Supplementary Materials for

Genetic Engineering of E. coli K-12 for Heterologous Carbohydrate Antigen Production

Caixia Li *et al.*

Supplementary methods

Suicide plasmid construction processes

The suicide plasmid construction processes, including each primer pair used, can be found in the supplementary Fig S2-S6.

Firstly, homologous arms of upstream (HA_UP) and downstream (HA_down) of the *E. coli* W3110 O-antigen locus were amplified. A fusion PCR was performed to combine these two homologous regions. Subsequently, the fused homologous arms and the linearized pHY093 (pRE112-Cm-GFP) vector were assembled end to end, resulting in pHY095 (pRE112-Cm-ΔOAg) (Supplementary Fig. S2). Similarly, pHY129 (pRE112-Cm-ΔOAg::*viaB* locus) was constructed by assembling three DNA fragments in order: the linearized pHY095 vector, the first half of the *S*. Typhi viaB locus, and the second half of the *S*. Typhi *viaB* locus. (Supplementary Fig. S3).

Secondly, an FRT-flanked antibiotic-resistant gene cassette was cloned from pCP3 and inserted into the middle of the homologous arms, resulting in pHY099 (pRE112-Km-ΔOAg::Cm_FRT2) (Supplementary Fig. S4). Subsequently, the linearized pHY099 vector was assembled with the intact O-antigen gene clusters of *S*. Typhimurium (OAg^{ST}) and *S*. Enteritidis (OAg^{SE}), resulting in pHY100 (pRE112-Km-ΔOAg::OAg^{ST}-Cm_FRT2) and pHY101 (pRE112-Km-ΔOAg::OAg^{SE}-Cm_FRT2), respectively (Supplementary Fig. S5). Notably, the OAg^{ST} and OAg^{SE} gene clusters were directly amplified by long-rang PCR using the *S*. Typhimurium and *S*. Enteritidis genomes as template, respectively. Since the OAg^{SA} and OAg^{SE} gene clusters are almost identical except that the tyv gene is frameshifted and non-functional in *S.* Paratyphi A, we cloned the OAg^{SE} gene cluster into two segments, deliberately bypassing the *tyv* gene (Supplementary Fig. S6). pHY102 (pRE112-Km-

ΔOAg::OAg^{SA}-Cm_FRT2) was then constructed by assembling linearized pHY099 vector and the PCR-amplified two segments of the OAg^{SE} gene cluster in order.

The mechanism of suicide plasmid-mediated allelic exchange in this study

χ7232 is a genetic engineered *E. coli* host strain that is capable of expressing the *pir* gene (due to the λpir genotype) and can be used to maintain and amplify plasmids with the R6K origin of replication.

The term "suicide" means the plasmid cannot survive on its own as they possess an origin of replication, such as R6K origin, that does not function in the target bacterial species. Generally, suicide plasmid contains a positive selectable marker (antibiotic resistance gene) and a counter-selectable marker (e.g., the *sacB* gene, which makes bacteria sensitive to sucrose).The fundamental mechanism of a suicide plasmid to construct a bacterial mutant is allelic exchange (homologous recombination). It involves the following basic steps: 1) Clone. The left and right homologous arm with (if insertion mutant is intended) or without (if deletion mutant is intended) a DNA fragment embedded are cloned into the suicide plasmid and electroporated into the χ7232. Positive recombinants are selected by PCR. 2) Conjugation. The successful recombinants are extracted from χ7232, and latter electroporated into a donor strain χ7213, which carries all the necessary elements for conjugation, enabling the transfer of suicide plasmids to a recipient cell through a rolling circle replication mechanism (due to the RP4-2-Tc::Mu λpir genotype). Meanwhile, the χ7213 donor strain is diaminopimelic acid (DAP) auxotrophy, which means it could be easily counter-selected after the conjugation. 3) The first Homologous Recombination. The suicide plasmid is integrated into the chromosome of a recipient cell via its endogenous homologous recombination system (RecA-dependent). The successfully integrants could be positively selected due to the antibiotic resistance marker. 4) The second Homologous Recombination. The second crossover event can excise the integrated plasmid backbone from the chromosome and results into two outcomes: either the desired mutation or the wild-type reversion. After the second recombination, bacteria are grown on media containing the counter-selective agent. Only those that have lost the plasmid backbone will survive. 5) Screening. Surviving colonies are screened by PCR to identify those that have retained the desired mutation and lost the plasmid backbone and selectable marker.

Supplementary Fig. S1

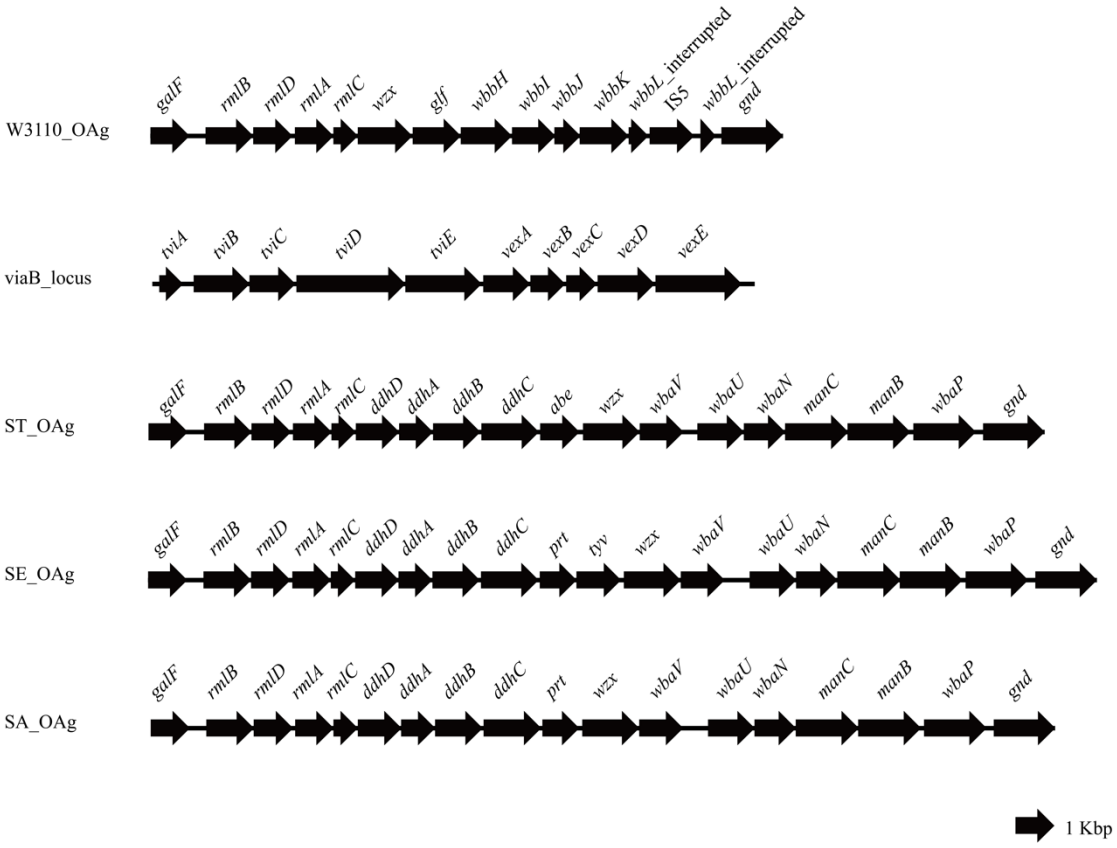


Fig. S1 Schematic representation of O-antigen gene cluster and *viaB*_locus.

The O-antigen gene clusters are within the region of *galF* and *gnd* genes on *E. coli* W3110, *S.* Paratyphi A, *S.* Typhimurium, or *S.* Enteritidis genomes, which could be accessed through the GenBank accession numbers NC_007779.1, NZ_CP019185.1, CP002614.1, and CP007361.1 respectively. The *viaB*_locus could be accessed through the GenBank accession numbers D14156.1. Diagrams were drawn to scale.

Supplementary Fig. S2

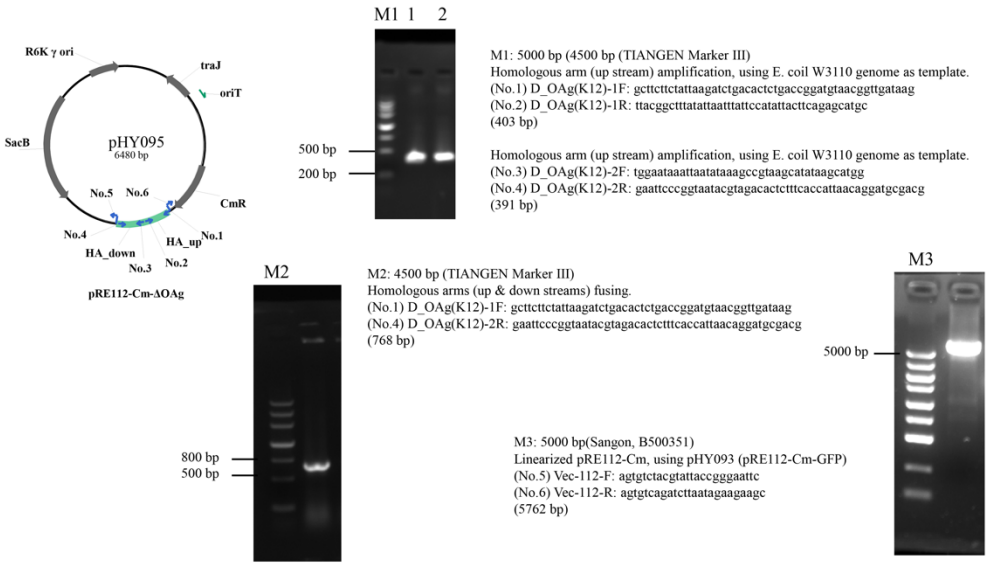


Fig. S2 pHY095 plasmid construction.

pHY095 (pRE112-Cm-ΔOAg) was derived from pHY093 (pRE112-Cm-GFP) by replacing the GFP gene with ΔOAg homologous arms.

Supplementary Fig. S3

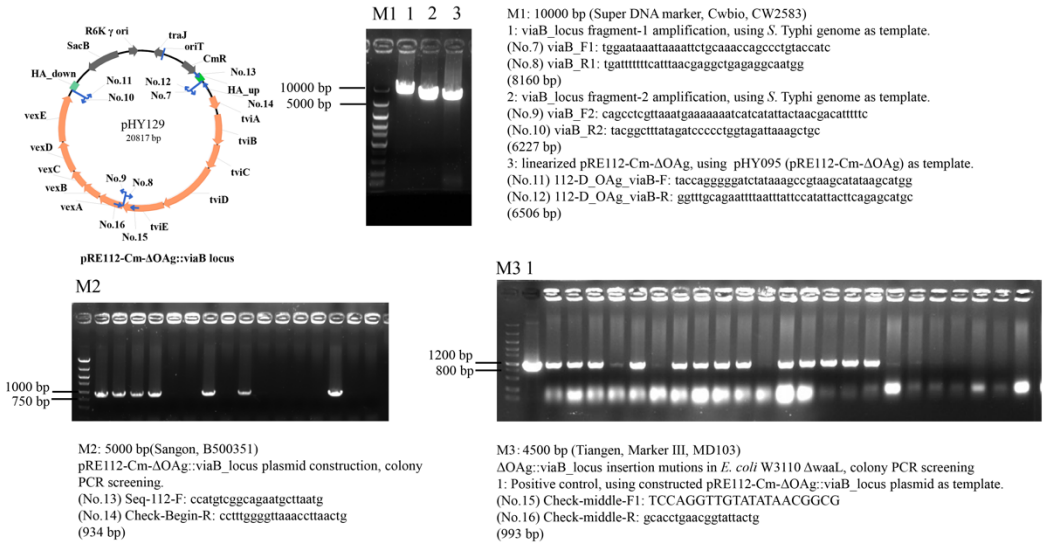


Fig. S3 pHY129 plasmid construction and ΔOAg::*viaB*_locus mutant screening.

pHY129 (pRE112-Cm-ΔOAg::viaB_locus) was derived from pHY095 (pRE112-Cm-ΔOAg) by embedding the viaB_locus into the ΔOAg homologous arms. Recombinants in χ7232 were first

screened by colony PCR and later confirmed by antisera (anti-Vi) glass-slide agglutination assays. ΔOAg::viaB_locus insertion mutations in *E. coli* W3110 background were also screened by colony PCR and later confirmed by antisera (anti-Vi) glass-slide agglutination assays. The final successful mutant was verified by whole genome sequencing.

Supplementary Fig. S4

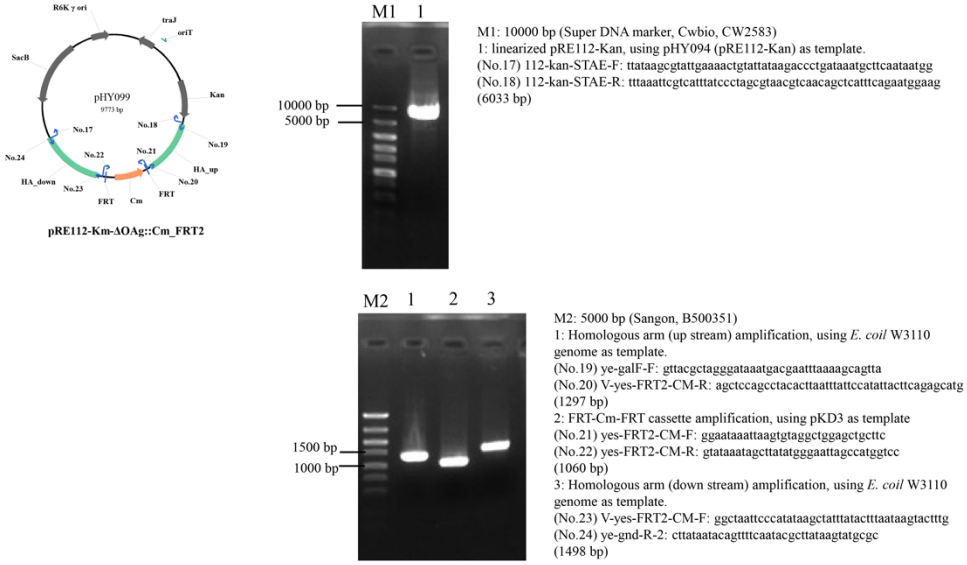
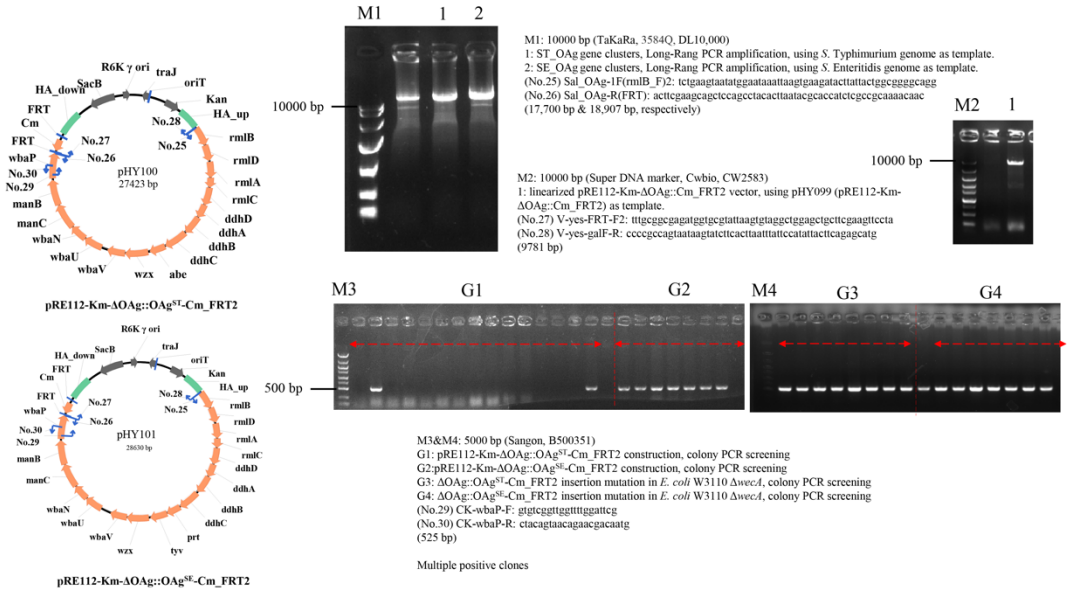


Fig. S4 pHY099 plasmid construction.

pHY099 (pRE112-Km-ΔOAg::Cm_FRT2) was derived from pHY094 (pRE112-Km-GFP) by replacing the GFP gene with the ΔOAg::Cm_FRT2 cassette. Specifically, linearized pHY094 vector, Homologous up and down fragments, and the FRT-flanked chloramphenicol-resistant (FRT-Cm-FRT) cassette were assembled in order *in vitro*.

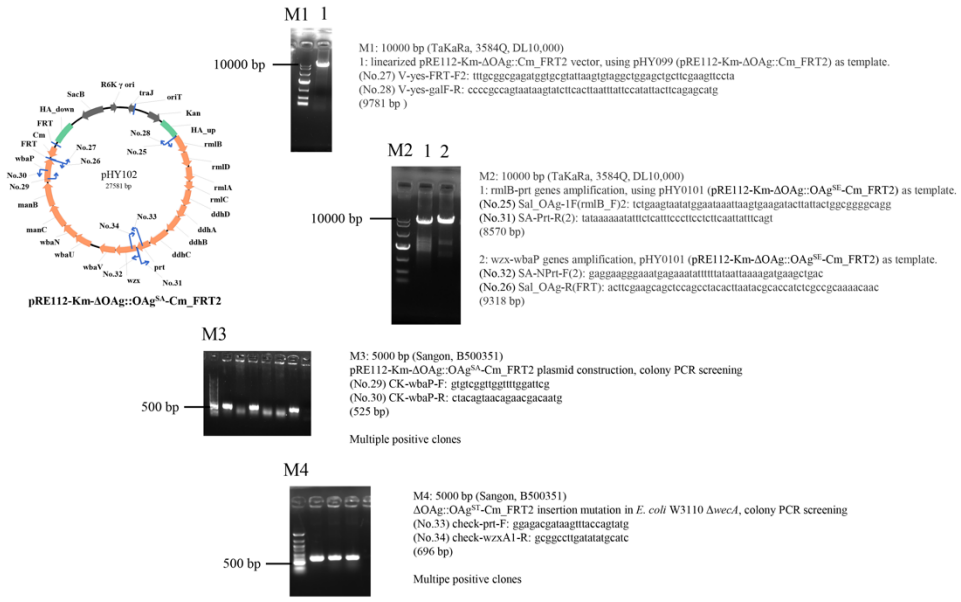
Supplementary Fig. S5



pHY100, pHY101 plasmids construction and ΔOAg::OAgST-Cm_FRT2, ΔOAg::OAg^{SE}-Cm_FRT2 mutants screening.

pHY100 (pRE112-Km-ΔOAg::OAgST-Cm_FRT2) and pHY101 (pRE112-Km-ΔOAg::OAg^{SE}-Cm_FRT2) were derived from pHY099 (pRE112-Km-ΔOAg::Cm_FRT2) by embedding the O-antigen gene clusters of *S.* Typhimurium (ST_OAg) and *S.* Enteritidis (SE_OAg) into the ΔOAg::Cm_FRT2 homologous arms. Specifically, the linearized pHY099 vector, and amplified ST_OAg or ST_OAg gene cluster were assembled in order *in vitro*. Recombinants in χ7232 were first screened by colony PCR and later confirmed by anti-O4 or anti-O9 antisera agglutination assays. ΔOAg::OAgST-Cm_FRT2 or ΔOAg::OAg^{SE}-Cm_FRT2 mutations in *E. coli* W3110 Δ*wecA* background were also screened by colony PCR first, and later confirmed by anti-Vi antisera agglutination assays. The FRT-flanked chloramphenicol-resistant cassette was excised by pCP20 plasmid, and the final successful mutants were verified by whole genome sequencing.

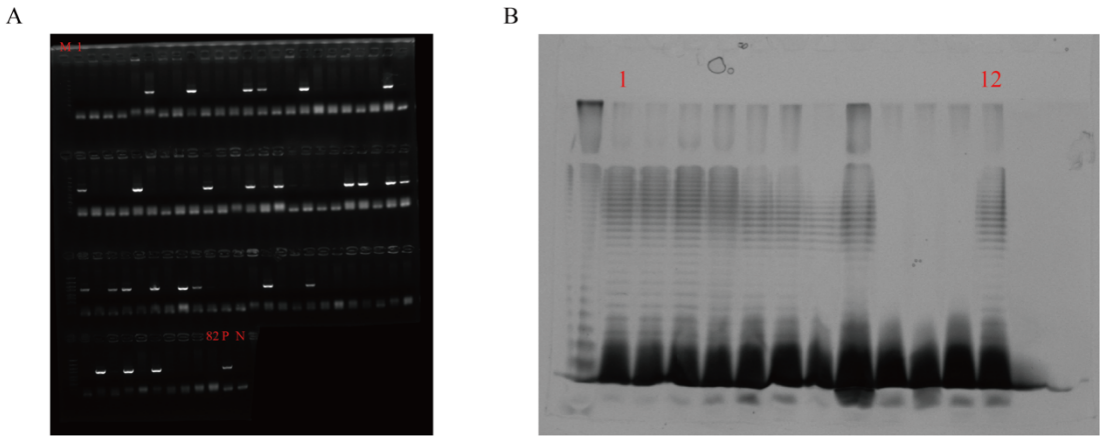
Supplementary Fig. S6



pHY102 construction and ΔOAg::OAg^{SA}-Cm_FRT2 mutant screening.

pHY102 (pRE112-Km-ΔOAg::OAg^{SA}-Cm_FRT2) was derived from pHY099 (pRE112-Km-ΔOAg::Cm_FRT2) by embedding the O-antigen gene cluster of S. Paratyphi A (SA_OAg) into the ΔOAg::Cm_FRT2 homologous arms. Specifically, the SA_OAg gene cluster was cloned from pHY101 (pRE112-Km-ΔOAg::OAg^{SE}-Cm_FRT2) by deliberately bypassing the *tyv* gene. The linearized pHY099 vector, the first half and the second half fragments of SE_OAg gene cluster were assembled in order *in vitro*. Recombinants in χ7232 were first screened by colony PCR and later confirmed by anti-O2 antisera agglutination assays. ΔOAg::OAg^{SA}-Cm_FRT2 mutation in *E. coli* W3110 Δ*wecA* background were also screened by colony PCR first and later confirmed by anti-O2 antisera agglutination assays. The FRT-flanked chloramphenicol-resistant cassette was excised by pCP20 plasmid, and the final successful mutants were verified by whole genome sequencing.

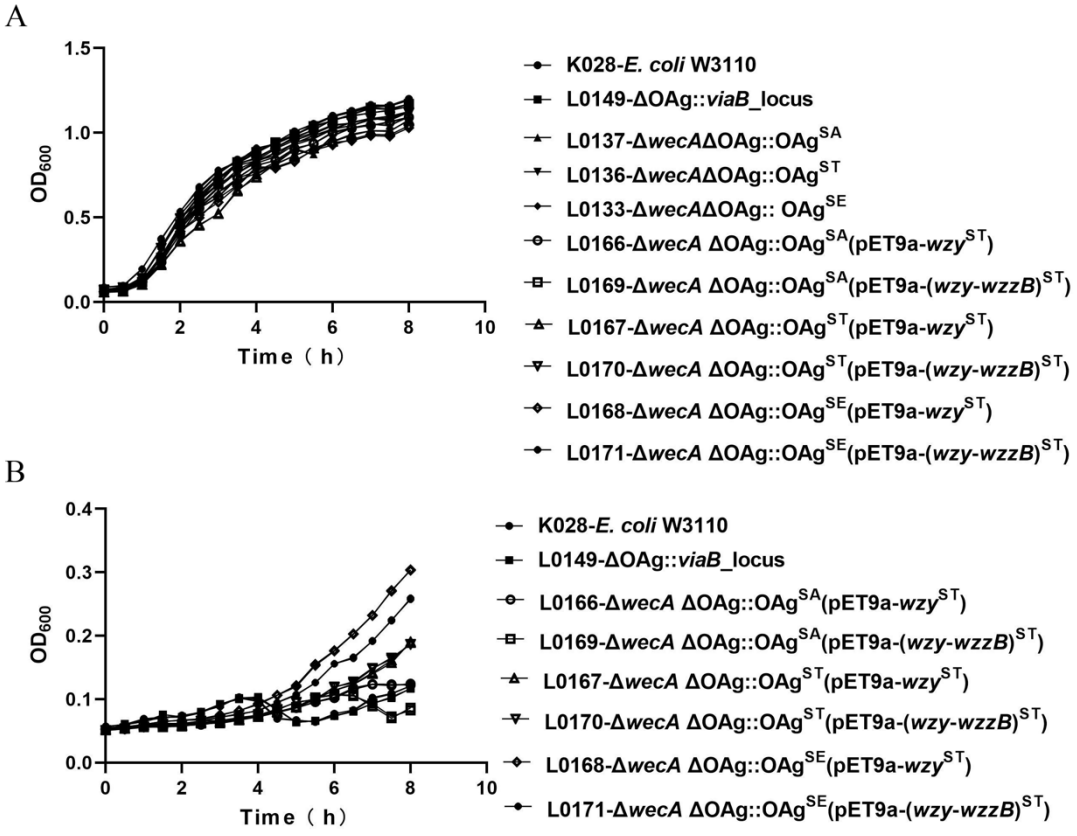
Supplementary Fig. S7



Editing efficiency and positive rates of scarless insertion mutations.

(A) PCR screening of the *E. coli* W3110 ΔOAg:: $(wzy\text{-}wzzB)^{ST}$-$(rmlB\text{-}wbaP)^{ST}$ scarless insertion mutations. A total of 82 colonies were randomly selected, and 26 colonies were PCR-positive, resulting in an editing efficiency of approximately 3.17×10^{-1}; (B) The LPS phenotype was confirmed by silver staining. 12 colonies were randomly selected from the 26 PCR-positive colonies, and 9 colonies successfully synthesized the LPS, yielding a positive rate of approximately 75%.

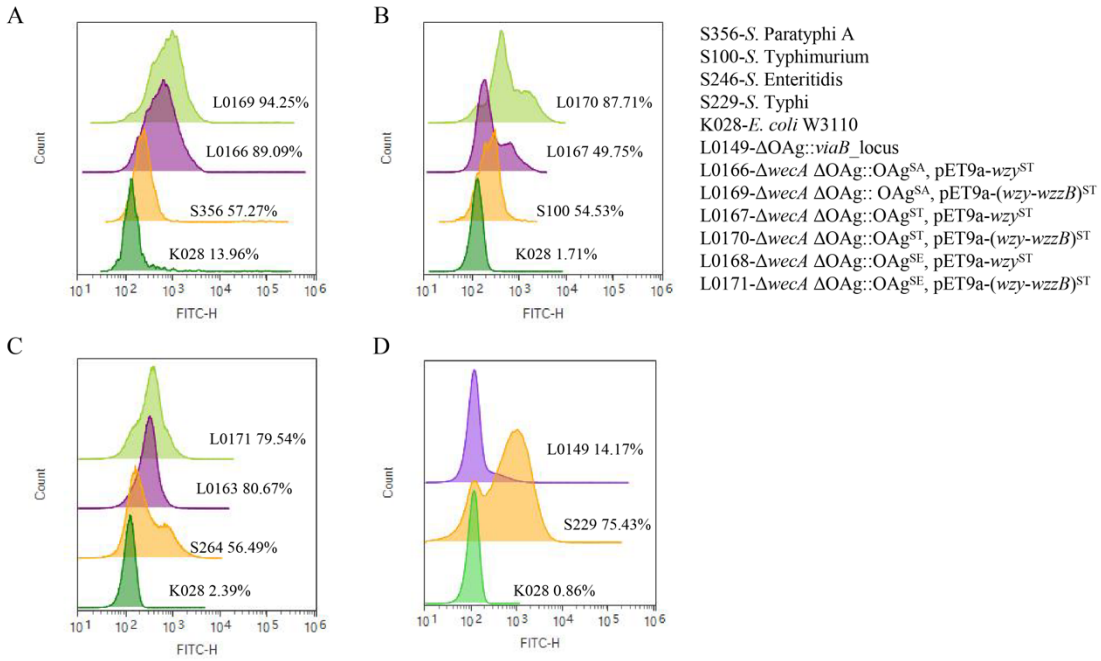
Supplementary Fig. S8



Growth curves of *E. coli* W3110 and its glycoengineered derivatives.

The A600 of each strain was measured at 30-minute intervals over an 8-hour period. Each data point represents the average A600 from two independent cultures. The identity of each strain is indicated by the legend to the right of the graph.

Supplementary Fig. S9



Bacterial surface polysaccharide antigen analysis

Flow cytometric analysis of intact cells using antiserum that recognizes *S*. Paratyphi A O2 (A), *S*. Typhimurium O4 (B), *S*. Enteritis O9 (C) O-antigen and *S*. Typhi Vi capsule (D). The identity of each stained cell was indicated on the right.

Supplementary Table S1. Primers used in this work

No	Primer name	Sequence 5'-3'
1	D_OAg(K12)-1F	gcttcttctattaagatctgacactctgaccggatgtaacggttgataag
2	D_OAg(K12)-1R	ttacggctttatattaatttattccatattacttcagagcatgc
3	D_OAg(K12)-2F	tggaataaattaatataaagccgtaagcatataagcatgg
4	D_OAg(K12)-2R	gaattcccggtaatacgtagacactctttcaccattaacaggatgcgacg
5	Vec-112-F	agtgtctacgtattaccgggaattc
6	Vec-112-R	agtgtcagatcttaatagaagaagc
7	viaB_F1	tggaataaattaaaattctgcaaaccagccctgtaccatc
8	viaB_R1	tgatttttttcatttaacgaggctgagaggcaatgg
9	viaB_F2	cagcctcgttaaatgaaaaaaatcatcatattactaacgacatttttc
10	viaB_R2	tacggctttatagatccccctggtagattaaagctgc
11	112-D_OAg_viaB-F	taccaggggatctataaagccgtaagcatataagcatgg
12	112-D_OAg_viaB-R	ggtttgcagaattttaatttattccatattacttcagagcatgc
13	Seq-112-F	ccatgtcggcagaatgcttaatg
14	Check-Begin-R	cctttggggttaaaccttaactg
15	Check-middle-F	gaatactattcgttggcgcc
16	Check-middle-R	gcacctgaacggtattactg
17	112-kan-STAE-F	ttataagcgtattgaaaactgtattataagaccctgataaatgcttcaataatg g
18	112-kan-STAE-R	tttaaattcgtcatttatccctagcgtaacgtcaacagctcatttcagaatggaa g
19	ye-galF-F	gttacgctagggataaatgacgaatttaaaagcagtta
20	V-yes-FRT2-CM-R	agctccagcctacacttaatttattccatattacttcagagcatg
21	yes-FRT2-CM-F	ggaataaattaagtgtaggctggagctgcttc
22	yes-FRT2-CM-R	gtataaatagcttatatgggaattagccatggtcc
23	V-yes-FRT2-CM-F	ggctaattcccatataagctatttatactttaataagtactttg
24	ye-gnd-R-2	cttataatacagttttcaatacgcttataagtatgcgc
25	Sal_OAg-1F(rmlB_F)2	tctgaagtaatatggaataaattaagtgaagatacttattactggcggggcag g
26	Sal_OAg-R(FRT)	acttcgaagcagctccagcctacacttaatacgcaccatctcgccgcaaaa caac
27	V-yes-FRT-F2	tttgcggcgagatggtgcgtattaagtgtaggctggagctgcttcgaagttc cta
28	V-yes-galF-R	ccccgccagtaataagtatcttcacttaatttattccatattacttcagagcatg
29	CK-wbaP-F	gtgtcggttggttttggattcg
30	CK-wbaP-R	ctacagtaacagaacgacaatg
31	SA-Prt-R(2):	tataaaaaatatttctcatttcccttcctcttcaattatttcagt
32	SA-NPrt-F(2)	gaggaagggaaatgagaaatatttttttataattaaaagatgaagctgac
33	check-prt-F	ggagacgataagtttaccagtatg
34	check-wzxA1-R	gcggccttgatatatgcatc

Supplementary Table S2

1. The DNA sequences of pET9a-*wzy*ST:

TTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATC
AGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAG
AAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCT
GCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCA
AAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGG
TGAGAATGGCAAAAGCTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCA
GCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCAT
TCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGAC
AATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGC
ATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGC
TGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTAC
GGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTT
AGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGT
TTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGT
CGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATC
AGCATCCATGTTGGAATTTAATCGCGGCCTCGAGCAAGACGTTTCCCGTTG
AATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTT
TATTGTTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGT
CAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGC
GCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTT
TGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTC
AGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGG
CCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAAT
CCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTT
GGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACG
GGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACT
GAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGG
AGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGC
GCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTC
GGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGG
GGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCT
GGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGAT
TCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGC
AGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGC
GCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCA
TATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCA
GTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACA
CCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCAT
CCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGG
TTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATC
AGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCA
GCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGG

CCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAG
GGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATG
CTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTG
TGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATC
ACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACA
GGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGG
GCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACC
ATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCAC
GTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGC
CAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGG
CCAGGACCCAACGCTGCCCGAGATGCGCCGCGTGCGGCTGCTGGAGATGG
CGGACGCGATGGATATGTTCTGCCAAGGGTTGGTTTGCGCATTCACAGTTC
TCCGCAAGAATTGATTGGCTCCAATTCTTGGAGTGGTGAATCCGTTAGCG
AGGTGCCGCCGGCTTCCATTCAGGTCGAGGTGGCCCGGCTCCATGCACCG
CGACGCAACGCGGGGAGGCAGACAAGGTATAGGGCGGCGCCTACAATCC
ATGCCAACCCGTTCCATGTGCTCGCCGAGGCGGCATAAATCGCCGTGACG
ATCAGCGGTCCAATGATCGAAGTTAGGCTGGTAAGAGCCGCGAGCGATCC
TTGAAGCTGTCCCTGATGGTCGTCATCTACCTGCCTGGACAGCATGGCCTG
CAACGCGGGCATCCCGATGCCGCCGGAAGCGAGAAGAATCATAATGGGG
AAGGCCATCCAGCCTCGCGTCGCGAACGCCAGCAAGACGTAGCCCAGCGC
GTCGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGG
TGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAA
TACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCT
CGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATG
ATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCA
CCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGACGCT
CTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGG
CCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGC
CCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAA
GCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTC
GGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCA
CGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATAC
GACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTT
AACTTTAAGAAGGAGATATACATccattagagtggaaagttggcgacgaaccctactatccagttaat
gataataaaaacatggagctttttaagaaatatagagagttagctagcagagaagacaaggttatatttggcgggcgtttggc
cgagtataaatattatgatatgcatcaagtgatatctgccgctctttatcaagtgaaaaatataatgagtacggattaatgcttat
aatttcatacattgcattatgcttattatttatagtttatctctatactctttccgtaagaattgaggggaaaataataaatgtaatggt
cccatacctgataataacagtccctacactgtacgtgtttgaaggtatatttgtatacctctcagaggtgcagaattatacagtg
gaatatttgtttttctatacttgctatataacatacatagcatcatttgttatttcttatctttacacacaaagaaaacccatatacaac
aaatcaaacacgaaaaataaaccaaggtatgtgtttacttcattgttattcaccttccttgcttttatcatttaccttccagtgttgat
ggagttcagagagtatatacttagcccaagaagaatatacgaattaaccagaacagggtatggtatatacttctatccttcatt
aatgttttctcttgtcgcttctatttgcgcgttctttacatacaaaaaatcaaagttatttttgtatttccatagttttatttaactgtatact
tattttcttgcatggtaacaaaggaccaatatttagtatatttatagcattcatcctttacctttcatatattgaaaataaaaaaattaa
atttatgttcctggtaaaatcgtttgctgttatagcagtcattgtaacggcattctttgcatatacgtttactgatgggaatccgata

gaaaatatggcgaattactcggattatacccgtaatgctgttcttgttgcttcctcaaactttgactttatgtacggaaaattacta
atggaaagcgaggtttactcgaggattccgagggctatttggcctgataagcctgaagattttggcgcattgtatctggcaaa
agtatttttccctgatgcattctacagaaatcagggcgctcctgctttcgggtatggtgaactatacgcagatttcgggcttttta
caccagtttggttagttatatctggagtatttaaaggcgtcctagctaagtatttctccaataaaactcaggaaacaaagtcagc
gcattatttcataatgttcctattttgcattggaataagtgtgattcctgttagcatgggatggttgttccctgagcatttgatgattg
cttttatggtatacattgcatcttcctttgttttttcagagcatataagattcgttttactaagaaacaataaataaaaatgtatttttg
aatgatttaaatttctctagacgcgatgctggatttaaagcaagaaaagatgcactggacattgcttcagattatgaaaacattt
ctgttgttaacattcctctatggggtggagtagtccagagaattattagttctgttaagcttagtacatttctctgcggtcttgaaa
ataaagatgttttaattttcaatttcccgCAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCC
ACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTT
GAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATATCCACAGGACG
GGTGTGGTCGCCATGATCGCGTAGTCGATAGTGGCTCCAAGTAGCGAAGC
GAGCAGGACTGGGCGGCGGCCAAAGCGGTCGGACAGTGCTCCGAGAACG
GGTGCGCATAGAAATTGCATCAACGCATATAGCGCTAGCAGCACGCCATA
GTGACTGGCGATGCTGTCGGAATGGACGATATCCCGCAAGAGGCCCGGCA
GTACCGGCATAACCAAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCG
AGGATGACGATGAGCGCATTGTTAGATTTCATACACGGTGCCTGACTGCG
TTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTATCGATGATAAG
CTGTCAAACATGAGAA

Uppercase letters represent the pET9a vector backbone, lowercase letters represent the *wzy*ST gene embedding into the upstream and downstream homologous arms of *E. coli* W3110 *wzy* locus, and the lowercase letters with underline represent the *wzy*ST coding sequence.

2. The DNA sequences of pET9a-(*wzy*-*wzzB*)ST:

TTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATC
AGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAG
AAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCT
GCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCA
AAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGG
TGAGAATGGCAAAAGCTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCA
GCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCAT
TCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGAC
AATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGC
ATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGC
TGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTAC
GGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTT
AGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGT
TTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGT
CGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATC
AGCATCCATGTTGGAATTTAATCGCGGCCTCGAGCAAGACGTTTCCCGTTG
AATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTT
TATTGTTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGT
CAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGC

GCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTT
TGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTC
AGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGG
CCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAAT
CCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTT
GGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACG
GGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACT
GAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGG
AGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGC
GCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTC
GGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGG
GGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCT
GGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGAT
TCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGC
AGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGC
GCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCA
TATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCA
GTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACA
CCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCAT
CCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGG
TTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATC
AGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCA
GCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGG
CCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAG
GGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATG
CTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTG
TGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATC
ACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACA
GGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGG
GCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACC
ATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCAC
GTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGC
CAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGG
CCAGGACCCAACGCTGCCCGAGATGCGCCGCGTGCGGCTGCTGGAGATGG
CGGACGCGATGGATATGTTCTGCCAAGGGTTGGTTTGCGCATTCACAGTTC
TCCGCAAGAATTGATTGGCTCCAATTCTTGGAGTGGTGAATCCGTTAGCG
AGGTGCCGCCGGCTTCCATTCAGGTCGAGGTGGCCCGGCTCCATGCACCG
CGACGCAACGCGGGGAGGCAGACAAGGTATAGGGCGGCGCCTACAATCC
ATGCCAACCCGTTCCATGTGCTCGCCGAGGCGGCATAAATCGCCGTGACG
ATCAGCGGTCCAGTGATCGAAGTTAGGCTGGTAAGAGCCGCGAGCGATCC
TTGAAGCTGTCCCTGATGGTCGTCATCTACCTGCCTGGACAGCATGGCCTG
CAACGCGGGCATCCCGATGCCGCCGGAAGCGAGAAGAATCATAATGGGG
AAGGCCATCCAGCCTCGCGTCGCGAACGCCAGCAAGACGTAGCCCAGCGC
GTCGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGG

TGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAA
TACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCT
CGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATG
ATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCA
CCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGACGCT
CTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGG
CCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGC
CCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAA
GCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTC
GGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCA
CGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAATAC
GACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTT
AACTTTAAGAAGGAGATATACATccattagagtggaaagttggcgacgaaccctactatccagttaat
gataataaaaacatggagctttttaagaaatatagagagttagctagcagagaagacaaggttatatttggcgggcgtttggc
cgagtataaatattatgatatgcatcaagtgatatctgccgctctttatcaagtgaaaaatataatgagtacggattaatgcttat
aatttcatacattgcattatgcttattatttatagtttatctctatactctttccgtaagaattgaggggaaaataataaatgtaatggt
cccatacctgataataacagtccctacactgtacgtgtttgaaggtatatttgtatacctctcagaggtgcagaattatacagtg
gaatatttgtttttctatacttgctatataacatacatagcatcatttgttatttcttatctttacacacaaagaaaacccatatacaac
aaatcaaacacgaaaaataaaccaaggtatgtgtttacttcattgttattcaccttccttgcttttatcatttaccttccagtgttgat
ggagttcagagagtatatacttagcccaagaagaatatacgaattaaccagaacagggtatggtatatacttctatccttcatt
aatgttttctcttgtcgcttctatttgcgcgttctttacatacaaaaaatcaaagttatttgtatttccatagttttatttaactgtatact
tattttcttgcatggtaacaaaggaccaatatttagtatatttatagcattcatcctttacctttcatatattgaaaataaaaaaattaa
atttatgttcctggtaaaatcgtttgctgttatagcagtcattgtaacggcattctttgcatatacgtttactgatgggaatccgata
gaaaatatggcgaattactcggattataccgtaatgctgttcttgttgcttcctcaaactttgactttatgtacggaaaattacta
atggaaagcgaggtttactcgaggattccgagggctatttggcctgataagcctgaagattttggcgcattgtatctggcaaa
agtatttttccctgatgcattctacagaaatcagggcgctcctgctttcgggtatggtgaactatacgcagatttcgggctttta
caccagtttggttagttatatctggagtatttaaaggcgtcctagctaagtatttctccaataaaactcaggaaacaaagtcagc
gcattatttcataatgttcctattttgcattggaataagtgtgattcctgttagcatgggatggttgttccctgagcatttgatgattg
cttttatggtatacattgcatcttcctttgttttttcagagcatataagattcgttttactaagaaacaataaataataaggcaatcgg
ggctgatgagtcccgatgatttattgaccaaatggaaataatgtctgatttttatcattaatcctatggcatatatttgctttatggc
tacactgtctccagcttcatccttttttttagttagggtatctatgacagtggatagtaatacgtcttccgggcgtgggaacgatcc
ggaacagattgatttgattgagttattgctacagttatggcgtgggaagatgaccattattgtagccgttattatcgccattttgct
ggctgtaggctacctgatgatagccaaagaaaaatggacatccacggcgattattacccaacctgatgccgcgcaggttgc
cacctataccaacgcgctcaacgtcttgtatggtgggaatgcccccaaaatctctgaagtgcaggcgaattttataagccgct
ttagctctgcgttttccgcattatcggaagtgctggataatcagaaagagcgggaaaagcttaccattgaacagtcggtaaaa
gggcaggcgctgccactctcggtttcttatgtgagtactaccgctgaaggggcgcagcgtcggctggcggaatatatccaa
caggtggatgaagaggtcgctaaggaactggaagttgacctgaaagataacatcacgctgcaaaccaaaacgttgcagg
agtcccttgagacgcaggaagttgtggcgcaggagcaaaaagatctgcgtattaagcaaatcgaagaagcgttgcgctat
gcggatgaggccaaaatcacgcagccgcagattcagcaaacccaggatgttacccaggacacgatgttcctgttggggga
gcgatgcgctaaaatcgatgatacagaacgaagcgacgcgtccactggtcttttctccggcctattaccagacgaagcaga
cactgctggacattaaaaatctgaaagtgactgccgatacggtgcacgtctatcgttatgtgatgaagccgacgctgccggt
ccgtcgcgatagcccgaaaacagccattacccttgtgctggctgtattgctgggtgggatgatcggtgccgggattgtgctg
gacgcaatgcgctacgtagctataagccaaaagccttgtaaaaatgtatttttgaatgatttaaatttctctagacgcgatgc
tggatttaaagcaagaaaagatgcactggacattgcttcagattatgaaaacatttctgttgttaacattcctctatggggtgga

gtagtccagagaattattagttctgttaagcttagtacatttctctgcggtcttgaaaataaagatgttttaattttcaatttcccgC
AAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACT
AGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAA
AGGAGGAACTATATCCGGATATCCACAGGACGGGTGTGGTCGCCATGATC
GCGTAGTCGATAGTGGCTCCAAGTAGCGAAGCGAGCAGGACTGGGCGGC
GGCCAAAGCGGTCGGACAGTGCTCCGAGAACGGGTGCGCATAGAAATTG
CATCAACGCATATAGCGCTAGCAGCACGCCATAGTGACTGGCGATGCTGT
CGGAATGGACGATATCCCGCAAGAGGCCCGGCAGTACCGGCATAACCAA
GCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAGCG
CATTGTTAGATTTCATACACGGTGCCTGACTGCGTTAGCAATTTAACTGTG
ATAAACTACCGCATTAAAGCTTATCGATGATAAGCTGTCAAACATGAGAA

Uppercase letters represent the pET9a vector backbone, lowercase letters represent the (*wzy-wzzB*)ST genes embedding into the upstream and downstream homologous arms of *E. coli* W3110 *wzy* locus, and the lowercase letters with underline represent the (*wzy-wzzB*)ST coding sequence.