CRISPR/Cas9-assisted seamless genome editing in *Lactobacillus plantarum* and its application in *N*-acetylglucosamine production

Running title: seamless genome editing of *Lactobacillus plantarum*

Ding Zhou[a], Zhennan Jiang[b], Qingxiao Pang[b], Yuan Zhu[a], Qian Wang[a,b] * Qingsheng Qi[b, c]

[a] National Glycoengineering Research Center, [b] State Key Laboratory of Microbial Technology, Shandong University, Qingdao 266237, P. R. China

[c] CAS Key Lab of Biobased Materials, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao 266101, P. R. China

Correspondence: qiqi20011983@gmail.com (Qian Wang)

Table S1. Strains and plasmids used in this study

Strain or plasmid	Relevant characteristics	References
Strains		
Lactobacillus plantarum WCFS1	wild type	(1)
Escherichia coli XL1-Blue	subcloning host	Transgen
E. coli DH5α	*ΔlacZ M15 recA*	Transgen
Lactococcus lactis MG1363	subcloning host	(2)
WG1	*Lactobacillus plantarum* WCFS1 *ΔnagB*	this study
WG2	*Lactobacillus plantarum* WCFS1 *ΔnagB*; P_{3a}:: *glmS1*	this study
WG3	*Lactobacillus plantarum* WCFS1 *ΔnagB*; P_{3a}:: $glmS1^{G472S}$	this study
Plasmids		
pNZ8148	source of chloramphenicol-resistance gene *cat*	lab stock
pSIP403	*spp*-based inducible expression vector; 256rep and pUC(pGEM)ori; Ery^R; P_{sppA}:: *gusA*	(3)
pSIP411	*spp*-based inducible expression vector; $SH71_{rep}$; Ery^R; P_{orfX}:: *gusA*	(3)
pLP-gba	pSIP411 derivative, P_{orfX}:: *lp_0640-lp_0641-lp_0642*	(4)
pLP-cre	pSIP411 derivative, P_{orfX}:: *cre*	(4)
p99S-Cas9	source of Cas9 gene and sgRNA	(5)
pNnsGM	pTrc99a, harboring nanA, slr1975, GNA1 and mutant glmS gene	(6)
pSIP-C9 (*nagB*)	pSIP403 derivative, P_{sppA}:: *Cas9* , P_{3a} :: sgRNA (*nagB*), Cm^R	this study
pSIP-C9 (lox)	pSIP403 derivative, P_{sppA}:: *Cas9* , P_{3a} :: sgRNA (lox), Cm^R	this study
pSIP-C9 (472)	pSIP403 derivative, P_{sppA}:: *Cas9* , P_{3a} :: sgRNA (472), Ery^R	this study
p411-RecT	pSIP411 derivative, P_{orfX}:: RecT, Cm^R	this study
p411-RecT-Dam	pSIP411 derivative, P_{orfX}:: RecT and Dam, Cm^R	this study

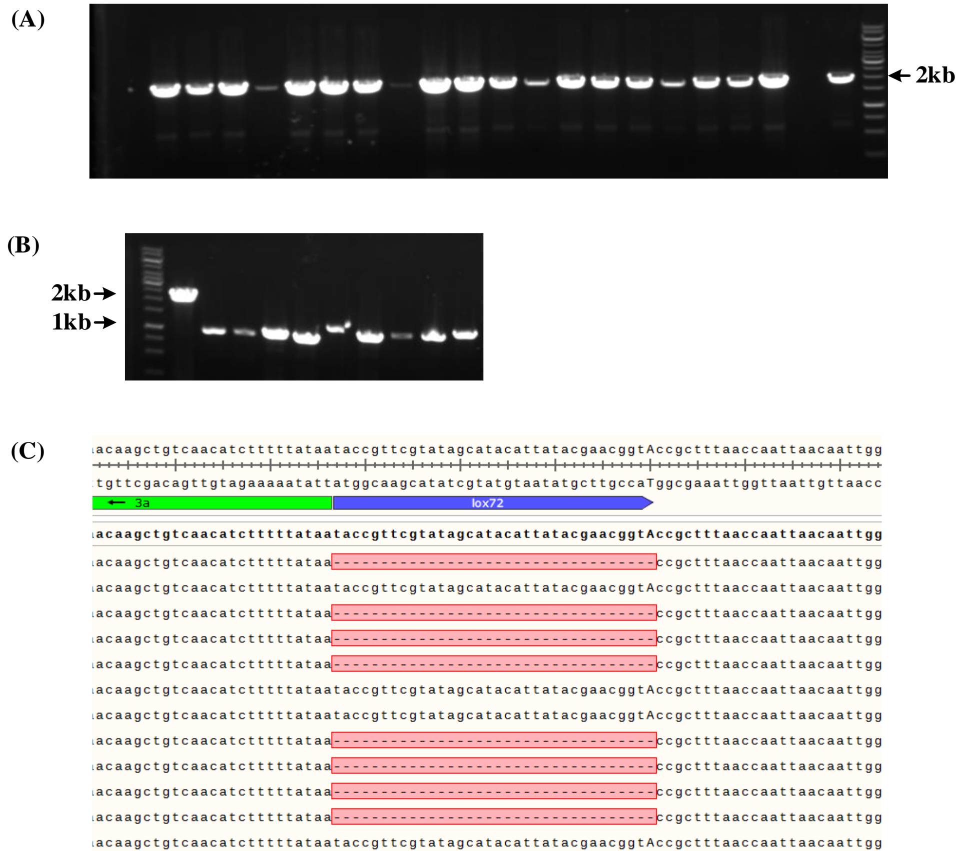
Table S2. Oligonucleotides used in this study.

Primer	Sequence(5'-3')
403-Cas9-f	AAGTATGCTTTATAAAATAATATATAGGAGTATGATTCCCcgagct TTTAGGAGGCAAAAATGGA
Cas9-3a-r	CGCAACTATTAAAGTTTATCATCATCAGAACAAGCTGTCAACATC TTTTTATAACCAGAAATCATCCTTAGCGAAAGCT
3a-gRNA (*nagB*)-f	CTGATGATGATAAACTTTAATAGTTGCGACTAGTACACCCAATTG ACTTACAAATCTTAGGGATGTTTTAGAGCTAGAAATAGCAAGTTA AAATAAGGCT
3a-gRNA(lox)-f	CTGATGATGATAAACTTTAATAGTTGCGACTAGTTACCGTTCGTAT AGCATACATTATACGAAGTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGCT
3a-gRNA(472)-f	CTGATGATGATAAACTTTAATAGTTGCGACTAGTCTTATTGAATAC GCCAAACGCCTTTTATATGTTTTAGAGCTAGAAATAGCAAGTTAA AATAAGGCT
sgRNA-403-r	TGGCGCCTTCGAACCCGGGGTACCGATAAAACGAAAGGCCCAG TCTTTCG
RecT-f	ATGAGTAATGAGCTAGTTACGATGGTTAAT
RecT-r	GGGTACCGAATTCCTCGAGTCTAGATTAGCTGGCGTCAAAGTCTC CG
p411-f	TCTAGACTCGAGGAATTCGGTACCC
p411-r	ATTAACCATCGTAACTAGCTCATTACTCATGGCTAAAATCTCCTTG TAATAGTATTTTATAGAAT
cm-f	AACAAATCGTTTAACTTCAGGAGAGATTACATGAACTTTAATAAA ATTGATTTAGACAATTGGAAGAGAA
cm-r	TTTCCAAATTTAAAAAAGCGACTCATAGAATTATAAAAGCCAGTC ATTAGGCCTATCTGA
p411-cm-f	TTCTATGAGTCGCTTTTTTAAATTTGGAAAGTT
p411-cm-r	GTAATCTCTCCTGAAGTTAAACGATTTGTT
Dam-f	CTTTGACGCCAGCTAATCTAGACTCGAGGTAACAGATTCGCAAG TAAGGAGTTTATATATCTGGCACAAGCAGAAACCG
Dam-r	TGCGTTCTCACTGGATGCTGTTCT
nagB-cf	GTGAACAAATAAATGGTTGGCGTAG
nagB-cr	CCTACCGCTACACCGCAAC
3a-cf	GAGACCCAGTGCCTCCCAAAC
3a-cr	GTGGTAACAAATGCGTCTGCAC
472-cr	AAGGTCGCTGGTATTCGCCA
472-cf	CAACGCTATCGCGCGAAGCAA
rpob-cf	GGATGGAACGGGTTGTGCG
rpob-cr	TTGCTGGACCATCGGCTAA
cm-cf	GGTGGCTACGTATTGACGGAAC
cm-cr	CTCTCCGTCGCTATTGTAACCAG

cm-cre-cf	GAGACCCAGTGCCTCCCAAAC
cm-cre-cr	TTGAAGCCTTCGCCGACCC
nagB-f	AGTGCTTGACCGATGGGAGA
nagB-r	TCTTTTCCGCCGGCAGCTTGATC
nagB-f2	AGGATCAAGCTGCCGGCGGAAAAGAGACGAAGCGGCTGCTAGCAA
nagB-r2	AACTGAAACCACTGCCACCAA
3a-f1	GCTGGAAGAAGCCTTGATTG
3a-r1	CCGCTTTAACCAATTAACAATTGGTC
3a-f2	GACCAATTGTTAATTGGTTAAAGCGGTACCGTTCGTATAATGTATGCTATACGAAGTTATTGATCCCCTTAGAAGCAAACTTAAGAGT
3a-r2	ATCATCATCAGAACAAGCTGTCAACATCTTTTTATAATACCGTTCGTATAGCATACATTATACGAAGTTATGCTCACCAACAACCTATCTTAACGAG
3a-f3	GTTGACAGCTTGTTCTGATGATGATAAACTTTAATAGTTGCGTGAGGTTGGTGACAGATATATTGACCACG
3a-r3	TTCGCCAAGTTTGTTGCCTCT
lox-r1	ATCATCATCAGAACAAGCTGTCAACATCTTTTTATAACCGCTTTAACCAATTAACAATTGGTC
lox-f2	GTTGACAGCTTGTTCTGATGATGATAAACTTTAATAGTTGCGTGAGGTTGGTGACAGATATATTGACCACG
g-f	TAAAATAAATATATAGGAGTATGATTCCCATGATACCAGCACAGGCAGGTTTAAAC
ssDNA-472-5	AAGGCCGTTTCTAATGAGACGGCGTAATCCAGTCCCCTTGATATATAAAAGGCGTTTGGCGTATTCAATAAGGCACTCT
ssDNA-472-3	AAGGCCGTTTCTAATGAGACGGCGTAATCCAGTCCCCGTGAAATATAAAAGGCGTTTGGCGTATTCAATAAGGCACTCT
ssDNA-rpob-1	TCAAACCACCAGGTCCTAAGGCTGATAAACGCCGCTTACGCGTTAATTCACCTAAGGGGTTGGTTTGGTCCATGAATTG
ssDNA-rpob-2	TCAAACCACCAGGTCCTAAGGCTGATAAACGCCGCTTCCGCGTTAATTCACCTAAGGGGTTGGTTTGGTCCATGAATTG
ssDNA-rpob-3	TCAAACCACCAGGTCCTAAGGCTGATAAACGCCGCTTCCTCGTTAATTCACCTAAGGGGTTGGTTTGGTCCATGAATTG
ssDNA-rpob-4	TCAAACCACCAGGTCCTAAGGCTGATAAACGCCGCTTCCTAGTTAATTCACCTAAGGGGTTGGTTTGGTCCATGAATTG
ssDNA-rpob-5	TCAAACCACCAGGTCCTAAGGCTGATAAACGCCGCTTCCTGCTTAATTCACCTAAGGGGTTGGTTTGGTCCATGAATTG
ssDNA-rpob-3+N	TCAAACCACCAGGTCCTAAGGCTGATAAACGCCGCTTNCGGCTTAATTCACCTAAGGGGTTGGTTTGGTCCATGAATTG

Underlined sequences indicate homology arms for Gibson assembly. Wavy underlined sequences indicate spacers for sgRNA. Yellow background sequence indicates RBS. Dark blue background sequences indicate lox site. Red coloured words indicate mutant bases.

Figure S1. The efficiency of the two-step gene insertion.

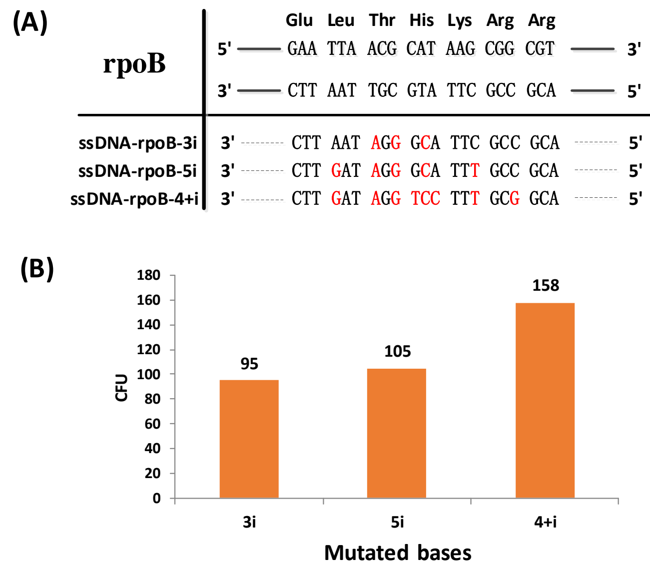


(A) The efficiency of the two-step gene insertion was also divided into two parts. The efficiency of the first step that insertion of the chloramphenicol resistance gene, loxP sites and the 3a promoter was 82% as shown in the Supplementary Figure 1A. We use primers cm-cf/cm-cr to verify this result, the primer cm-cr is located on the inserted chloramphenicol gene and there is no band if it is not inserted successfully. (B) The results of removing the chloramphenicol label using the loxP / Cre system. We verify this result with primers cm-cre-cf/cm-cre-cr. The strains was cultured for 24 hours after transformation of Cre enzyme, all of the nine tested colonies showed the mutant genotype and could not grow on chloramphenicol-containing media, the first band is a negative control. (C) The second step is CRISPR-assisted seamless genome knockout of loxP sites, 57 colonies were obtained on the plate and in 24 tested colonies, 14 were correct mutants identified by colony PCR and sequenced. The positive rate is 58.3%. Part of the sequencing results are shown in the Supplementary Figure 1C.

Figure S2. Partial DNA sequencing results of point mutations G472S.

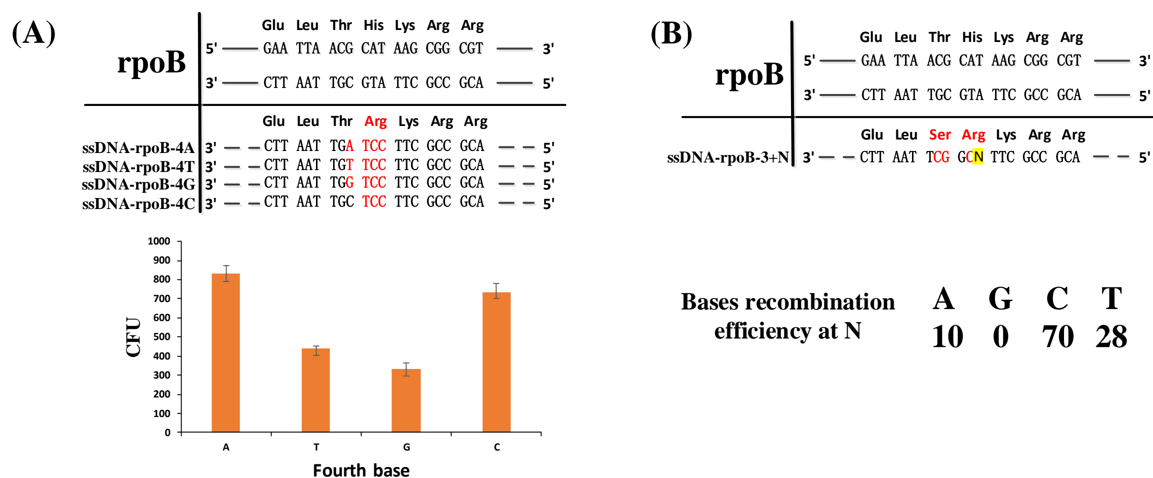
The ssDNA recombineering result was verified by PCR using primers 472-cf/472-cr. The sequencing results was shown above.

Figure S3. Mutation efficiency of interval mismatches ssDNA.



(A)The ssDNA with various interval mutant bases marked red were used to change the genome. (B)The efficiency of ssDNA recombination resulting in various mismatches.

Figure S4. The recombination efficiency of different bases.



(A)The fourth base is a degenerate base, we design single-stranded DNAs named ssDNA-rpob-4A/G/C/T to verify the mismatch efficiency of each base at this site. When the fourth base is A, the mismatch efficiency is the highest and single colonies can achieve about 800 on rifampicin-resistant plates which is probably twice the other bases. (B) We designed an ssDNA named ssDNA-rpoB-3+N (N=A/G/C/T) to calculate the mismatch efficiency of this site. After transformation of ssDNA, we got single colonies growing on plates containing rifampicin resistance and sequenced the genome of single colonies to calculate the number of bases that appeared in the position of yellow bottom N. The mismatch efficiency towards base C is the highest, while base G is the lowest.

References

1. Kleerebezem M, Boekhorst J, van Kranenburg R, Molenaar D, Kuipers OP, Leer R, Tarchini R, Peters SA, Sandbrink HM, Fiers MW, Stiekema W, Lankhorst RM, Bron PA, Hoffer SM, Groot MN, Kerkhoven R, de Vries M, Ursing B, de Vos WM, Siezen RJ. 2003. Complete genome sequence of *Lactobacillus plantarum* WCFS1. Proc Natl Acad Sci U S A 100:1990-5.
2. Wegmann U, O'Connell-Motherway M, Zomer A, Buist G, Shearman C, Canchaya C, Ventura M, Goesmann A, Gasson MJ, Kuipers OP, van Sinderen D, Kok J. 2007. Complete genome sequence of the prototype lactic acid bacterium *Lactococcus lactis* subsp. cremoris MG1363. J Bacteriol 189:3256-70.
3. Sorvig E, Mathiesen G, Naterstad K, Eijsink VG, Axelsson L. 2005. High-level, inducible gene expression in Lactobacillus sakei and *Lactobacillus plantarum* using versatile expression vectors. Microbiology 151:2439-49.
4. Yang P, Wang J, Qi Q. 2015. Prophage recombinases-mediated genome engineering in *Lactobacillus plantarum*. Microb Cell Fact 14:154.
5. Su T, Jin H, Zheng Y, Zhao Q, Chang Y, Wang Q, Qi Q. 2018. Improved ssDNA recombineering for rapid and efficient pathway engineering in *Corynebacterium glutamicum*. Journal of Chemical Technology & Biotechnology 93:3535-3542.
6. Kang J, Gu P, Wang Y, Li Y, Yang F, Wang Q, Qi Q. 2012. Engineering of an N-acetylneuraminic acid synthetic pathway in *Escherichia coli*. Metab Eng 14:623-9.