Electronic Supplementary Material (ESI) for Green Chemistry.
This journal is © The Royal Society of Chemistry 2020

Supplementary Information

Biological upgrading of 3,6-anhydro-L-galactose from agarose to a new platform chemical

Dong Hyun Kim,[a] Jing-Jing Liu,[b] Jae Won Lee,[bc] Jeffrey G. Pelton,[d] Eun Ju Yun,[a] Sora Yu,[a] Yong-Su Jin,*[bc] and Kyoung Heon Kim *[a]

[a]Department of Biotechnology, Graduate School, Korea University, Seoul 02841, South Korea

[b]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[c]Department of Food Science and Human Nutrition, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[d]QB3 Institute, University of California, Berkeley, CA, USA

*Correspondence should be addressed to K.H.K. (khekim@korea.ac.kr) or Y.-S.J. (ysjin@illinois.edu)

Supplementary Materials and Methods

NMR analysis of AHGol and ABol. A gradient COSY experiment of AHGol was recorded using the pulse sequence cosygpprqf. A total of 2048 and 256 complex points were collected in the direct and indirect dimensions, respectively, with spectral widths of 11.1 ppm, a carrier frequency of 4.79 ppm, and a recycle delay of 2 sec. Pre-saturation during the recycle delay was used to reduce the intensity of the residual water signal. A 2D H2BC experiment of AHGol was recorded using the pulse sequence h2bcetgpl3. A total of 1024 and 64 complex points were recorded in the ^{1}H and ^{13}C dimensions, respectively, with spectral widths of 8 ppm for ^{1}H and 70 ppm for ^{13}C, using carrier frequencies of 4.78 ppm for ^{1}H and 70 ppm for ^{13}C, and a recycle delay for 1.5 s. 1D ^{1}H, 2D COSY, and 2D ^{1}H-^{13}C edited HSQC spectra of ABol were recorded in the same manner as the corresponding experiments for AHGol, except that the 2D COSY was acquired without pre-saturation of the water resonance using the pulse sequence cosygpprqf. To obtain ^{13}C-^{13}C correlations, a 1,1 Adequate spectrum was recorded using the pulse sequence adeq11etgprdsp. A total of 1024 and 128 complex points were recorded in the ^{1}H and ^{13}C dimensions, respectively, with spectral widths of 11 ppm (^{1}H) and 60 ppm (^{13}C), using carrier frequencies of 4.69 ppm (^{1}H) and 82 ppm (^{13}C), and a recycle delay for 2 s.

GRE3 **expression of D452-L124 with AB.** Strain D452-L124 was grown in YPD, harvested, and transferred to synthetic complete medium containing 50 g/L glucose and 10 g/L AB, respectively. Total RNA from yeast cells was prepared using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. Each cDNA was prepared from each extracted RNA using the SuperScript II Reverse

Transcriptase kit (Thermo Fisher Scientific) according to the provided instructions. To investigate expression of *GRE3*, PCR was performed using cDNA with the following primer pairs c*GRE3*-F and c*GRE*-R. *ACT1* was used as a house keeping gene, and it was also confirmed using PCR with the following primer pairs c*ACT1*-F and c*ACT1*-R. Same volume of PCR products were loaded on a DNA gel.

Table S1 Primers, plasmids, and strains used in this study

Primer	Primer sequence	Source
GRE3-F	5'-CATATGTCTTCACTGGTTACTCTTAATAACGGT-3'	This study
GRE3-R	5'-GCGGCCGCGGCAAAAGTGGGGAATTTACCATCCAA-3'	This study
LAC12-F	5'-tctagagcggccgcactagtgccaccatggcaga tcattcgagcag-3'	This study
LAC12-R	5'-tctagagcggccgcgtcgacttaaacagattctg cctctg-3'	This study
LAC4-F	5'-tctagagcggccgcactagtgccaccatgtcttg ccttattcctgagaat-3'	This study
LAC4-R	5'-tctagagcggccgcgtcgacttattcaaaagcga gatcaaactc-3'	This study
gCS8-U	TGATTCAATCATTCTTATTGgttttagagctagaaatagcaag	Liu, *et al.*[1]
gCS8-D	CAATAAGAATGATTGAATCAgatcatttatctttcactgcgga	Liu, *et al.*[1]
CS8-IU	caaaattacctacggtaattagtgaaaggccaaaatctaatgttacaataAATTAACCCTCACTAAAGGGA	Liu, *et al.*[1]
CS8-ID	gaccgttcccttgtgttgtaccagtggtagggttcttctcggtagcttctGTAATACGACTCACTATAGGGC	Liu, *et al.*[1]
CS8-CKU	agtggaacatagaagggg	Liu, *et al.*[1]
CS8-CKD	Taagcagcccagtgaac	Liu, *et al.*[1]
gCS6-U	GATACTTATCATTAAGAAAAgttttagagctagaaatagcaag	This study
gCS6-D	TTTTCTTAATGATAAGTATCgatcatttatctttcactgcgga	This study
CS6-IU	aacctcgaggagaagttttttacccctctccacagatcCAGGAAACAGCTATGACCATG	This study
CS6-ID	taattaggtagaccgggtagatttttccgtaaccttggtgtcTGTAAAACGACGGCCAGT	This study
CS6-CKU	gtctgccgaaattctgtg	This study
CS6-CKD	cggtcagaaagggaaatg	This study
c*GRE3*-F	5'- TCCTCCTTCGGTCCTCAATC-3'	This study
c*GRE3*-R	5'-CCTTCAATTCTTGCTCCGTT-3'	This study
c*ACT1*-F	5'- AGGAATTATACGGTAACATCGTTATGTC-3'	This study
c*ACT1*-R	5'-TTGTGGTGAACGATAGATGGA-3'	This study

Plasmid	Description	Source
p423-pGPD	pSR423-pTDH3-tCYC1	Mumberg[2]
p425-pGPD	pSR425-pTDH3-tCYC1	Mumberg[2]
pRS423-	pRS423-pGPD harboring *LAC12* gene from *K. lactis* Y-8279	This study

pGPD-*LAC12*		
pRS425-pGPD-*LAC4*	pRS425-pGPD harboring *LAC4* gene from *K. lactis* Y-8279	This study
Cas9-NAT	p414-TEF1p-Cas9-CYC1t-NAT1	Zhang *et al.*[3]
p42K-gCS8	pRS42K carrying guide RNA for integration at CS8 locus	Liu*, et al.*[1]
p42H-gCS6	pRS42H carrying guide RNA for integration at CS6 locus	This study

Strain	**Description**	
D452-2	*MAT*α *leu2 ura3 his3 can1*	Hosaka *et al.*[4]
D452-L12	D452-2 with CS8-*LAC12* integration	This study
D452-L124	D452-2 with CS8-*LAC12, CS6-LAC4* integration	This study

Supplementary References

1. Liu, J.J. et al. Metabolic engineering of probiotic *Saccharomyces boulardii*. *Appl. Environ. Microbiol.* **82**, 2280-2287 (2016).
2. Mumberg, D., Müller, R. & Funk, M. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* **156**, 119-122 (1995).
3. Zhang, G.C. et al. Construction of a quadruple auxotrophic mutant of an industrial polyploidy *Saccharomyces cerevisiae* using RNA-guided Cas9 nuclease. *Appl. Environ. Microbiol.* 02310-02314 (2014).
4. Hosaka, K., Nikawa, J.I., Kodaki, T. & Yamashita, S. A dominant mutation that alters the regulation of INO1 expression in *Saccharomyces cerevisiae*. *J. Biochem.* **111**, 352-358 (1992).

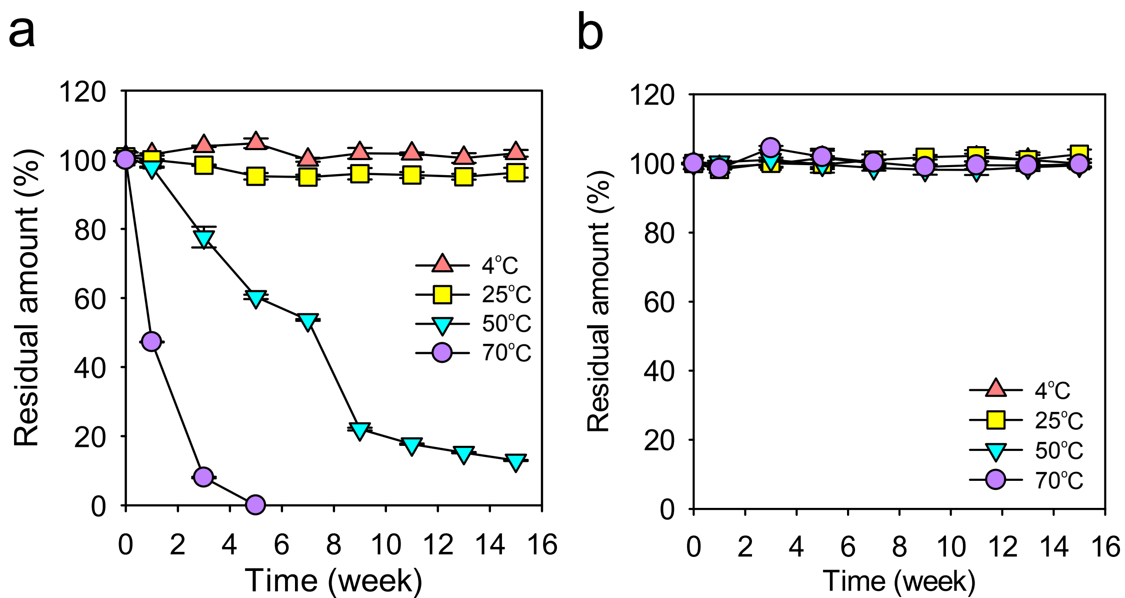

Fig. S1 Thermal stability of AHG and AHGol. a AHG and **b** AHGol were incubated at various temperatures ranging from 4 to 70 °C for 15 weeks. Values and error bars represent means and standard deviations of duplicate experimental data.

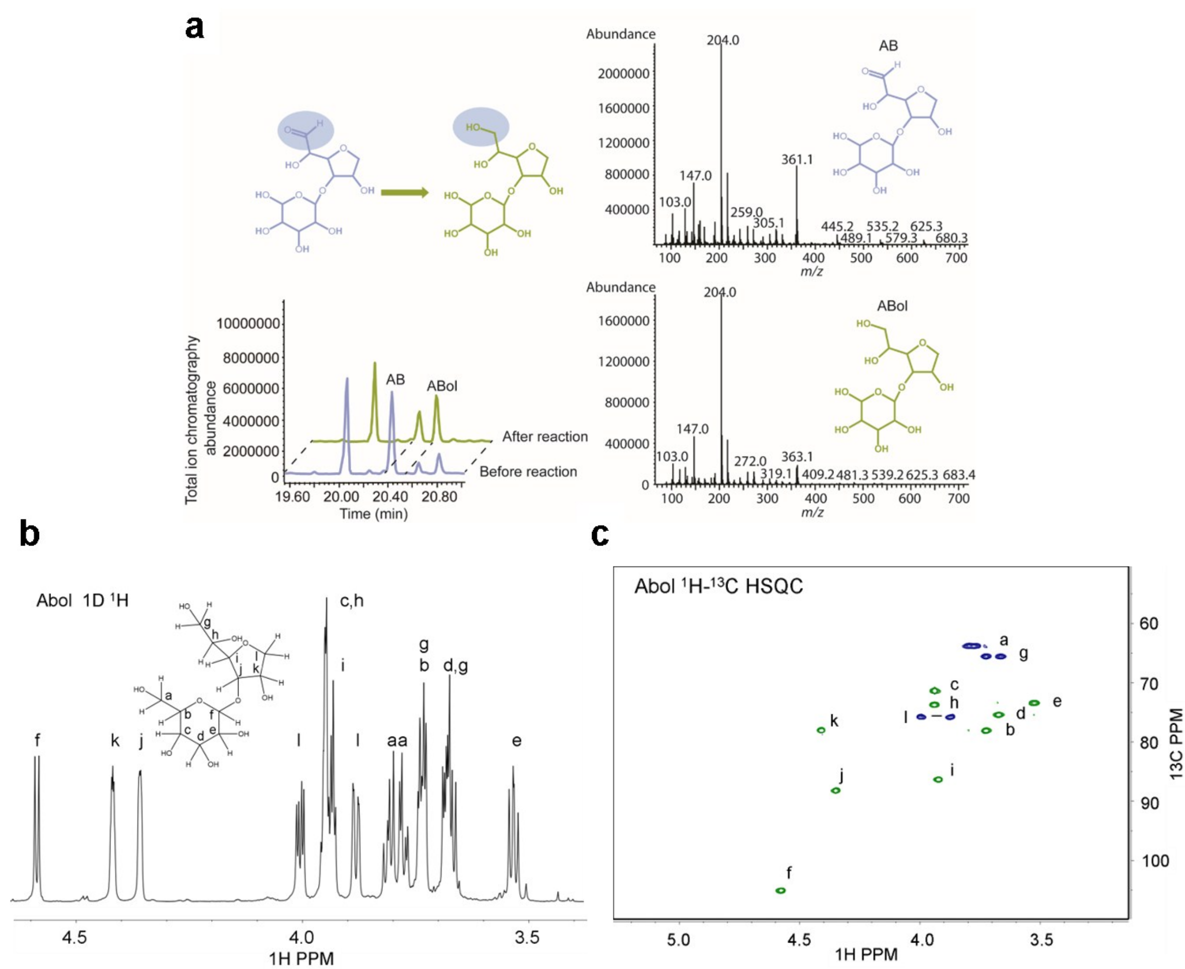

Fig. S2 *In vitro* enzymatic conversion of AB to ABol with AR. a Total ion chromatograms and spectra of AB and ABol obtained for GC/MS analysis of AR enzymatic reaction against AB. **b**–**c** NMR spectra of ABol produced from AB by AR. **b** 1D H NMR spectrum. **c** ^{1}H-^{13}C HSQC NMR spectrum.

Table S2 ^{1}H-and ^{13}C chemical shifts from the NMR spectroscopy of AHGol produced by AR

	^{1}H chemical shift (ppm)	^{13}C chemical shift (ppm)
1	3.63 / 3.70	65.6
2	3.83	74.1
3	3.73	87.4
4	4.13	81.0
5	4.24	79.5
6	3.99 / 3.83	75.4

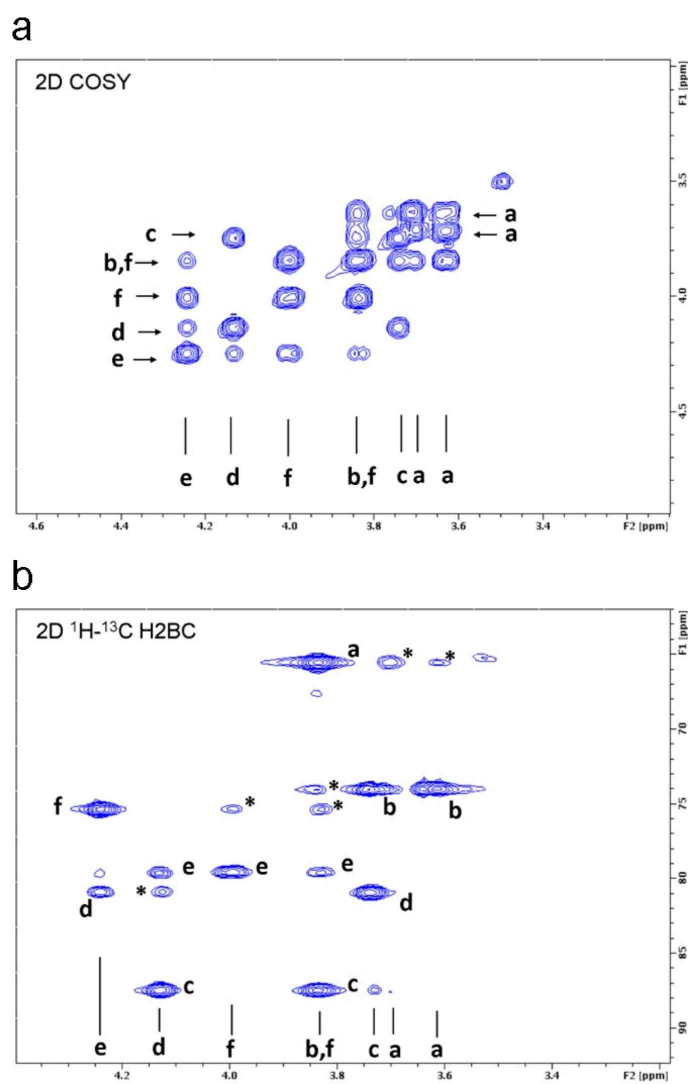

Fig. S3 NMR analysis of AHGol produced by AR. a 2D proton gradient COSY spectrum of AHGol. **b** 2D ^{1}H-^{13}C H2BC spectrum of AHGol showing ^{13}C-^{13}C neighbors for the four CH and two CH_2 groups.

Table S3 ^{1}H-and ^{13}C chemical shifts from the NMR spectroscopy of ABol produced by AR

	^{1}H chemical shift (ppm)	^{13}C chemical shift (ppm)
1	3.80 / 3.77	63.8
2	3.72	78.0
3	3.94	71.3
4	3.67	75.3
5	3.52	73.4
6	4.58	105.1
7	3.72 / 3.66	65.6
8	3.94	73.7
9	3.92	86.3
10	4.35	88.2
11	4.41	78.0
12	3.99 / 3.87	75.8

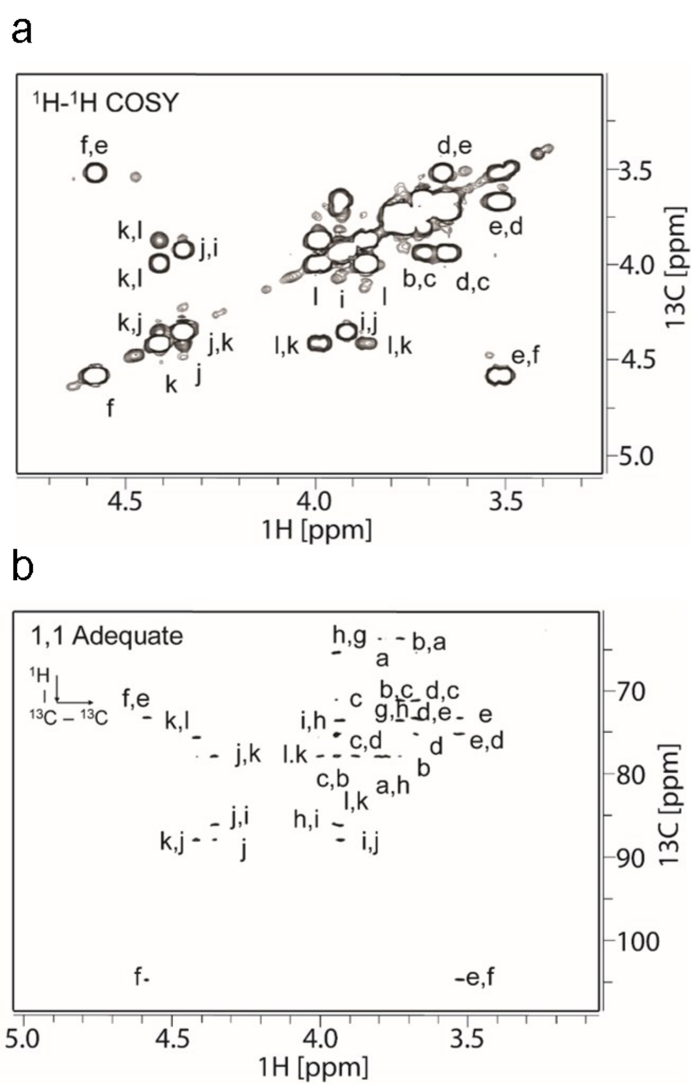

Fig. S4 NMR analysis of ABol produced by AR. a ^{1}H-^{1}H COSY NMR spectrum. **b** 1,1 Adequate NMR spectrum.

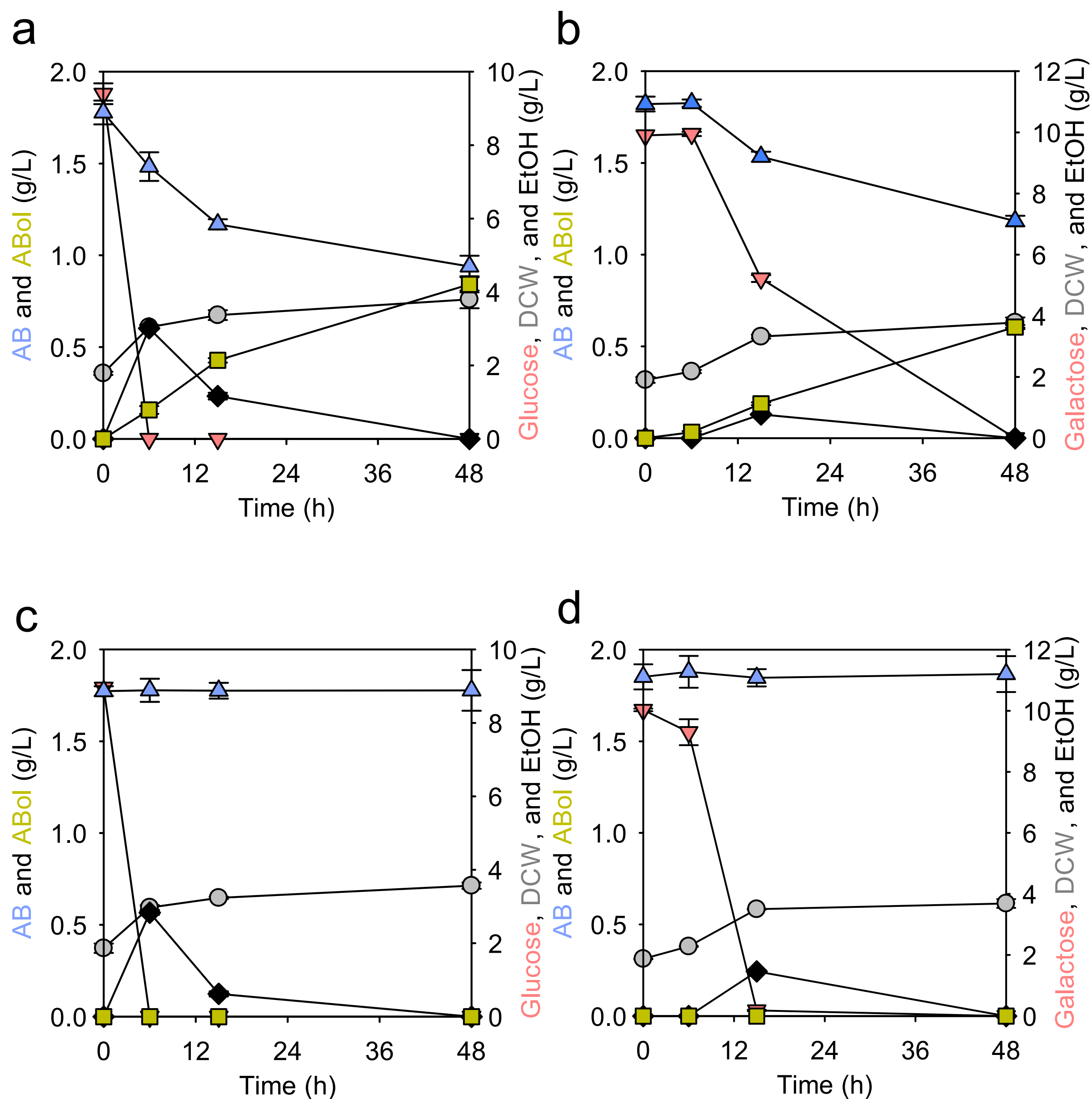

Fig. S5 Comparison of the AB uptake abilities and culture profiles of strain D452-L12 and the parental strain D452-2. a D452-L12 with glucose, **b** D452-L12 with galactose, **c** D452-2 with glucose, and **d** D452-2 with galactose. The strains were cultivated in Verduyn medium containing 1.8 g/L of AB and glucose or galactose as the carbon source. Values and error bars represent means and standard deviations of duplicate experiments.

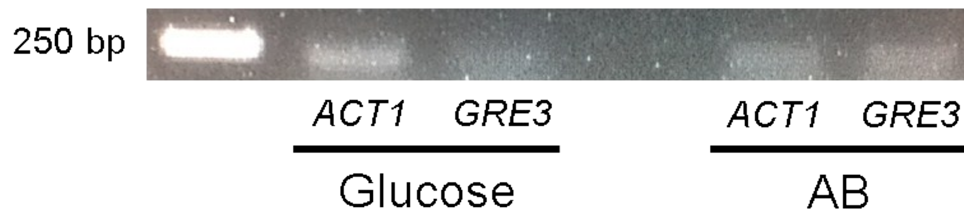

Fig. S6 Expression of the endogenous gene *GRE3* encoding AR in strain D452-L124 on AB compared to glucose. *ACT1* was used as a house-keeping gene.

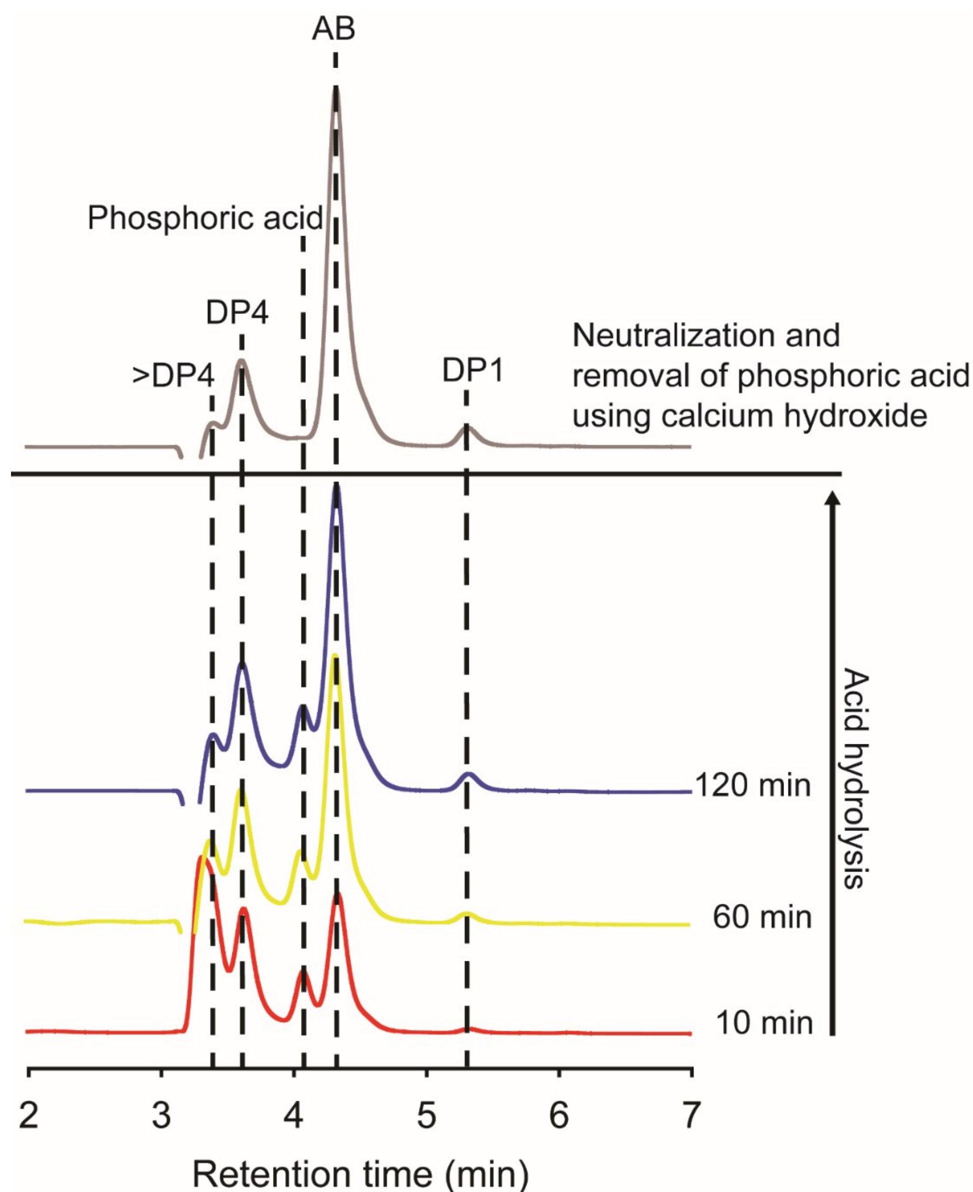

Fig. S7 HPLC analysis of the reaction products during acid hydrolysis of agarose to produce agarose hydrolysate containing mainly AB. Acid hydrolysis was performed on 20% (w/w) agarose at 95°C for 120 min using 2% (w/v) H_3PO_4. Reaction products were analyzed by HPLC equipped with an RID detector and a Rezex ROA-organic acid H^+ (8%) column using 0.005 N H_2SO_4 as a mobile phase at a flow rate of 0.6 ml/min at 65 °C.

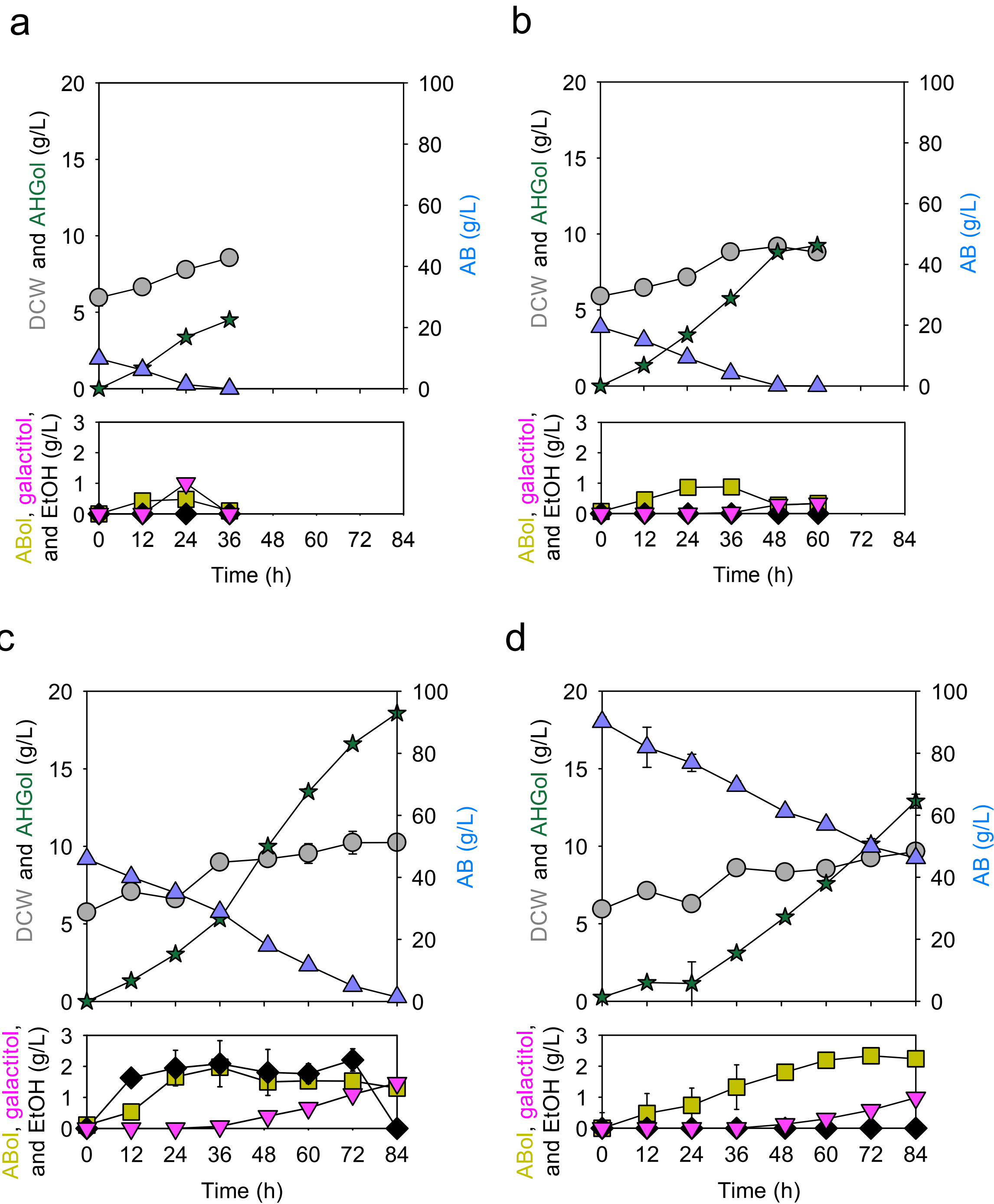

Fig. S8 Batch fermentation profiles of strain D452-L124 at the increased initial cell density at various AB loadings. At 6 g/L of the initial cell density, the initial AB loadings were **a** 10 g/L, **b** 20 g/L, **c** 45 g/L, and **d** 90 g/L AB in medium. Values and error bars represent means and standard deviations of duplicate experiments.

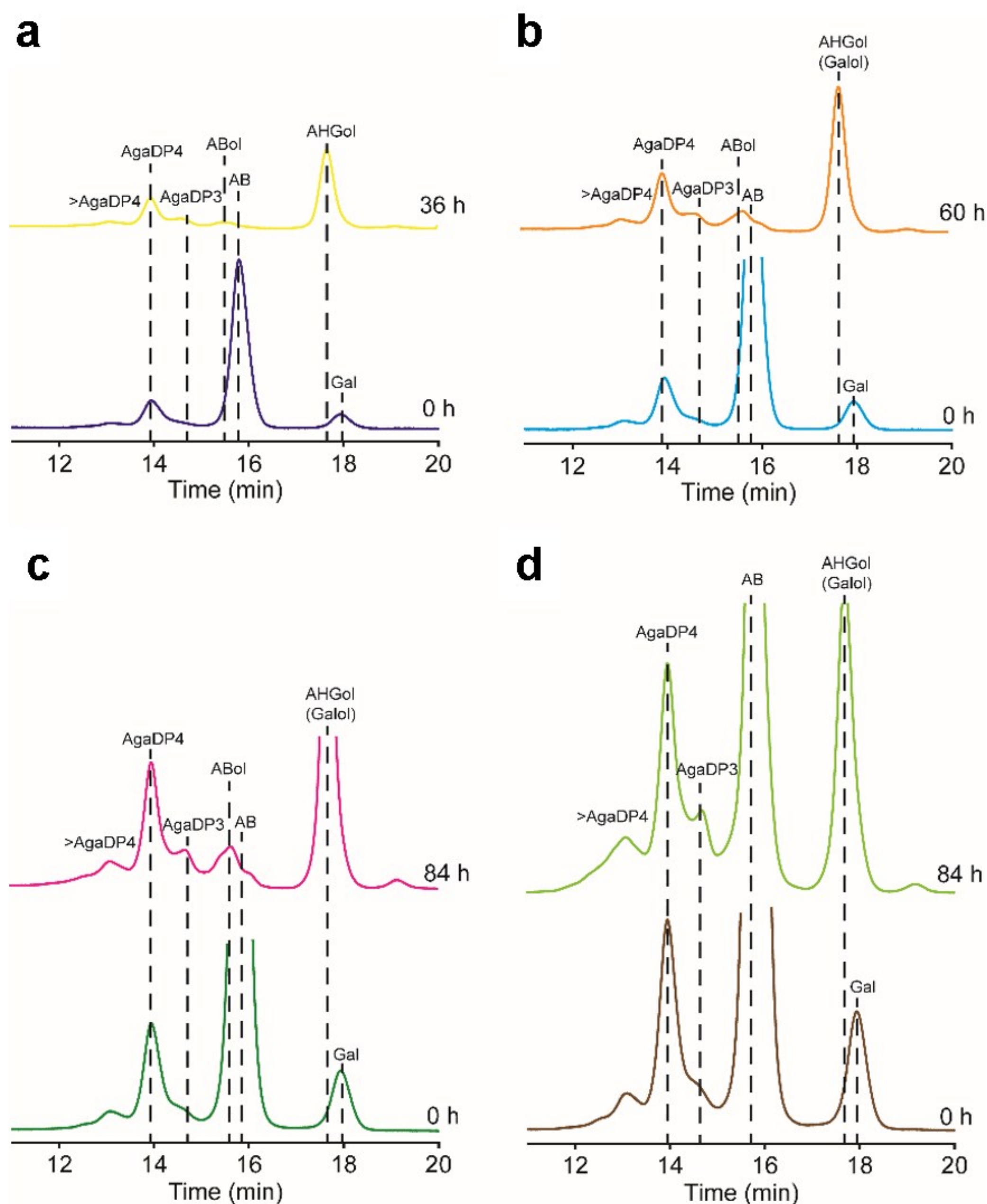

Fig. S9 Transglycosylation activity of β-galactosidase in D452-L124 during the batch fermentation for AHGol production. For each fermentation, **a** 10 g/L, **b** 20 g/L, **c** 45 g/L, and **d** 90 g/L of AB were added to its culture media and analyzed at 0 h and final point using HPLC equipped with the RID detector and a KS-802 column using distilled water as a mobile phase at a flow rate of 0.5 mL/min at 85 °C.

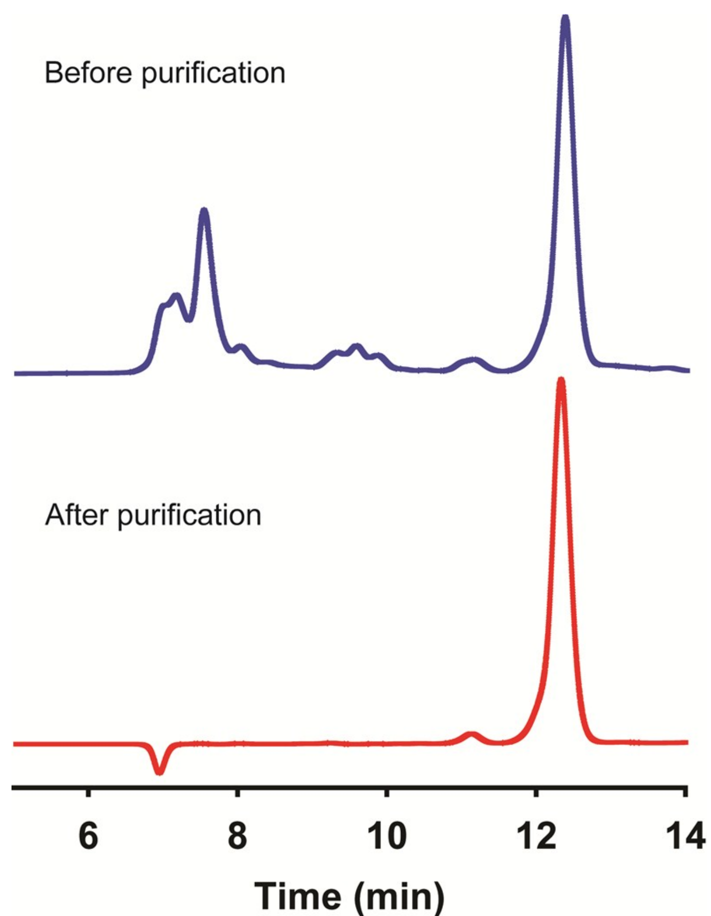

Fig. S10 Purification of AHGol from the fermentation culture broth of agarose hydrolysate. Purification was performed using size-exclusion chromatography with G-10 resin. The chromatograms were obtained using HPLC analysis with a KS-802 column.

Supplementary information of the intergenic sites for Cas9-based integration

NNNN: CDS of functional gene

NNNN: CDS of not essential protein, or protein with no function (or unknown function)

NNNN: tRNA

NNNN: δ sequence

nnnn: Non-coding region

nnnn: PAM sequence

nnnn: Target sequence for gRNA

nnnn: Homologous region for integration

nnnn: sequencing primer site

Supplementary information of the intergenic site CS8 for CRISPR-Cas9-based integration

[CS8 region, Chromosome XVI]

......TAGATCTTATACAAAAGCAACTGCGCGCTGTGGTAAAGCTACGGAAACAATGTCCTATCTG
TGGGAAGGTTTGTTCGAGACCTTCAACACTGAGGACTCATTACTTAATACATACGGGAGACAC
ACCTTTCAAATGTACTTGGGAGCATTGCAACAAATCTTTCAATGTCAAGAGTAACATGTTAAGG
CATTTAAGAACCCATCAAAAGAAAATAGCAAAGAAAAAACATCAGTGAaagcattgtttagaatattttgtttt
cagtgtgattttgatgtaggtgtgacactttttaccaagtaaaatgagtatagatatgtattagttccatataatatattacatgtagccaacaatcaat
ttttactgacttccgatttttgaatagtgagaggatttttgttcgaattataatttctacaagaaatttgtttgcgaatgcatcagcatttgtaatacggttct
atactgccgcacgatacattttattctgtctagttcgtaagacaaaggaccttttcatttagtacaacgttgccgacacggatgtcttgatgaatctct
gtaccaggaattcaattacctgactagtatgcagttcactatcaaaaaaaaaaaaaaggttaaattggtataatcatggttaaagaacaaaca
gcctctttcctctacaaaagtatgttaaatgtaggtagttcaaattgcggtcacgtatatgccaacgttgttaatatataatgatccattggaacaatg
aagtttacagaagctcaaaattaatgcaagaataattacgtcaagggagtcataacaaactaagaaaacaaaggaatttgactaagtttgag
cgcaatgatatatccatgcttgatattctgattcatacatatcgtttctgtcatttcaatggaagtctttatttatcaattaccacaacctttagtattgtgctt
attattggtcaaaaggagttcatgcgctagtgatagtcaacccatgaagtgattagtggaacatagaaggggataaatttcctatcggataaca
aagaaaaaccaggattattgttttaccaccacattttcaaactttcccgctgaatttcctactaccctcattgaaagaacgctttccactctttaggtt
ggcgttgtatcatcttttcttccattcaatgcaccaagacttaattttgttggttaaactcttattctttgtcagtttaaccaacaaaattacctacggtaatt
agtgaaaggccaaaatctaatgttacaatatgattcaatcattcttattgcggagaagctaccgagaagaaccctaccactggtacaacacaa
gggaacggtctcattgacctcgtgcataaccgggaaggtaacacatgaatatgctgtgtttgaatcatctccaatccaaactataggaatgaaa
aattttttttgggtacaatgtcttctcacaatatcgtattcttggtcagcaatgtttccaaggatgacagactctttctaagaaaagatttcaagttcaaa
atttacataacagtgaactatgcagaaaattttacttctagaccacgcatttaacaataatgtttgcccgcaacaataaagtatttggtcgctttaag
acgttatcagttattttatgagaaggtgtttaccattcgtgctcccatactaaaaccaaggggtaaataggacccacatttcgtttctgacagccaa
caacgtgccactagcaggacagttattggaggtttatattatgtaattattgttagctgataaatgatttacagaaagcgctgatgttcactgggctg
cttattcagcaatccccacgcatatttacttcaaagagaaagaatttttgatgtaagttttatttgaacgtggaactttggcgtggggcatttatggctg
ctatcacacaaatcgcggagcagagtaccccctggattatataacacaactcactaaaatcgctcaaaattggggggtacgggttagcgcggc
agctcatcgagggaacagcacctagtgcacgtttaattgatagtatcttcagaacgaaatcaaatttttcgcagtatcatttacgtttcaagaaactt
aaagtgtttggagaaattaggaacgaaaaaaagctactatagaaattgtagcctgtcttcaactgctgtcctcggcttttgacttgcctaagtatat
gccgcaatgggcgataagcgggagtgtcggtcagatctgtggtaaaaaaaagatagcaataaaaaattatgaatttaaacagtagtcttgaat

ttaaacaagtgatcttgacttgaatacttccatcaatgagtcgctaaaatgaaaatggctgcacaatctcctccgcatatcttaaaaggcagcaat
tagccagttgatcaccctacatttccacagaaacgaaaaagtgtaatagttatttttgcggccaatatcgtaagctctagcggtgccttagctttatg
attgttgcaagaaagtttgccttttgactctccttcaggtgtcagtattaacaaacggcgttgaatgtttaagttattattttattcatcaaatcgacttgg
atactttctcggtgatgtcgctaattggattataccattaaggttgtcaccagtgtaaaaattttgcatcggtacttgcaactacgtgtgaaggtcaag
ttactcaacgcaaagaaacgttaacatttttacaaattttagctgcagtactatgttttaagtaatccaaagggaactatttgttatatcccagaattat
ttacattcgtttcttagtttcataaacaatgaatacctattgaatggatagaaattctgacttgattttacgagttattattgctgacattagtccaaagac
atctcagttttgtttcctctacaacccaatgaggaggctagccagggctgtcgcccaaaaaaatagccaataaagcggcaactttctgt**ATGT**
CTTGCATTTTTTCCGCTGACTTGGGAGTAGAGTACAGCTGCGCGGAGTCGCGTATTACAAATCT
TGTACTTTGCATTCTGTGCATACGTGAGGAAAAAGCAGCACCTGTTGTGAAAAAAGATAAATTT
CTTTTTTTCATTTTTCTATCCTCAAAGGAAAATCTTTTCTGCGAAATTTCTAGCCGCTCTTCAGTT
TCCGAAGTAAAGTCGCTGAGGGCGAAAAACAACTTTATTTGCCCACACGCTGTTATAGGCTTC
CAGGGAGCAAAAGCGCAATTGGGAATAAGGTTTCCTCAATTGAATGGTTTGCTTCGCCTACAG
CATTAAccagaaggtaatttgatctcttgtatgtccgctaacagatcttgtccattaattgtttattttcttcaggtaggctgagcccactttgtagta
gctccctaccattacttttggccttgtagttattggctttgatttttctttctatgacaatcccgttgagcaaccttgcgaagggctctttttgaaaaaaag
gcgtgggcaagatctctgtat……

Supplementary information of the intergenic site CS6 for CRISPR-Cas9-based integration

[CS6 region, Chromosome VII]

ataaaa**TGAAACGGACAGGAATTGAACCTGCAACCCTTCGATTGCAATCTTATTCCGTGGAATTT
CCAAGATTTAATTGGAGTCGAAAGCTCTACCATTGAGCCACCGCTTC**atcttgaaatatcgaagatataaca
ttttacttgtctgccgaaattctgtgtttgctataatgtttgaattagaatctcttaaaatagctactcatacttcttcataactaatccattagtgaccatat
gaagtaatcggacgccacacatcattgatgtttcacgatggagaatgataacacactaagtggcattgtgggcaaagtaagttaaacacctatt
gctcaaatgatcaacttggtgtttgcacatatacggatgtaagtcatgacattgaaatcataatatgctttcatgataatcatatgagcatgtttaatat
tactaataaggctgtattctatacttctcttatatagaataagaagatctgcatttattcttgattgacactacagttcaacaattaattaccaacagaa
ttaattactacctatcgttcacaagctactacgttatcatacacaatgtaaaaatatgacacaaaaatggaaaaccgtcatcagatttaatggggc
tgaaacacaagaattcataatgtgatagaataatgggtgaagtgtataaagaagaatatataatattactgtgtagaaatatcaatttccctttgtg
agttctcataacctcgaggagaagtttttttacccctctccacagatcgatacttatcattaagaaaatgggacaccaaggttacggaaaaatcta
cccggtctacctaattactctcttggcgcactagttttccgaaaaaaacaggtaaattcttctttagataaagataaatataaaacttcacagccatt
cactcacacaaactagtcccttagggtgcgtataatgatctgtacatcttatttctatatatcttaccgtgtatttttttcttttctcaattcttgttcgcaaata
aaaagatattcgtgtttgtggaagaacactagttccgttttgtattcaacctggaaatttacaatagatcttcatcatcgtatgtctaccatgttaatctc
ccgttaaactgtttcacgttatcaagattatgtcatctattcctgggcgaacataattccttacaaaaacatttgtcattacacaagtgtaaggggtaa
tgaaaagtaattttgttacaagtacgcaaaattcgtttatttcaagaaacactaaggatcgtcatttccctttctgaccgatgttccttctttttgctatttttt
tcccgagtcatctcatcgttttgagtttttcctagtccattaaattgtcaccttactctcggaaaaaagaaacgacaaatgctcctagtgccgtttttcg
aagcttgaaaaaaaaaattgcaaattatttaattttgctgctaaggagttgaagtaggtgcattccgccttattgatcaccctgttagatttgttgcgat
cgttatagtgctagtttgtccattgttgtgtcataaaagatagctttgggagaaaattca**TCAAAACAACATATCATCAGCGTTAT
TACAATTCATTGTCCTTCCCAAGTTTTTTTGACGTATAATATTATCGCTATCTGACTCATTAGTAC
ACAAATACAGATATACAACCTCAAAATCAAAAATGCCTAGAAACCCATTGAAAAAGGAATATT
GGGCAGATGTAGTTGACGGATTCAAGCCGGCTACTTCTCCAGCCTTCGAGAATGAAAAAGAAT
CTACTACATTTGTTACCGAACTAACTTCCAAAACCGATTCTGCATTTCCATTAAGTAGCAAGGA
TTCACCTGGCATAAACCAAACCACAAACGATATTACCTCTTCAGATCGCTTCCGTCGTAATGAA
GACACAGAGCAGGAAGACAT**caacaacacca