This study
pEASY-rDNA	Cloning *rDNA* site into pEASY-Blunt Simple	This study
pEASY-DS-GFP	Cloning P_{TEF1}-*synDS-GFP*-T_{CYC1} cassette into pEASY-Blunt Zero	This study
pEASY-tHMGR	Cloning P_{PGK1}-*tHMG1*-T_{ADH1} cassette into pEASY-Blunt Zero	This study
pEASY-UGTPg1	Cloning P_{TDH3}-*synUGTPg1*-T_{ADH2} cassette into pEASY-Blunt Zero	This study
pEASY-PgUGT74AE2	Cloning P_{TDH3}-*synPgUGT74AE2*-T_{ADH2} cassette into pEASY-Blunt Zero	This study
pEASY-BiP	Cloning P_{TEF1}-*BiP*-T_{CYC1} cassette into pEASY-Blunt Zero	This study
pEASY-HAC1	Cloning P_{TEF1}-*HAC1*-T_{CYC1} *c*assette into pEASY-Blunt Zero	This study

pEASY-PDI1	Cloning P_{TEF1}-*PDI1*-T_{CYC1} cassette into pEASY-Blunt Zero	This study
pEASY-IDI1	Cloning P_{TDH3}-*IDI1*-T_{TPI1} cassette into pEASY-Blunt Zero	This study
pEASY-ERG20	Cloning P_{PGK1}-*ERG20*-T_{ADH1} *c*assette into pEASY-Blunt Zero	This study
pEASY-ERG9	Cloning P_{TEF1}-*ERG9*-T_{CYC1} cassette into pEASY-Blunt Zero	This study
pEASY-ERG1	Cloning P_{PGK1}-*ERG1*-T_{ADH1} cassette into pEASY-Blunt Zero	This study
pEASY-ERG7^{-}	Cloning P_{TEF1}-*ERG7*$^{-}$-T_{CYC1} cassette into pEASY-Blunt Zero	This study
pEASY-1	Cloning *overlap*-P_{TDH3}-*synPgUGT74AE2*-T_{ADH2}-*HIS*-*δ1-2* into pEASY-Blunt Zero	This study
pEASY-2	Cloning *overlap*-P_{TDH3}-*synUGTPg1*-T_{ADH2}-*HIS*-*δ1-2* into pEASY-Blunt Zero	This study
pEASY-3	Cloning *δ1-1*-P_{TEF1}-*synDS-GFP*-T_{CYC1}-P_{PGK1}-*tHMG1*-T_{ADH1} into pEASY-Blunt Zero	This study
pEASY-S28	Cloning *δ4-1*-P_{TDH3}-*IDI1*-T_{TPI1}-P_{PGK1}-*ERG20*-T_{ADH1} into pEASY-Blunt	This study
pEASY-S813	Cloning *overlap*-P_{TEF1}-*ERG9*-T_{CYC1}-*overlap* into pEASY-Blunt	This study
pEASY-S1319	Cloning P_{PGK1}-*ERG1*-T_{ADH1}-P_{TEF1}-*ERG7*$^{-}$-T_{CYC1}-*LEU*-*δ4-2* into pEASY-Blunt	This study
p-Cas9-gRNA	Cas9 and gRNA co-expressing plasmid, URA3	This study
p-TRP-rDNA	Scaffold plasmid with *rDNA*, *TRP* marker gene and multiple cloning sites	This study
p-TRP-BiP	Cloning P_{TEF1}-*BiP*-T_{CYC1} cassette into p-TRP-rDNA	This study
p-TRP-HAC1	Cloning P_{TEF1}-*HAC1*-T_{CYC1} cassette into p-TRP-rDNA	This study
p-TRP-PDI1	Cloning P_{TEF1}-*PDI1*-T_{CYC1} cassette into p-TRP-rDNA	This study

Table S2 Primers used for construction of plasmids and genome integration modules

Primer	Sequence（5' to 3'）
UGTPg1-F	CGGGATCCCGATGAAGTCCGAATTAATTTTC
UGTPg1-R	GCGTCGACGTCTTGGCCATAGTTACATAATCTCCTCAAATAATTTG
UGT74AE2-F	CGGGATCCCGATGTTGTCCAAGACTCACATC
UGT74AE2-R	GCGTCGACGTCTTGGCCATAGTTAGGAGGAAACTAGCTTGG
Delta1-2F	GGAAGCTGAAACGTCTAACGGATC
Delta1-TEF1-1R	GGTGTGTGGGGGATCACTGAGAACTTCTAGTATATTCTGTATACCTAATATTATAG
A-TEF1-Delta1-F	CTATAATATTAGGTATACAGAATATACTAGAAGTTCTCAGTGATCCCCCACACACC
TEF1-DS-R	GGGCTACCTTTAGCTTCCACATTTTTGTAATTAAAACTTAGATTAGA
DS-TEF1-F	TCTAATCTAAGTTTTAATTACAAAAATGTGGAAGCTAAAGGTAGCCC
DS-CYC1-R	CCTTCCTTTTCGGTTAGAGCGGATTTACTTGTACAGCTCGTCCATGCCG
CYC1-DS-F	CGGCATGGACGAGCTGTACAAGTAAATCCGCTCTAACCGAAAAGGAAGG
CYC1-PGK1-R	GCAGATGTTATAATATCTGTGCGTCGAGCGTCCCAAAACCTTCTC
PGK1-CYC1-F	GAGAAGGTTTTGGGACGCTCGACGCACAGATATTATAACATCTGC
PGK1-tHMG1-R	CTTCAGTTTTCACCAATTGGTCCATTTTTGTTTTATATTTGTTGTAAAAAGTA
tHMG1-PGK1-F	TACTTTTTACAACAAATATAAAACAAAAATGGACCAATTGGTGAAAACTGAAG
tHMG1-ADH1-R	GTATACACTTATTTTTTTTATAACTCACATGGTGCTGTTGTGCTTC
ADH1-tHMG1-F	GAAGCACAACAGCACCATGTGAGTTATAAAAAAAATAAGTGTATAC
ADH1-R	TCGGCATGCCGGTAGAGGTGTGGTC
B400-F	TACCGCTCCTGGTACCAAC
ADH1-TDH3-R	GACGCTAACATTCAACGCTAGTATTTACCGTTATCTCCCTTATACTTCTC
TDH3-ADH1-F	GAGAAGTATAAGGGAGATAACGGTAAATACTAGCGTTGAATGTTAGCGTC

TDH3-UGTPg1-R	GGCAAGAAAATTAATTCGGACTTCATTTTTGTTTGTTTATGTGTGTTTATTCG
UGTPg1-TDH3-F	CGAATAAACACACATAAACAAACAAAAATGAAGTCCGAATTAATTTTCTTGCC
UGTPg1-ADH2-R	AAGACATAAGAGATCCGCTTACATAATCTCCTCAAATAATTTGGC
ADH2-UGTPg1-F	GCCAAATTATTTGAGGAGATTATGTAAGCGGATCTCTTATGTCTT
ADH2-HIS-R	GTCATCACCGAAACGCGCATTTACCGTTATCTCCCTTATAC
HIS-ADH2-F	GTATAAGGGAGATAACGGTAAATGCGCGTTTCGGTGATGAC
HIS-Delta2-R	TTTGAGGATTCCTATATCCTCGAGGTGTCACTACATAAGAACACCT
Delta2-HIS-2F	AGGTGTTCTTATGTAGTGACACCTCGAGGATATAGGAATCCTCAAA
Delta2-1R	TGTTGGAATAGAAATCAACTATCATC
UGT74AE2-TDH3-F	CGAATAAACACACATAAACAAACAAAAATGTTGTCCAAGACTCACATCATG
UGT74AE2-ADH2-R	CAAAACAAATAAAACATCATCACAAAAATGGCAGCAGCAATGGTGTTG
TDH3-UGT74AE2-R	CATGATGTGAGTCTTGGACAACATTTTTGTTTGTTTATGTGTGTTTATTCG
ADH2-UGT74AE2-F	CCAAGCTAGTTTCCTCCTAAGCGGATCTCTTATGTCTT
TY4-F1	TGTTGGAACGAGAGTAAT
TY4-1-TDH3-R	CTAACATTCAACGCTAGTATTATGTACCTACCATAACATG
TDH3-TY4-1-F	CATGTTATGGTAGGTACATAATACTAGCGTTGAATGTTAG
TDH3-IDI1-R	CTATTGTTGTCGGCAGTCATTTTGTTTGTTTATGTGTG
IDI1-TDH3-F	CACACATAAACAAACAAAATGACTGCCGACAACAATAG
IDI1-TPI1-R	TTTTATATAATTATATTAATCTTATAGCATTCTATGAATTTG
TPI1-IDI1-F	CAAATTCATAGAATGCTATAAGATTAATATAATTATATAAAA
TPI1-PGK1-R	GATGTTATAATATCTGTGCGTCTATATAACAGTTGAAATTTG
PGK1-TPI1-F	CAAATTTCAACTGTTATATAGACGCACAGATATTATAACATC
PGK1-ERG20-R	CCTAATTTCTTTTTCTGAAGCCATTGTTTTATATTTGTTGTAA
ERG20-PGK1-F	TTACAACAAATATAAAACAATGGCTTCAGAAAAAGAAATTAGG

ERG20-ADH1-R	ACTTATTTTTTTTATAACTCTATTTGCTTCTCTTGTAAAC
ADH1-ERG20-F	GTTTACAAGAGAAGCAAATAGAGTTATAAAAAAAATAAGT
ADH1-TEF1-R	GGTGTGTGGGGGATCACTTCGGCATGCCGGTAGAGGTG
ERG1-PGK1-F	TTACAACAAATATAAAACAATGTCTGCTGTTAACGTTGC
ERG1-ADH1-R	CTTATTTTTTTTATAACTTTAACCAATCAACTCACCAAAC
PGK1-CYC1-F	GAAGGTTTTGGGACGCTCGACGCACAGATATTATAAC
PGK1-ERG1-R	GCAACGTTAACAGCAGACATTGTTTTATATTTGTTGTAA
ADH1-ERG1-F	GTTTGGTGAGTTGATTGGTTAAAGTTATAAAAAAAATAAG
ERG7-TEF1-F	CTAAGTTTTAATTACAAATCCAAAAAAAAGTAAGAATTTTTG
ERG7-CYC1-R	CCTTTTCGGTTAGAGCGGATCCTCTGGCGAAGAATTGTTA
TEF1-ADH1-F	CACCTCTACCGGCATGCCGAAGTGATCCCCCACACACC
TEF1-ERG7-R	CAAAAATTCTTACTTTTTTTTGGATTTGTAATTAAAACTTAG
CYC1-ERG7-F	TAACAATTCTTCGCCAGAGGATCCGCTCTAACCGAAAAGG
CYC1-LEU-R	CCTGACCATTTGATGGAGCGAGCGTCCCAAAACCTTCTC
TY4-R2	TGTTGATAATTAGAGGTT
TY4-2-LEU-F	GCGTAAGGAGAAAATACCTATGAGGAATATGAGTCGTC
LEU-CYC1-F	GAGAAGGTTTTGGGACGCTCGCTCCATCAAATGGTCAGG
LEU-TY4-2-R	GACGACTCATATTCCTCATAGGTATTTTCTCCTTACGC
S28-400F	GACTACTTAGACTGCTTCGG
TEF1-ERG9-R	CAATTGTAATAGCTTTCCCATTTTGTAATTAAAACTTAGATTAGA
PGK1-CYC1-F	GAAGGTTTTGGGACGCTCGACGCACAGATATTATAAC
S1319-400R	GTGAGAACAGGCTGTTGTTG
TEF1-ERG9-R	CAATTGTAATAGCTTTCCCATTTTGTAATTAAAACTTAGATTAGA
ERG9-TEF1-F	TCTAATCTAAGTTTTAATTACAAAATGGGAAAGCTATTACAATTG

ERG9-CYC1-R	CTTTTCGGTTAGAGCGGATTCACGCTCTGTGTAAAGTG
CYC1-ERG9-F	CACTTTACACAGAGCGTGAATCCGCTCTAACCGAAAAG
CYC1-PGK1-R	GTTATAATATCTGTGCGTCGAGCGTCCCAAAACCTTC
rDNA1-MQWD-F	GCGGATCCCGGCGGCCGCACGAGCTCGTCCCGCGGGGAATGAGAGTAGCAAACGTAAG
rDNA1-MQWD-R	CCCTCGAGGGAAAAGTACTTTTCGTCGACGCGTTGTCGCTCTGCCTTAACTACG
GJ-rDNA1-2μ-F	CACGAGGCCCTTTCGTCGCGGATCCCGGCGGCCGC
GJ-rDNA2-PUC-R	CACATGTTCTTTCCTGCGTCCCCGCGGGACGAGCTCGTG
PUC-GJ-rDNA2-F	CACGAGCTCGTCCCGCGGGGACGCAGGAAAGAACATGTG
2μ-GJ-rDNA1-R	GCGGCCGCCGGGATCCGCGACGAAAGGGCCTCGTG
rDNA2-TRP-F	GAATAGACCGAGATAGGGTTGGAGAGGGCAAAAGAAAATAAAAG
rDNA2-MQWD-R	TCCCCGCGGGACGAGCTCGTGCGGCCGCCGGGATCCGCTTTCCTCTAATCAGGTTCCAC
TRP-MQWD-F	TTAAGGCAGAGCGACAACGCGTCGACGAAAAGTACTTTTCCCTCGAGGGTCGCGCGTTTCGGTGATGACGGTG
TRP-rDNA2-R	CTTTTATTTTCTTTTGCCCTCTCCAACCCTATCTCGGTCTATTC
GJ-F	GGCAGAGCGACAACGCGTCGACAGTGATCCCCCACACACCATAG
GJ-R	TAATATCTGTGCGTCCCTCGAGCGAGCGTCCCAAAACCTTCTC
TEF1-BIP-R	CGCTTAGTCTGTTGAAAAACATTTTTGTAATTAAAACTTAGATTAGA
BIP-TEF1-F	TCTAATCTAAGTTTTAATTACAAAAATGTTTTTCAACAGACTAAGCG
BIP-CYC1-R	TTCCTTTTCGGTTAGAGCGGATCTACAATTCGTCGTGTTCGAAATAATC
CYC1-BIP-F	GATTATTTCGAACACGACGAATTGTAGATCCGCTCTAACCGAAAAGGAA
TEF1-HAC1-R	CAAAATCAGTCATTTCCATTTTTGTAATTAAAACTTAGATTAGA
HAC1-TEF1-F	TCTAATCTAAGTTTTAATTACAAAAATGGAAATGACTGATTTTG
HAC1-CYC1-R	TTCCTTTTCGGTTAGAGCGGATTCATGAAGTGATGAAGAAATC
CYC1-HAC1-F	GATTTCTTCATCACTTCATGAATCCGCTCTAACCGAAAAGGAA
TEF1-PDI1-R	GCACCAGCAGAAAACTTCATTTTTGTAATTAAAACTTAGATTAGA

PDI1-TEF1-F	TCTAATCTAAGTTTTAATTACAAAAATGAAGTTTTCTGCTGGTGC
PDI1-CYC1-R	TTCCTTTTCGGTTAGAGCGGATTTACAATTCATCGTGAATGGC
CYC1-PDI1-F	GCCATTCACGATGAATTGTAAATCCGCTCTAACCGAAAAGGAA
TEF1-SUP4t-MSC-F	GTGCTTTTTTTGTTTTTTATGTCTGTCGACAAGCTTAGTGATCCCCCACACAC CATAG
TEF1-Cas9-R	CGTCGTGGTCCTTATAGTCCATTTTTGTAATTAAAACTTAGATTAG
Cas9-TEF1-F	CTAATCTAAGTTTTAATTACAAAAATGGACTATAAGGACCACGACG
Cas9-ADH2-R	CTATAAATCGTAAAGACATAAGAGATCCGCTTACTTTTTCTTTTTTGCCTGG
ADH2- Cas9-F	CCAGGCAAAAAAGAAAAAGTAAGCGGATCTCTTATGTCTTTACGATTTATA G
ADH2-pESC-R	CGTTGGCCGATTCATTAATGCAGCTGTTACCGTTATCTCCCTTATAC
pESC-ADH2-F	GTATAAGGGAGATAACGGTAACAGCTGCATTAATGAATCGGCCAACG
pESC-SNR52P-MCS-R	CATACATTATCTTTTCAAAGAGCGGCCGCGGATCCCAGCTGGCGTAATAGCG AAGAGG
SNR52p-MSC-pESC-F	CCTCTTCGCTATTACGCCAGCTGGGATCCGCGGCCGCTCTTTGAAAAGATAA TGTATG
SUP4t-MSC-TEF1-R	CTATGGTGTGTGGGGGATCACTAAGCTTGTCGACAGACATAAAAAACAAAA AAAGCAC

Table S3 Primers used for construction of gene deletion cassettes

Primer	Sequence（5' to 3'）
HXK2-1F	CCAAAAAAACCACAAGCCAGAAAGG
HXK2-1R	ATATTAAGGGTTGTCGACCTGCCAAAGAGTCGGCAATAAATTCCCAC
HXK2-2F	CAGATCCACTAGTGGCCTATGCCCCAGCCAGAATCGAGGAAG
HXK2-2R	AATAACAGCGGCACCAGCAC
SER3-1F	AATCTTTCATGAATACCGTTCCACAGC
SER3-1R	ATATTAAGGGTTGTCGACCTGCGGAGAAAGGCGAGTTGAAAACAGCA
SER3-2F	CAGATCCACTAGTGGCCTATGCACATTCCATCTTTGATCCAAGCCGT
SER3-2R	CGGTCTTCAAAACACCTGGTACATT
SOR1-1F	TCGAGCAAAGACCAATCCCTACCAT
SOR1-1R	ATATTAAGGGTTGTCGACCTGCCGACACAAGCGCCCTCTTCATAACT
SOR1-2F	CAGATCCACTAGTGGCCTATGCGCTACAGAGAGCAAAAGATTTCGGA
SOR1-2R	ACAGCGTCACGATAATCACCGAATG
Kan-F	CAGCTGAAGCTTCGTACGC
Kan-R	GCATAGGCCACTAGTGGATCTG

Table S4 Primers used for qPCR

Primer	Sequence（5′→3′）
DS-qF	GGGCAGAAGATCCCAATGGT
DS-qR	TCAGAGAAAGTCCACGCACC
actin-qF	ACCCTGTTCTTTTGACGG
actin-qR	GCAGTGGTGGAGAAAGAG

Table S5 The HPLC conditions used in this study

Method	Gradient	Product
1	0 min, 32%; 10 min, 32%; 20 min, 85%; 25 min, 100%; 26 min, 32%; 35 min, 32% ACN-H_2O	3β-*O*-Glc-DM 20*S*-*O*-Glc-DM (products of enzyme reactions)
2	0-40 min, 58% ACN-H_2O	3β-*O*-Glc-DM 20*S*-*O*-Glc-DM (extracts of yeast fermentations)

Table S6 ^{1}H-NMR (600 MHz) and ^{13}C-NMR (150 MHz) spectral data of 3β-*O*-Glc-DM produced by *in vitro* reaction in methanol-d_4

C	δ_C	δ_H (mult, *J* in Hz)
1	40.19	1.67 (1H, m); 0.96 (1H, m)
2	27.03	1.93 (1H, m); 1.73 (1H, m)
3	90.58	3.18 (1H, m)
4	41.54	
5	57.49	0.78 (1H, m)
6	19.07	1.44 (2H, m)
7	36.28	1.56 (1H, m); 1.28 (1H, m)
8	37.82	
9	51.94	1.38 (1H, m)
10	40.13	
11	32.11	1.51 (1H, m); 1.07 (1H, m)
12	25.69	1.53 (2H, m)
13	43.33	1.73 (1H, m)
14	51.30	
15	28.61	1.86 (1H, m); 0.96 (1H, s)
16	25.57	1.71 (1H, m); 1.54 (1H, m)
17	50.33	1.74 (1H, m)
18	16.80	0.91 (3H, s)
19	16.67	0.84 (3H, s)
20	75.82	
21	22.51	1.12 (3H, s)
22	42.22	1.43 (2H, m)
23	23.38	2.03 (2H, m)
24	125.77	5.10 (1H, t, *J* = 7.2 Hz)
25	131.72	
26	25.00	1.68 (3H, s)
27	17.53	1.62 (3H, s)
28	28.21	1.04 (3H, s)
29	16.60	0.85 (3H, s)
30	15.84	1.00 (3H, s)
3-*O*-Glc-1′	106.52	4.32 (1H, d, *J* = 7.8 Hz)
2′	76.02	3.18 (1H, m)
3′	78.09	3.34 (1H, m)
4′	71.47	3.24 (1H, m)
5′	77.48	3.17 (1H, m)
6′	62.61	3.84 (1H, dd, *J* = 12.0, 2.4 Hz); 3.66 (1H, br d, *J* = 12.0 Hz)

Table S7 ^{1}H-NMR (600 MHz) and ^{13}C-NMR (150 MHz) spectral data of 20*S*-*O*-Glc-DM produced by *in vitro* reaction in methanol-d_4

C	δ_C	δ_H (mult, *J* in Hz)
1	40.19	1.69 (1H, m); 0.91 (1H, m)
2	27.88	1.53 (2H, m)
3	79.45	3.13 (1H, m)
4	40.65	
5	57.21	0.74 (1H, m)
6	19.25	1.38 (2H, m)
7	36.30	1.54 (1H, m); 1.28 (1H, m)
8	41.44	
9	51.99	1.37 (1H, m)
10	39.87	
11	22.56	1.48 (2H,m)
12	25.77	1.73 (2H, m)
13	43.30	1.75 (1H, m)
14	51.30	
15	32.08	1.47 (1H, m); 1.01 (1H, m)
16	28.65	1.83 (1H, m); 1.58 (1H, m)
17	49.23	1.95 (1H, m)
18	16.64	0.88 (3H, s)
19	15.91	0.77 (3H, s)
20	83.55	
21	21.52	1.24 (3H, s)
22	40.65	1.49 (2H, m)
23	23.56	2.10 (2H, m)
24	126.10	5.08 (1H, t, *J* = 7.2 Hz)
25	131.50	
26	25.71	1.66 (3H, s)
27	17.74	1.61 (3H, s)
28	28.41	0.95 (3H, s)
29	15.89	1.00 (3H, s)
30	16.85	0.91 (3H, s)
20-*O*-Glc-1′	98.33	4.48 (1H, d, *J* = 7.8 Hz)
2′	75.37	3.13 (1H, m)
3′	78.36	3.34 (1H, m)
4′	71.62	3.28 (1H, m)
5′	77.23	3.20 (1H, m)
6′	62.78	3.80 (1H, dd, *J* = 12.0, 2.4 Hz); 3.64 (1H, br d, *J* = 12.0 Hz)

Table S8 ^{1}H-NMR (600 MHz) and ^{13}C-NMR (150 MHz) spectral data of 3β-*O*-Glc-DM produced by engineered yeast in methanol-d_4

C	δ_C	δ_H (mult, *J* in Hz)
1	40.19	1.67 (1H, s); 0.96 (1H, m)
2	27.03	1.91 (1H, m); 1.67 (1H, m)
3	90.58	3.18 (1H, m)
4	41.54	
5	57.49	0.78 (1H, m)
6	19.07	1.49 (2H, m)
7	36.28	1.57 (1H, m); 1.28 (1H, m)
8	37.82	
9	51.94	1.38 (1H, m)
10	40.06	
11	32.11	1.47 (1H, m); 1.05 (1H, m)
12	25.57	1.51 (2H, m)
13	43.33	1.73 (1H, m)
14	51.30	
15	28.61	1.84 (1H, m); 0.96 (1H, s)
16	25.70	1.68 (1H, m); 1.47 (1H, m)
17	50.33	1.74 (1H, m)
18	16.67	0.90 (3H, s)
19	16.80	0.84 (3H, s)
20	75.82	
21	23.39	1.68 (3H, s)
22	42.22	1.46 (2H, m)
23	22.51	2.03 (2H, m)
24	125.77	5.10 (1H, t, *J* = 7.2 Hz)
25	131.72	
26	25.00	1.12 (3H, s)
27	17.53	1.62 (3H, s)
28	28.21	1.04 (3H, s)
29	16.60	0.91 (3H, s)
30	15.84	1.00 (3H, s)
3-*O*-Glc-1′	106.52	4.32 (1H, d, *J* = 7.8 Hz)
2′	76.00	3.18 (1H, m)
3′	78.09	3.34 (1H, m)
4′	71.47	3.25 (1H, m)
5′	77.48	3.17 (1H, m)
6′	62.61	3.84 (1H, dd, *J* = 12.0, 2.4 Hz); 3.65 (1H, br d, *J* = 12.0 Hz)

Table S9 ^{1}H-NMR (600 MHz) and ^{13}C-NMR (150 MHz) spectral data of 20*S*-*O*-Glc-DM produced by engineered yeast in methanol-d_4

C	δ_C	δ_H (mult, *J* in Hz)
1	40.19	1.70 (1H, m); 0.91 (1H, m)
2	27.88	1.53 (2H, m)
3	79.45	3.12 (1H, m)
4	40.65	
5	57.21	0.75 (1H, m)
6	19.25	1.45 (2H, m)
7	36.30	1.54 (1H, m); 1.28 (1H, m)
8	41.44	
9	51.99	1.38 (1H, m)
10	39.87	
11	22.56	1.53 (2H, m)
12	25.77	1.73 (2H, m)
13	43.30	1.74 (1H, m)
14	51.30	
15	32.08	1.48 (1H, m); 1.01 (1H, m)
16	28.65	1.83 (1H, m); 1.26 (1H, m)
17	49.08	1.94 (1H, m)
18	16.64	0.88 (3H, s)
19	15.91	0.77 (3H, s)
20	83.55	
21	21.52	1.24 (3H, s)
22	40.65	1.49 (2H, m)
23	23.56	2.10 (2H, m)
24	126.10	5.09 (1H, t, *J* = 7.2 Hz)
25	131.50	
26	25.71	1.68 (3H, s)
27	17.74	1.61 (3H, s)
28	28.41	0.96 (3H, s)
29	15.89	1.00 (3H, s)
30	16.85	0.91 (3H, s)
20-*O*-Glc-1′	98.33	4.47 (1H, d, *J* = 7.8 Hz)
2′	75.37	3.13 (1H, m)
3′	78.36	3.33 (1H, m)
4′	71.62	3.28 (1H, m)
5′	77.23	3.20 (1H, m)
6′	62.78	3.78 (1H, dd, *J* = 12.0, 2.4 Hz); 3.63 (1H, br d, *J* = 12.0 Hz)