Construction of a Novel Recombinant *Escherichia coli* Strain Capable of Simultaneously Utilizing Glucose and Xylose

Seong Hun Jeong

Department of Biomedical Engineering
School of Nano-Bioscience and Chemical Engineering
Graduate School of UNIST

2013

Construction of a Novel Recombinant *Escherichia coli* Strain Capable of Simultaneously Utilizing Glucose and Xylose

Seong Hun Jeong

Department of Biomedical Engineering
School of Nano-Bioscience and Chemical Engineering
Graduate School of UNIST

Construction of a Novel Recombinant *Escherichia coli* Strain Capable of Simultaneously Utilizing Glucose and Xylose

A thesis
Submitted to the School of Nano-Bioscience and Chemical Engineering and the Graduate School of UNIST in fulfillment for the degree of
Master of Science

Seong Hun Jeong

11. 11. 2013

Approved by

Major Advisor
Sung Kuk Lee, Ph. D.

Construction of a Novel Recombinant *Escherichia coli* Strain Capable of Simultaneously Utilizing Glucose and Xylose

Seong Hun Jeong

This certifies that the thesis of Seong Hun Jeong is approved.

11. 11. 2013

signature

Thesis Supervisor: Sung Kuk Lee, Ph. D.

signature

Choel Min Ghim, Ph. D.

signature

Robert J. Mitchell, Ph. D.

Abstract

Lignocellulose-derived carbon sources are attractive to produce valuable biofuels and biochemicals. Despite many approaches have been intensively attempted for the production of valuable bioproducts in *Escherichia coli*, there still problems remaining such as carbon catabolite repression (CCR). Since non-engineered *E. coli* cannot simultaneously utilize glucose and xylose, it leads to longer fermentation times and lower productivities. Recently, we have found the enhanced glucose-xylose co-fermenting strains by using xylose metabolic operon with constitutive promoters, in order to exclude CCR in *E. coli*. Further, we performed adaptive evolution to strengthen the xylose metabolism mediated by the synthetic promoters. Analysis of whole genome re-sequencing of the adapted *E. coli* strains showed that four mutations were found in each adapted strain resulting in metabolic changes. Consequently we confirmed the significance of these mutations for xylose metabolism by using precise genome engineering tools. This study would be helpful to design an efficient platform strain for valuable bioproducts from lignocellulosic biomass containing large amount of glucose and xylose.

Contents

List of figures

List of tables

Nomenclature

CCR: Carbon catabolite repression

E. coli: *Escherichia coli*

PTS: Phosphotransferase system

HPr: Histidine protein

PEP: Phosphoenolpyruvate

cAMP: Cyclic adenosine monophosphate

CRP: cAMP receptor protein

PPP: Pentose phosphate pathway

CP: Constitutive promoter

MAGE: Multiplex automated genome engineering

OD: Optical density

PCR: Polymerase chain reaction

SOE-PCR: Splicing by overlapping extension PCR

1. Introduction

In the modern world, interest in alternative energy source is increasing due to the major concerns on global warming and depletion of fossil fuels[1]. Despite the development of alternative energy sources their supply is limited and hence conventional fossil fuels still remain as essential element to maintain the existing traditional transportations and other key industries. To overcome these problems, the conversion of lignocellulosic biomass to biofuels and valuable products using microorganisms has been intensively studied nowadays[2]. The advantages of lignocellulosic biofuels are 1) low cost of substrate, 2) eco-friendly, and 3) sustainable supply of biomass[3, 4]. However, lignocellulosic biomass has several problems to be overcome.

1.1 Lignocellulosic biomass

The ideal substrate for the production of bulk commodity should satisfy the following advantages including low production and preprocessing costs, consistent supply, and not adversely affect the environment[4]. In this aspect, lignocellulosic biomass is the most ideal because lignocellulosic biomass is very cheap and most abundant substrate in the world. Moreover, this biomass does not affect global food supply that was the major cause of the decline of sugar cane and corn based biofuels. These benefits suggest that lignocellulosic biomass is attractive substrate to produce biofuels or other valuable materials[1].

The lignocellulosic biomass is composed of three main components accounted for 45% of cellulose, 30% of hemicellulose, and 25% of lignin[5]. The lignin is composed of polyphenolic compounds, so it functions to maintain the structure and waterproofing in plant. However, lignin and its byproducts are toxic to most microbes. The cellulose is a polymer of D-glucose by β-1,4-linkage. Due to its crystalline structure, cellulose is generally water-insoluble and highly resistant to hydrolysis. The hemicellulose is heteropolymer constituent of the plant cell wall including several sugars such as mannose, galactose, rhamnose, arabinose, and xylose[6, 7] and physically weaker than cellulose and lignin because it has amorphous structure. Therefore, hemicellulose is well hydrolyzed by acid and base than cellulose. Most predominant sugar in hemicellulose is D-xylose from xylan, commonly comprising approximately 1/3 of the total lignocellulosic biomass[8].

1.2 Metabolism of glucose, arabinose and xylose in *E. coli*

The microorganism *Escherichia coli* uses phosphotransferase system (PTS) energized by high energy phosphate group from phosphoenolpyruvate (PEP) for uptake of glucose into the cell. Eventually glucose is transported inside the *E. coli* cell in the form of glucose-6-phosphate and can directly enter glycolysis pathway for metabolism[9].

Arabinose and xylose are most abundant pentose sugars in hemicellulose[10]. Arabinose is transported into the cell by *araFGH* and *araE* transporters. Both *araFGH* and *araE* operons are induced by arabinose. The AraE is a low-affinity arabinose transporter, and uses proton motive force as the energy. On the other hand, a high affinity arabinose transporter encoded by the *araFGH* uses ATP as an energy source. The genes for arabinose catabolism are coded in *araBAD* operon. The *araA* encodes L-arabinose isomerase; *araB* encodes L-ribulokinase; *araD* encodes L-ribulose-5-phosphate-4-epimerase[11]. The expression of these genes is activated by cAMP-CRP complex and arabinose-bound AraC. When the arabinose is present in the cell, transcriptional activator AraC activates *araBAD*, *araFGH*, and *araE* transcription for up-regulation.

The xylose is transported into the cell by both transporters encoded by *xylFGH* and *xylE*[12]. The XylE is a low affinity xylose transporter, and uses proton motive force as the energy for the transport of xylose into the cell. On the other hand, XylFGH is a high affinity xylose transporter using the energy of ATP[12, 13]. Metabolism of xylose is mediated by *xylA and xylB* genes encoding D-xylose isomerase and xylulokinase, respectively[14]. When D-xylose binds to XylR and the cAMP concentration in cell is high, XylR increase the transcription of *xylAB* operon[15, 16]. Both xylose and arabinose are converted to D-xylulose 5-phosphate finally, and then enter the pentose phosphate pathway (PPP) (Fig. 1)[17]. Interestingly, xylose catabolic operon transcription is repressed by arabinose-AraC complex[11].

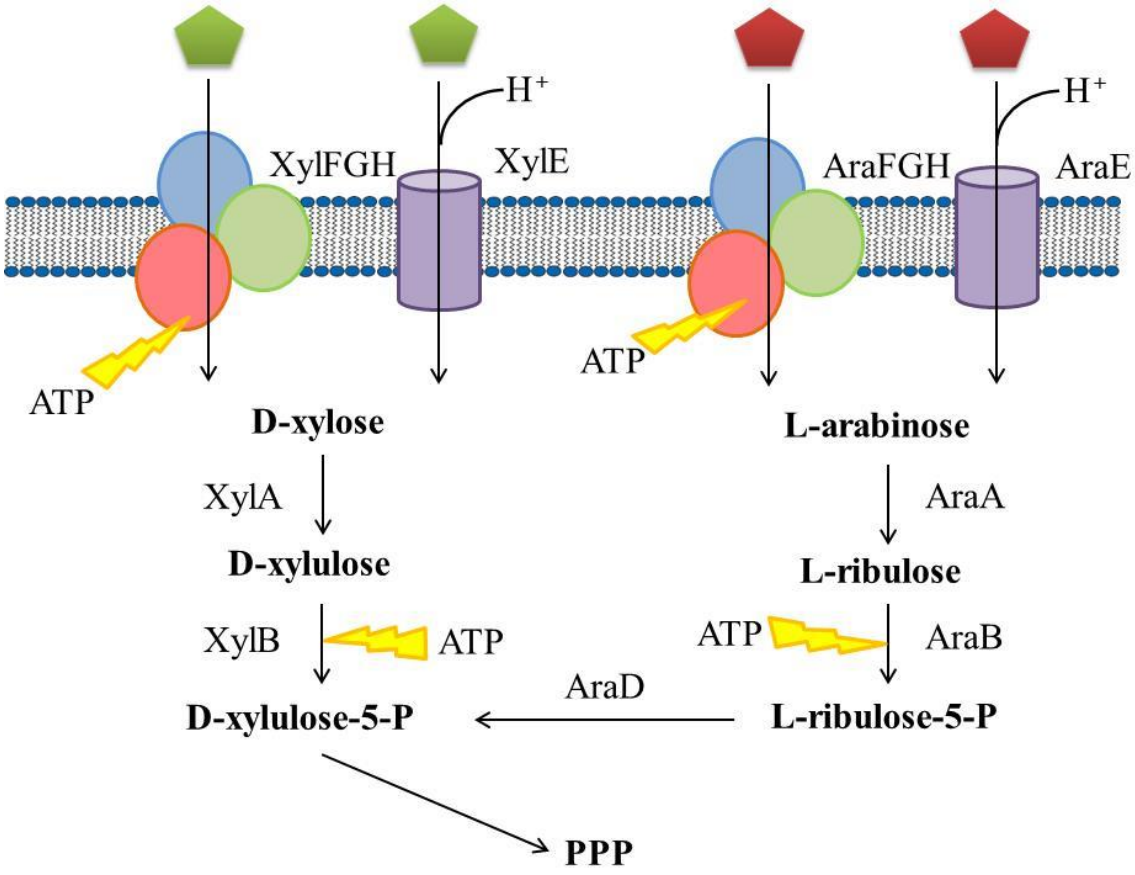

Figure 1. A schematic of arabinose and xylose uptake and catabolism in *E. coli*

PPP (pentose phosphate pathway), XylA (xylose isomerase), XylB (xylulokinase), XylFGH (xylose specific transporter), XylE (low affinity xylose transporter), AraA (L-arabinose isomerase), AraB (L-ribulokinase), AraFGH (arabinose specific transporter), AraE (low affinity arabinose transporter)

1.3 Carbon catabolite repression in *E. coli*

E. coli can utilize several sugars like glucose, arabinose, galactose, mannose and xylose. But if theses sugars are present in a mixture only glucose will be preferentially utilized[18]. Such a phenomenon is termed as carbon catabolite repression (CCR)[19]. When *E. coli* is growing in a media containing glucose, the expression of catabolic genes and transporter genes of xylose and arabinose remains low. This repression is achieved by PTS system, cAMP, CRP (cAMP binding protein), and other components acting cooperatively. By CCR, glucose effectively controls carbohydrate uptake and metabolism[19].

PEP (phosphoenolpyruvate) dependent glucose PTS system acts as glucose transportation and produces glucose-6-phosphate through a cascade of reactions. First step is that E1 enzyme catalyzes conversion PEP to pyruvate and transfers the resulting phosphate group to Hpr protein. Second step is that phosphorylated Hpr transfers the phosphate group to histidine residue of $EIIA^{Glu}$ domain. Attached phosphate group at $EIIA^{Glu}$ is transferred to EIIB and phosphate group is attached to glucose with transportation of glucose into the cell (Fig. 2)[20].

The CCR regulation of *E. coli* is caused by alteration in the level of phosphorylated $EIIA^{Glu}$. $EIIA^{Glu}$ is readily dephosphorylated when cell is in the exponential phase in glucose containing media. But when glucose is depleted in the media, $EIIA^{Glu}$ remains phosphorylated and activates adenylate cyclase on inside of inner membrane[21, 22]. The ATP is converted to cAMP by activated adenylase cyclase. Produced cAMP binds to the receptor protein, CRP. After that, the cAMP-CRP complex can activate several operons relating to sugar catabolism and transport such as arabinose and xylose[23, 24]. In summary, PTS system blocks cAMP-CRP complex production in glucose media, so that the expression of transporters and metabolic genes of other sugars is inhibited[18]. As a result, *E. coli* cannot utilize pentose sugars with glucose simultaneously. The mechanism and effects of CCR in sugar metabolism is extensively studied[23-25]. However, we cannot fully understand the process of regulation yet, and most suitable technologies for sugar co-utilization are not developed.

1.4 Co-metabolism of glucose and pentose sugars in *E. coli*

Glucose, arabinose, and xylose are most abundant sugars of lignocellulosic biomass. When these three sugars are present together, wild-type *E. coli* utilizes glucose first followed by arabinose, and finally xylose. Even worse case of sequential utilization is that there is a lag phase before beginning of the utilization of second sugar, ultimately leading to inefficient utilization of sugar mixture. Several studies have described the presence of hierarchy among glucose, arabinose, and xylose utilization and possible ways to remove the hierarchy. Despite several decades of research, the hierarchy of glucose and xylose is not yet fully understood.

Previous researches have concentrated on *ptsG* knockout strategy (glucose-specific PTS transporter) or inactivation of other sugar related transcription factor elements like *crp*, *mlc*, *mgsA* to reduce the effect of CCR[26, 27]. The *ptsG* mutation that commonly introduced to many *E. coli* strains has been investigated intensively and successful results are obtained. Although glucose uptake rate was lower than wild type, CCR was partially removed in *ptsG* deleted strain[28]. For example, Nichols et al reported that *ptsG* deleted strain showed partial co-metabolism of arabinose and xylose in presence of glucose[28]. The total time to ferment sugar mixture completely was also dramatically reduced in rich media[29]. On the other hands, the *E. coli* strain with *ptsH*, *ptsI*, and *crr* genes being deleted showed insufficient PTS causing impaired growth using both PTS sugars and non-PTS sugars[30]. The ability to utilize glucose in strains with impaired PTS was achieved using several methods such as evolving in glucose media, activation of GalP, and overexpression of glucokinase[31]. Resulting strains showed co-utilization of glucose with pentose sugars. These results indicate that inactivation of PTS system leads to relief of CCR. However, similar results were not achieved in mineral salt media which especially demanded for industrial process[32].

The *E. coli* strains containing constitutively active cAMP independent CRP mutation also show co-utilization of glucose and xylose[12]. As a global regulator, CRP controls expression of approximately 200 promoters and constitutively active CRP might lead to changes in cellular physiology in large scale[33]. However, glucose effect was not eliminated with constitutively active CRP when glucose concentration was very high. The significance of these physiological alterations is not widely known. Finding new approach to engineer co-utilization of sugars is more efficient way than to analyze whole possible physiological changes of the strain due to mutant CRP

Previous studies induced co-metabolism of various sugars using the major glucose transport system and global regulatory system. However, these approaches cause other problems also such as impaired glucose metabolism by *ptsG* inactivation that inhibits efficient glucose and xylose co-utilization[27]. Moreover, CRP mutation strains could not overcome glucose effect. Most of all, up to now, many studies ignored arabinose effect in co-metabolism of glucose and xylose[11].

We tried to solve the problem of simultaneous fermentation problem between the glucose and xylose. In this work, to increase co-metabolism between glucose and major two pentose sugars and to avoid the problems of previous studies, 1) native promoters are exchanged with synthetic constitutive promoter for sugar metabolic operon (total 5 promoters)[34], 2) *araC* was deleted to reduce down regulation and to avoid regulatory interference for improvement of xylose metabolism[11] 3) evolutionary adaptation was carried out for more improved sugar co-utilizing strain[35]. 4) Next generation sequencing and precise genome engineering were used to identify critical mutations of the evolution. This technology will reduce the inefficient substrate utilization of *E. coli* and would help to increase the productivity[36].

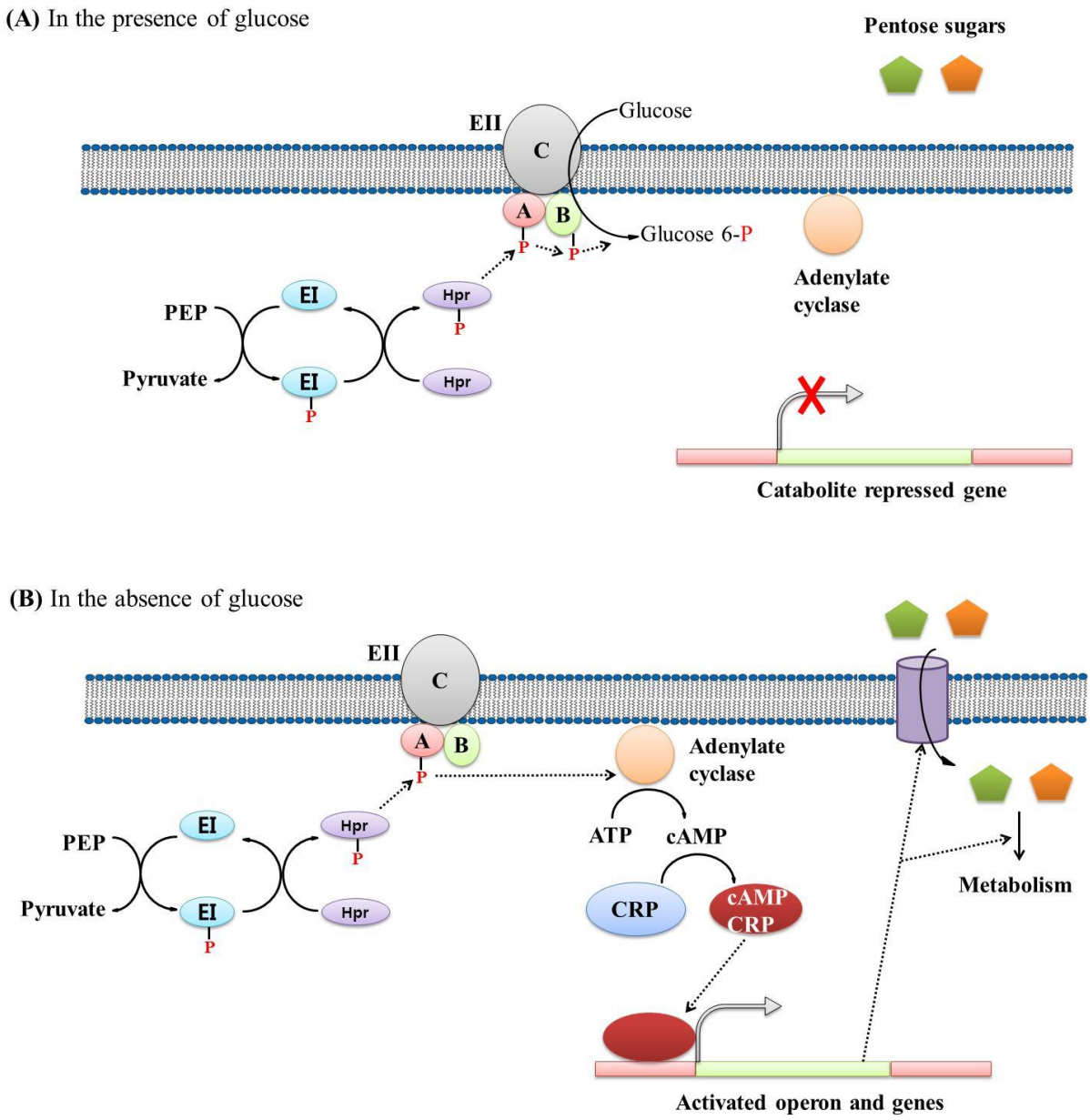

Figure 2. Glucose-mediated carbon catabolite repression in *E. coli*

Enzyme I (EI), Enzyme IIA (EIIA), Enzyme IIB (EIIB), Enzyme IIC (EIIC) and Histidine protein (Hpr). The phosphate group was transferred from PEP (phosphoenolpyruvate) to the incoming glucose by these five domains. (A) In the presence of glucose, phosphate group is attached to glucose, so adenylate cyclase cannot activate. (B) In the absence of glucose, phosphorylated EIIA activates adenylate cyclase and signal cascade is activated. As a result, transporter and metabolic genes of pentose sugars are expressed[36, 37].

2. Materials and methods

2.1 Bacterial strains and media

The bacterial strains used in this study are listed in Table 1. *E. coli* MG1655 was used as a parental strain. Luria-Bertani (LB) medium containing suitable antibiotics was used for general cell growth (50 μg/mL of kanamycin, 100 μg/mL ampicillin, or 30 μg/mL chloramphenicol). Strains possessing temperature sensitive plasmids were grown at 30℃. The others were grown at 37℃. Engineered strains were cultured in M9 minimal medium with appropriate sugars and antibiotics for cell growth.

Table 1. *E. coli* strains and plasmids used in this study

Strains/plasmids	Genotype/description	Reference
Strains		
E. coli MG1655	Wild type	[38]
AXcp	MG1655 with P_{CP25}-*araB*, P_{CP6}-*araF*, P_{CP6}-*araE*, P_{CP25}-*xylA*, P_{CP6}-*xylF*, *ΔaraC*::FRT	In this study
AXcpA50	AXcp adapted for 50days in arabinose minimal media	In this study
AXcpX50	AXcp adapted for 50days in xylose minimal media	In this study
AXcpAX50	AXcp adapted for 50days in arabinose/xylose minimal media	In this study
AXcpM	AXcp with *mutS*::pRED2	In this study
Plasmids		
pSIM5	pSC101 *ori.* temperature sensitive replication *exo, bet, gam* gene - λ-RED recombinase expression is induced by temperature	[39]
pCP20	Yeast FLP recombinase gene controlled by *c*I repressor and temperature sensitive replication	[40]
pKD13	Template of FRT-flanked kanamycin resistant gene. R6K gamma *ori.* Requiring the *pir* $^{+}$ *E. coli*	[41]
pKD46	Temperature sensitive replication *exo, bet, gam* gene - λ RED recombinase expression is induced by arabinose	[41]
pRED2	R6K gamma *ori.* Requiring the *pir* $^{+}$ *E. coli* *mutS* homologous region with *exo, bet, gam* gene - λ-RED recombinase expression is induced by temperature	In this study

2.2 Strain construction

λ RED recombination system of pSIM5 plasmid was used to for several genes deletion and promoter replacement[39]. Chromosomal exchanges were identified by PCR analysis. Primers used to construct strains are listed in Table 2, 3, and 4[42]. Two synthetic constitutive promoters, CP6 and CP25, were used for promoter exchange[43].

AXcp strain was constructed by replacing five promoters of arabinose and xylose operons and deletion of *araC*. The arabinose and xylose catabolic genes of *araBAD* and *xylAF* promoters were replaced with CP25 synthetic promoter. And transporter gene promoters such as *araFGH*, *araE*, and *xylFGH* were replaced with CP6 synthetic promoter (Table 1).

The engineered strains were adapted in minimal medium with 4 g/L arabinose, 4 g/L xylose, and 2 g/L arabinose and 2 g/L xylose, respectively. When cells reached to OD_{600} of 1.0, they were subcultured into the fresh medium, allowing for selection for a faster growing subpopulation [35]. After adaptation of 50 days, the predominant strains are designated AXcpA50, AXcpX50, and AXcpAX50 subcultured in a medium containing arabinose, xylose, and arabinose and xylose, respectively.

To construct AXcpM strain for genome engineering using MAGE method, pRED2 plasmid (not published) carrying λ RED recombination system was integrated into *mutS* site of AXcp strain[39]. Strains carrying the pRED2 plasmid in *mutS* site were selected by using LB chloramphenicol plate and PCR confirmation of genomic DNA from the integrated strain. The primers used for strain validation are listed in Table 4.

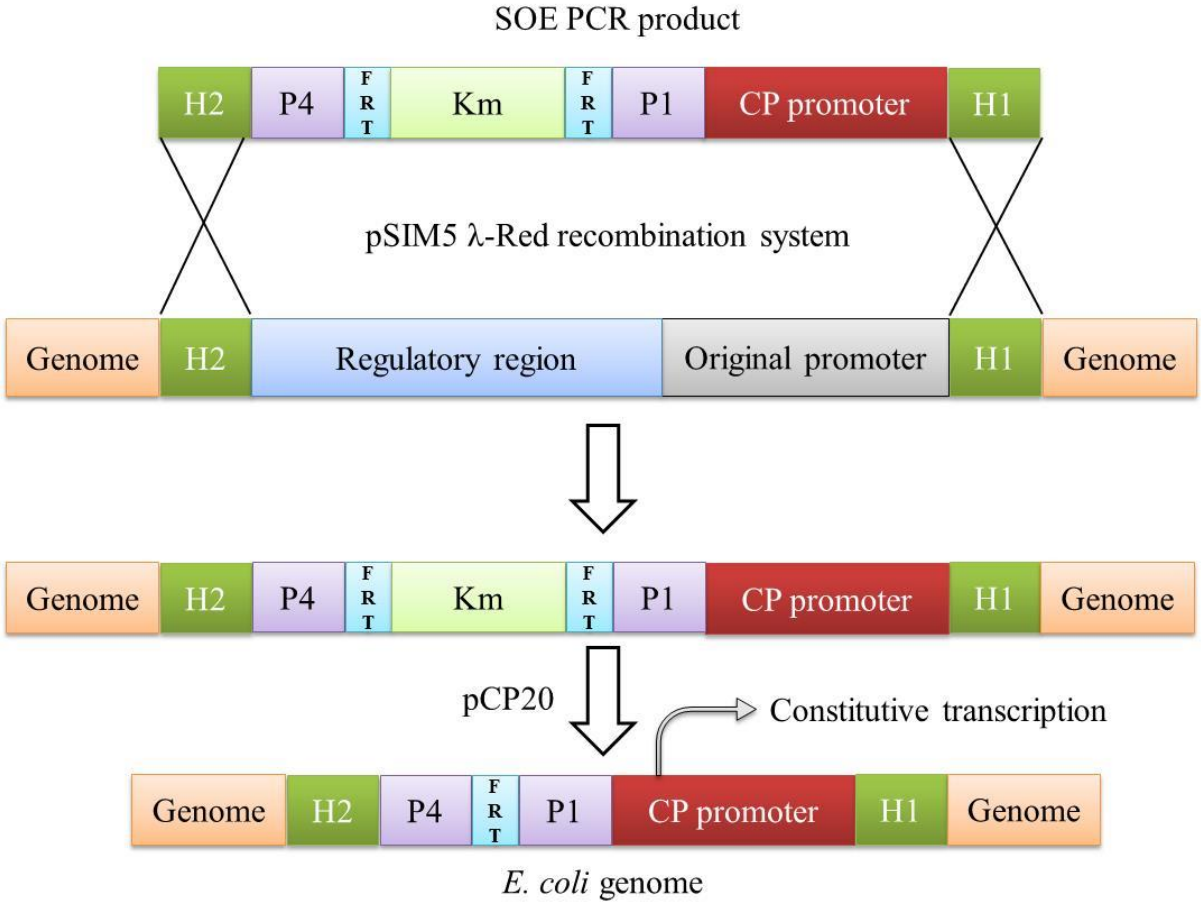

Figure 3. The strategy for exchange from native promoter to synthetic constitutive promoter

SOE PCR products are transformed to wild type MG1655 strain[42]. After recombination and selection, kanamycin cassette is deleted by using Flp recombinase from pCP20 plasmid for further promoter exchange.

Table 2. Primers for SOE-PCR in this study

	Primer sequences	References
CP6 and Kanamycin cassette	5′ -CATAGCTGTTTCCTGTGTGAACAGTACTCAGTTATTATATCATCCGG-3′ 5′ -TCAGTTATTATATCATCCGGAAATATCTGTGTCAAGAATAAACTCC-3′ 5′ -CTGTGTCAAGAATAAACTCCCACATGATTCCGGGGATCCGTCGACC-3′	[44]
CP25 and Kanamycin cassette	5′ -CATAGCTGTTTCCTGTGTGAACAGTACTATGTGATTATACCAGCCCCC-3′ 5′ -ATGTGATTATACCAGCCCCCTCACTACATGTCAAGAATAAACTGC-3′ 5′ -ACATGTCAAGAATAAACTGCCAAAGATTCCGGGGATCCGTCGACC-3′	[44]

Table 3. Primers for promoter exchange and gene deletion in this study

	Primer sequences	References
P_{CP25}-*araB, ΔaraC*	5′ -CCCTATGCTACTCCGTCAAGCCGTCAATTGTCTGATTCGTTACCAA GTGTAGGCTGGAGCTGCTTCG-3′ 5′ -TCGCACAGAATCACTGCCAAAATCGAGGCCAATTGCAATCGCCAT AGCTGTTTCCTGTGTGAAC-3′	[44]
P_{CP6}-*araF*	5′ -GGTAATGCGGCCTATTGACTGGTTAAAAAGAAGACATCCCGCATG GGTAGTGTAGGCTGGAGCTGCTTCG-3′ 5′ -GACATAACGGCTGCCAGACCAATGGCTGCCAGGGCTTTAGTAAA TTTGTGCATAGCTGTTTCCTGTGTGAACAGTACT-3′	[44]
P_{CP6}-*araE*	5′ -TATCTGCTGTAAAATTAGGTGGTTAATAATAATCTCAATAATTCAAC GTGTAGGCTGGAGCTGCTTCG-3′ 5′ -CCCGCAAAGAACGTGGCGTTAAAGCAGATTCCGTATTGATAGTA ACCATAGCTGTTTCCTGTGTGAACAGTACT-3′	In this study
P_{CP25}-*xylA*	5′ -TGAGCCTTCATAACGAACGCGATCGAGCTGGTCAAAATAGGCTTG CATAGCTGTTTCCTGTGTGAA-3′ 5′ -TTGTTGCGCAATTGTACTTATTGCATTTTTCTCTTCGAGGAATTACC CAGTTTCATCAATTCCGGGGATCCGTCGACC-3′	In this study
P_{CP6}-*xylF*	5′ -TGATGAAACTGGGTAATTCCTCGAAGAGAAAAATGCAATAAGTAC AATTGCGCAACAAGTGTAGGCTGGAGCTGCTTCG-3′ 5′ -AAGCAGGAGTGAGGTGCAAAGGGTGAGTAGAATGTTCTTTATTTT CATAGCTGTTTCCTGTGTGAAC-3′	In this study

Table 4. Sequencing primers for confirmation of genomic variations

	Primer sequences	Confirmation
araE-F	5′ -CCGAGCATCATGCTACTAACC-3′	Promoter exchange
araE-R	5′ -GGCGCGTTAATTCATTCTTCT-3′	
araF-F	5′ -GCTCTCATTATACGTGTTCTG-3′	Promoter exchange and MAGE
araF-R	5′ -CCTCAAACCCTAAATCCTTCC-3′	
araC-F	5′ -AACTGGTTATTCGGGGCATC-3′	Promoter exchange
araC-R	5′ -AGGCGTGCCAGAAACTTAA-3′	
xylA-F	5′ -AACTCAAATGCGACATCTGC-3′	Promoter exchange and MAGE
xylA-R	5′ -ATGCCTTCTTGTTTGGCTTC-3′	
araE SNP-F	5′ -GCTATAACTGAACGCTGTATC-3′	MAGE
araE SNP-R	5′ -CTGCTTTAACGCCACGTTCT-3′	
ybjG-F	5′ -CCACGATTGCAGACGTTGAT-3′	MAGE
ybjG-R	5′ -CGCCAGACTCGGCTCCGTGG-3′	
thiC-F	5′ -TAAATGCGTTTTGAGTTGGG-3′	MAGE
thiC-R	5′ -ATGGATTACTACGATTCCAG-3′	
pyrE-F	5′ -GGGCCAAACAGCAGATCGAAC-3′	MAGE
pyrE-R	5′ -GTCGGAATTGTGAACGGCGA-3′	
MutS A1R	5′ -CTCTCATCCGCCAAAACAGCCCAT AACCCATGAGTGCAATAG-3′	AXcpM construction
Cm R	5′ -CCGTTTTCACCATGGGCAAATATTATACG-3′	
MutS A5F	5′ -GTATAATCACATAGTACTGTTTTACA CCAGGCTCTTCAAGCGATA-3′	AXcpM construction
pSIM5 UP	5′ -CAGTGCGTCCTGCTGATGTGC-3′	

2.3 Batch culture

For seed culture, 50 µL of cells grown overnight in LB were inoculated (1:100) into 5 mL of M9 media supplemented with suitable concentration of sugars such as glucose, xylose, and arabinose in a test tube and grown at 37℃ in the shaking incubator. After 12 hours of seed culture, 500 µL grown cells were inoculated (1:100) into fresh 50 mL of M9 media supplemented with the same concentration of sugar as in seed culture in 250 mL shaking flasks and grown at 37℃ in the shaking incubator. Cell density and residual sugar concentration was regularly measured using Biochrom Libra S22 spectrophotometer at OD_{600} and high-pressure liquid chromatography (HPLC), respectively.

2.4 HPLC analysis

The total amount of remaining sugars in the medium during the bacterial growth period was regularly measured using a Shimadzu HPLC station, equipped with a refractive index detector (Shimadzu), a SIL-20A auto-sampler (Shimadzu), and an HPX-87P column (Bio-Rad). The column was operated at 0.6 mL/min with HPLC grade water as the mobile phase at 80℃ of oven temperature. 1 mL of culture medium was collected every 2 or 4 hours and centrifuged to remove cells. The supernatant were collected and heated at 80℃ for 1hour, and then the samples were centrifuged again for 30 minutes to remove proteins. 20 µL of appropriately diluted and processed samples was injected in the column and analyzing time was 30 minutes.

2.5 Whole genome re-sequencing

Whole genome re-sequencing was conducted by Next generation sequencing (Macrogen Inc, Korea) for AXcpX50 and AXcpAX50 strains using AXcp as the standard strain. All reported mutation regions were PCR amplified and verified by DNA sequencing analysis.

2.6 Multiplex automated genome engineering (MAGE)

For effective and fast introduction of point and deletion mutations, MAGE method was used[45-47]. AXcpM strain was grown overnight at 30℃ in LB supplemented with chloramphenicol. And then, seed was inoculated using 1/100 dilution into 3 mL modified LB - Lennox and grown till the OD_{600} reached to 0.5~0.6 at 30℃. Cells were induced at 42℃ water bath for 15 minutes to express λ-recombination system. After induction, cells were cooled on ice for 1 hour to prepare electrocompetent. After cooling, approximately 1 mL ~ 2 mL cells were centrifuged in pre-cooled centrifuge at 4℃ and then supernatant were removed. Cell pellet was washed using 4℃ chilled autoclaved water, and then spin down and removed supernatant. Using same procedure, repeat at least 3 times for washing the cell pellet. After washing step, the cell pellet was resuspended in 50 μL RNAse free water containing each seven MAGE primers diluted up to 0.5 μM concentration, respectively. Re-suspended cells were transferred to electrocuvette and transformed via electroporation. Transformed cells were recovered in 3 mL modified LB supplemented with appropriate chloramphenicol in test tube till the OD_{600} reaches 0.5~0.6 at 30℃. The whole procedure was repeated 4 times. After final round, recovery was carried out for over 10 hours. After recovery, cells were properly diluted (~10^3 cells) and then spread in normal LB chloramphenicol plate or M9 minimal media plate supplemented with 4 g/L xylose and appropriate chloramphenicol. Colonies from M9-xylose plates were used for mutation validation using their genomic DNA.

Table 5. MAGE primers for introduction of mutations

Target region	Primer sequences
ybjG (882316) †	5′-AAAATATCGGCTATAACTTCCTGCATCATGCGGCGGATGGCTCATTCCCAAGCGATCACGGTACGGTGATTTTCACCTTT-3′
araF (1984108~1984122) †	5′-CGAGCTTCAGGTTCTCCGCCATAGCGGATTGTGACATAACGGCTGCCAGGGCTTTAGTAAATTTGTGCATAGCTGTTTCC-3′
araE in AXcpAX50 (2979933) †	5′-TATAGAACCGAGTACAAACAGGATGGCCCCCGCCATCAGGATGTATTTACGCCCCAGGCGGAACGACAGCCAACCATTAA-3′
araE in AX cpX50 (2979938) †	5′-GCCTTCGCTCCTCATCTTACTTTTCTACAGACAAAAAAAAAGCGACTCATCAGTCGCCTTAAAAATCAGTTTGCCAGCGCC-3′
pyrE upstream (3813833) †	5′ -AACCGAGTACAAACAGGATGGCCCCCGCCATCAGGCTGTATTACGCCCCAGGCGGAACGACAGCCAACCATTAAACAGC-3′
thiC upstream (4194272) †	5′-GCGGGGCTAATTTCTTGTCGGAGTGCCTTAACTGGCTGATACCGTTTATTCGGGATCCGCGGAACCTGATCAGGCTAATA-3′
CP25 *xylA* (57th base of CP25)	5′-GCTGGTCAAAATAGGCTTGCATAGCTGTTTCCTGTGTGAAAAGTACTATGTGATTATACCAGCCCCCTCACTACATGTCA-3′

† Base number in NCBI reference sequence NC_000913.2

3. Results

3.1 Adaptive evolution of *E. coli* strain AXcp

To improve arabinose or xylose co-utilization with glucose, 5 promoters of *ara* and *xyl* operons were exchanged with synthetic constitutive promoters and *araC* was deleted, designated AXcp strain (Table 1). To investigate sugar utilization ability of AXcp strain, the cells were cultured in M9 minimal media supplemented with 4 g/L glucose, 4 g/L xylose, and 4 g/L arabinose, respectively. In 4 g/L glucose M9 minimal media, AXcp showed a cell growth similar to wild-type strain (Fig. 4A). Unfortunately, AXcp showed no cell growth in 4 g/L arabinose and 4 g/L xylose as a sole carbon source during 16 hours (Fig. 4B and C).

To restore the cell growth of AXcp strain in arabinose and xylose M9 minimal media, adaptive evolution was carried out using fed-batch culture in 4 g/L arabinose, 4 g/L xylose, and 2 g/L arabinose and 2 g/L xylose, respectively, during 50 days. The resulting strains were named as AXcpA50 for adaptation on M9-arabinose medium, AXcpX50 for adaptation on M9-xylose medium, and AXcpAX50 for adaptation on M9- arabinose/xylose medium.

3.2 Cell growth of engineered strains

The difference of cell growth after 50 days adaptation was investigated by comparing grow rate between un-adapted strain AXcp and three adapted strains in M9 minimal media supplemented with 4 g/L glucose, 4 g/L arabinose, or 4 g/L xylose. Samples were collected every two hours for analysis of the sugar concentration and OD_{600}. Glucose utilization rates were not impaired in these strains (Fig. 4A). AXcpA50 strain showed 15% lower OD_{600} value than wild type in M9 minimal media with 4 g/L glucose after cultivation of 16 hours. Cell growth rate of AXcpAX50 strain is similar to wild type strain. Interestingly, AXcpX50 strain exhibited slightly higher OD_{600} value than the wild-type after 16 hours of cultivation (Fig. 4A).

The AXcp, AXcpX50, and AXcpAX50 strain showed almost no cell growth during 16 hours in M9 minimal media with 4 g/L arabinose. Interestingly, AXcpA50 strain grew well like wild-type strain by 8 hours, and then cell stopped growing. This might indicate that *araC* might have additional roles except for the activation of *ara* operon (Fig. 4B). Like AXcp strain, the AXcpA50 strain showed almost no cell growth in M9 minimal media with 4 g/L xylose and another arabinose adapted strain AXcpAX50 could not show improved growth on arabinose medium also. Both AXcpX50 and

AXcpAX50 strains showed improved cell growth in xylose medium than wild type cell. In particular, AXcpAX50 strain showed most fast growth by initial 10 hours. AXcpX50 strain showed best biomass formation after 16 hour growth (Fig. 4C).

Interestingly, the strain adapted in arabinose-xylose medium could not grow efficiently on arabinose but can grow even better than wild type cells in xylose minimal medium. This might be due to the initial difference in the ability to grow on xylose or arabinose by the strain, AXcp. It could also be because of the regulatory proteins like *araC* and *xylR*[15, 16]. While AXcp had *araC* deleted, *xylR* was still present in this strain. This result might indicate the need for the operon-specific regulators on sugar metabolism that operates beyond the promoters. Yet another insight on this regulation is despite being pentose sugars arabinose and xylose may have different rate-limiting steps in their metabolism.

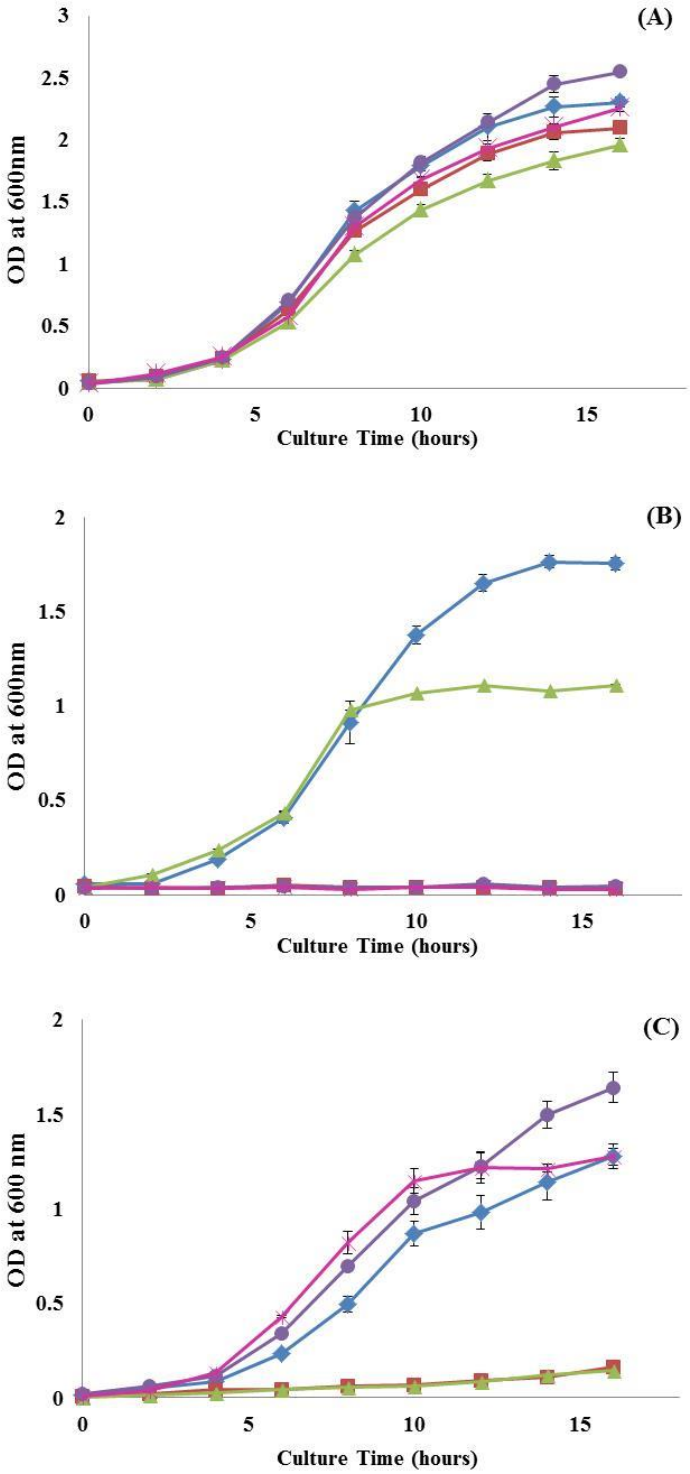

Figure 4. Cell growth pattern in 4 g/L glucose (A), 4 g/L arabinose (B), and 4 g/L xylose (C) supplemented minimal media. Wild type (diamond), AXcp (square), AXcpA50 (triangle), AXcpX50 (circle), and AXcpAX50 (star) were used. Error bars indicate the standard deviation of experiments performed in triplicate.

3.3 Co-metabolism of glucose and xylose by AXcp strains

The restoration of xylose metabolism in AXcpX50 and AXcpAX50 strains was expected to enhance co-metabolism of glucose and xylose. To identify whether the engineered strains can simultaneously co-metabolize glucose and xylose or not, we compared cell growth and sugar utilization between MG1655, AXcp, AXcpX50, and AXcpAX50 strains in M9 minimal media with glucose 2 g/L and xylose 2 g/L (Fig. 5). In M9 media supplemented with 2 g/L glucose and 2 g/L xylose, both AXcpX50 and AXcpAX50 strains showed complete utilization of glucose and xylose by 16 hours, while wild type had remaining xylose above 50% (Fig. 5). Despite suffering from slight catabolite repression by glucose, AXcpX50 rapidly consumed xylose after glucose utilization (i.e. 10 hours) (Fig. 5C). While wild type strains could not start xylose consumption for up to 4 hours after glucose consumption (i.e. 14 hours), AXcpX50 strains could finish the completion of xylose by this time. So even if AXcpX50 strains could not overcome completely CCR, its xylose utilization rate after glucose depletion is faster than wild-type cell and cell mass is higher than others strains after 16 hours. Also, glucose metabolism of AXcpX50 strain is recovered because its growth rate in glucose is similar to wild-type. On the other hand, AXcpAX50 strains showed simultaneous co-utilization of glucose and xylose (Fig. 5D). Both glucose and xylose were completely consumed within 14hours. Interestingly, AXcpAX50 strain utilized xylose more rapidly than glucose (Fig. 5B).

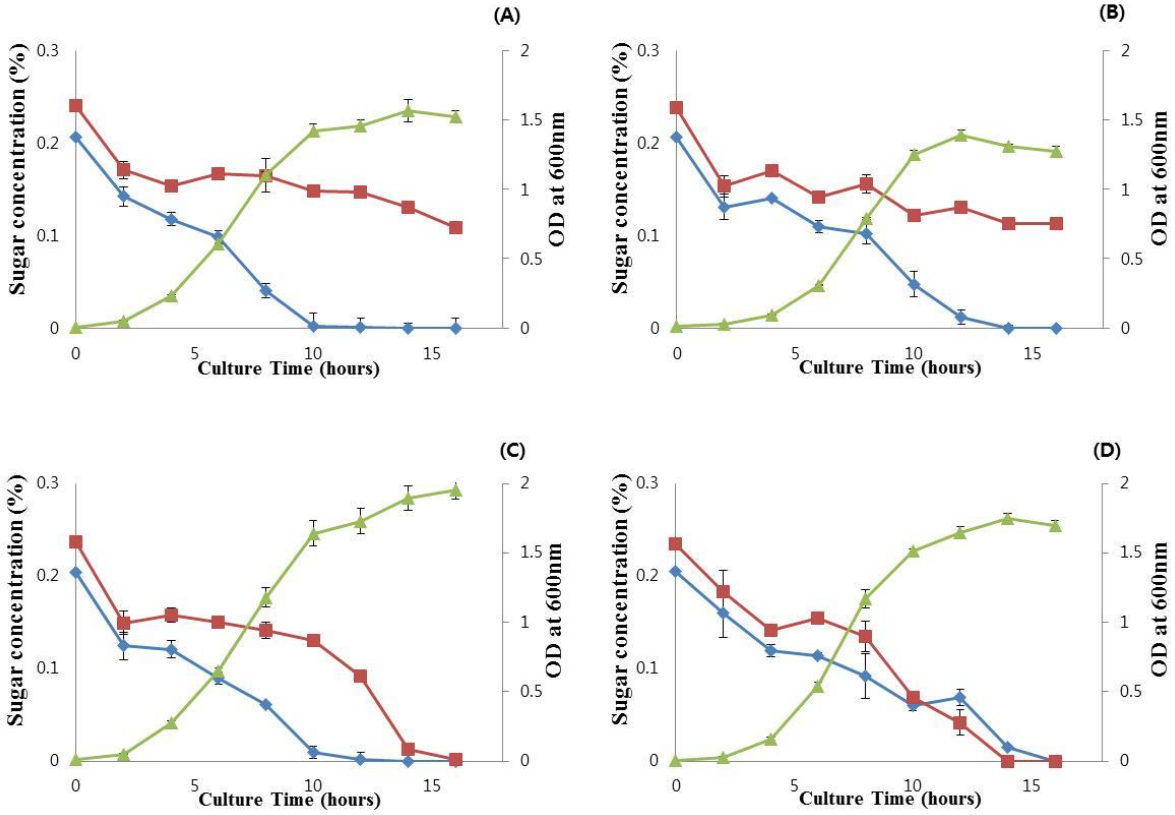

Figure 5. Glucose and xylose utilization pattern in wild type (A), AXcp (B), AXcpX50 (C), and AXcpAX50 (D) in M9 minimal media supplemented with 2 g/L glucose and 2 g/L xylose. Diamond indicates the concentration of glucose, square indicates the concentration of xylose, and triangle indicates cell concentration in OD_{600}. Error bars indicate the standard deviation of experiments performed in triplicate.

To investigate the sugar utilization pattern in a real biomass oil palm empty fruit bunch (EFB) containing approximately 60% glucose and 40% xylose[48], M9 minimal media with glucose 3.5 g/L xylose 2.5 g/L was used to test co-metabolism of these sugars. AXcpX50 strain did not show complete co-metabolizing pattern, but xylose metabolism is improved than wild type cell after glucose depletion. So, in this sugar composition, AXcpX50 strain could not overcome CCR. Compared with wild type cells, AXcpX50 consumed all xylose within 28 hours but wild type cell could not utilize all xylose until 40 hours. Biomass production is best in AXcpX50 strain (Fig. 6C).

Interestingly, AXcpAX50 strain preferentially consumed xylose more rapidly than glucose. It seems that impaired glucose metabolism is not cured, but xylose metabolism is dramatically improved (Fig. 6D). Although most sugars were utilized, cell mass of AXcpAX50 strain is similar to the wild-type. AXcp strain showed poor utilization of both glucose and xylose in this sugar composition (Fig. 6B).

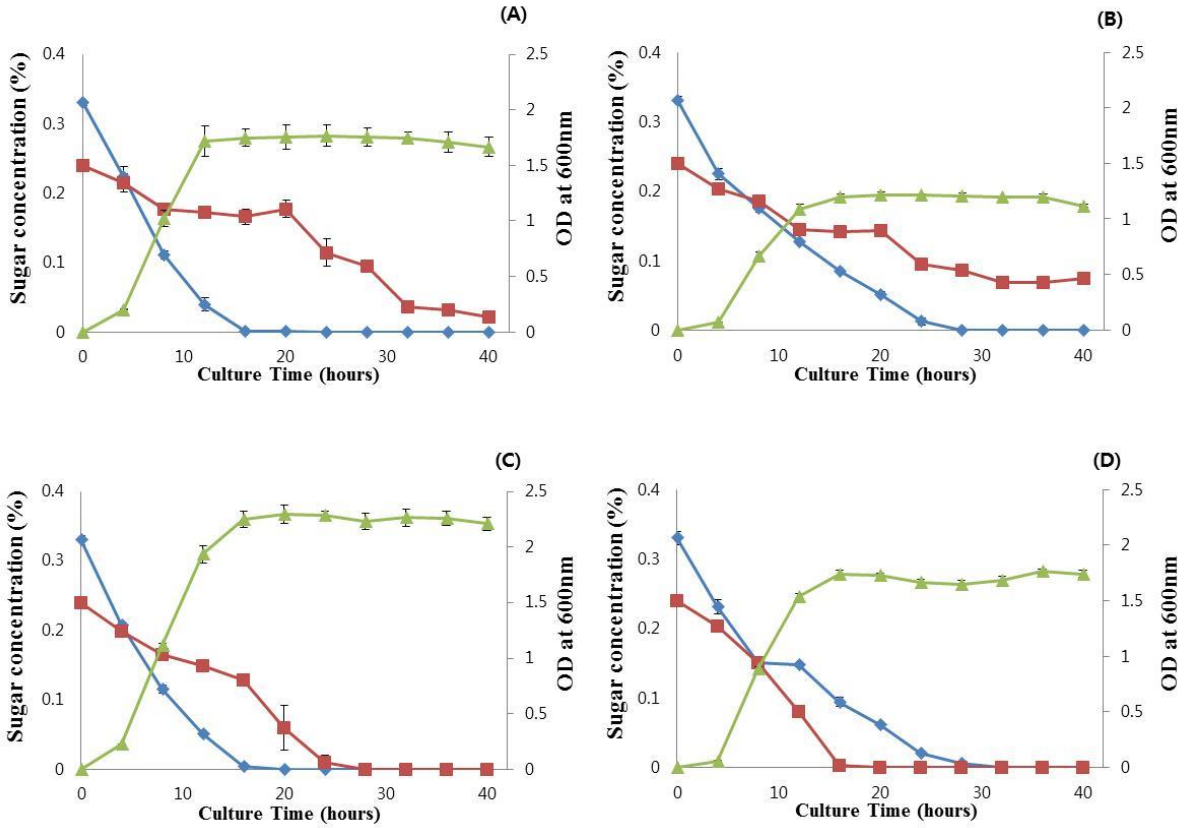

Figure 6. Glucose and xylose utilization pattern in wild type (A), AXcp (B), AXcpX50 (C), and AXcpAX50 (D) in M9 minimal media supplemented with 3.5 g/L glucose and 2.5 g/L xylose. Diamond indicates the concentration of glucose, square indicates the concentration of xylose, and triangle indicates cell concentration in OD_{600}. Error bars indicate the standard deviation of experiments performed in triplicate.

In order to investigate the effect of acidic compound products on the biomass formation of the engineered strains, pH of the broth after fermentation was measured. The pH affects *E. coli* growth substantially and there is no cell growth for pH less than 4[49]. The pH was measured after 48 hours growth in M9 minimal media with glucose 3.5 g/L and xylose 2.5 g/L.

Except for AXcp strain, wild-type, AXcpAX50, and AXcpX50 utilized most sugars in 48 hrs. AXcpX50 showed highest pH 5.07 and cell mass. This indicates that more carbon source may be used for cell growth rather than production of acidic byproducts. In contrast, wild type and AXcpAX50 strain showed lower pH 4.5 and cell mass, indicating that AXcpAX50 strain may have active side pathways for acidic compound production like wild-type (Fig. 7).

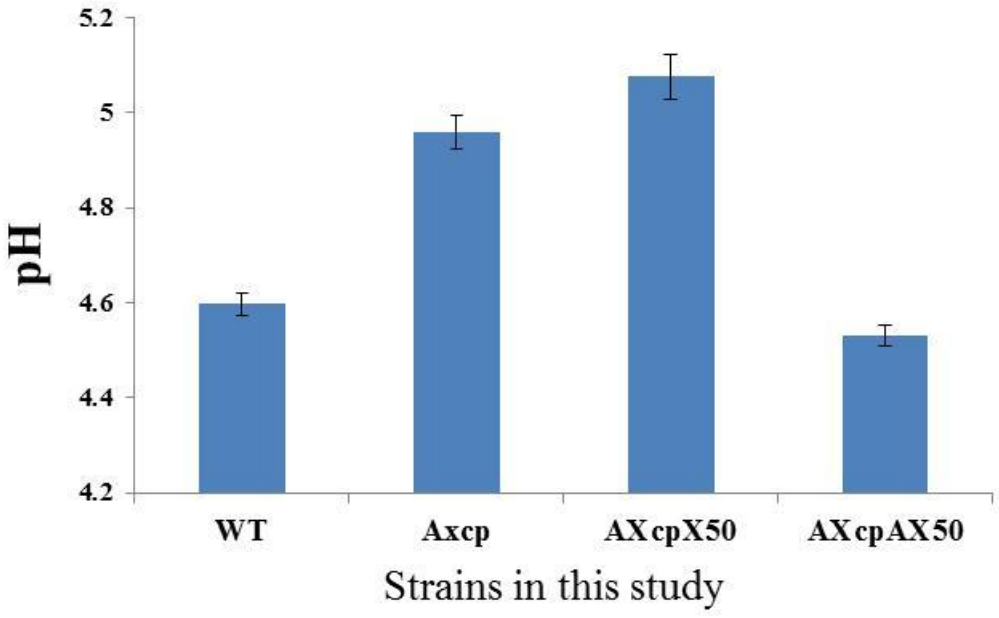

Figure 7. The pH profile after 48 hours of incubation in 50 mL M9 minimal media supplemented with 3.5 g/L glucose and 2.5 g/L xylose. Error bars indicate the standard deviation of experiments performed in triplicate. X-axis indicates *E. coli* strains using this study. Y-axis indicates pH of culture media after 48hour.

3.4 Whole genome re-sequencing analysis

To identify the underlying mutations for the observed phenotype changes in the adapted strains, whole genome re-sequencing analysis of AXcpX50 and AXcpAX50 strains was performed (Macrogen, Korea). Both AXcpX50 and AXcpAX50 strains had meaningful four mutations on their genome, respectively, compared with AXcp (Table 6). AXcpX50 strain had two mutations in coding region and two mutations in non-coding region. First, the 15 base pairs of *araF* coding region were deleted, resulting in the deletion of signal peptide region from 11th amino acid to 15th amino acid in the N-terminal of the protein[50, 51]. The *araF* encodes an L-arabinose-binding periplasmic protein that is a component of the AraFGH for arabinose transporter. The partial deletion of the signal peptide region may affect its function in the periplasmic space through inhibition of secretion of the protein. Second, the deletion of single nucleotide in *araE* coding region causes non-sense mutation of AraE at 126th amino acid, (leucine to stop codon) leading to premature termination of the protein which otherwise would be 472 amino acids long[52]. The *araE* encodes an arabinose-proton symporter that transports arabinose and proton simultaneously into *E. coli* cytoplasm. Third, a single point mutation of *thiC* upstream region where adenine replaces cytosine is located at 151 base upstream from start codon[53]. The thiamine diphosphate binds this region of mRNA and blocks the translation of thiamin diphosphate biosynthase ThiC. The mutation site is -151 region of the regulatory region. Fourth, the exchanged constitutive CP25 promoter for *xylAB* overexpression has mutation at 57th base from guanine to thymine. This region is included in -35 region of CP25 promoter (Table 6.)[43]. Interestingly, xylose adapted strains had all mutations related to transporters for arabinose.

AXcpAX50 has two coding region mutations and two non-coding region mutations. First, a single point mutation in *ybjG* coding region replaces thymine with cytosine, causing a missense mutation in the protein (aspartic acid 99 glycine) (Fig. 9). The YbjG is a predicted protein with undecaprenyl-diphosphatase activity and bacitracin resistance activity when over-expressed[54]. From potentially annotation, the 99th amino acid residue is located in middle of cytoplasmic loop between transmembrane helix. Second, *araE* coding region has a missense mutation in codon 91 causing serine to be replaced by isoleucine. The amino acid residue of AraE protein is the end of cytoplasmic loop between transmembrane helix[52]. Third, a single base pair upstream of *pyrE* region is deleted. The function of this intergenic region is unknown. Fourth, exchanged constitutive CP25 promoter for *xylAB* overexpression has a single base substitution from guanine to thymine at the 57th base. Interestingly, the same mutation in CP25 promoter of *xylAB* was found in AXcpX50 strain. This indicates that xylose metabolism is not optimized by the original CP25 promoter.

Table 6. Whole genome re-sequencing results

Strain	Region	Nucleotide mutation	Base number [a)]	Protein mutation
AXcpX50	*araF* coding region	Deletion GGCTGCCAGACCAAT→(-)	1984108~ 1984122	Deletion of signal peptide region (11^{th}~15^{th} residue)
	araE coding region	Deletion T→(-)	2979938	Nonsense mutation (L126Stop)
	thiC upstream region	Transversion C→A	4194272	Unknown
	xylA CP25 promoter (57^{th} base)	Transversion G→T	None	Unknown
AXcpAX50	*ybjG* coding region	Transition T→C	882316	Missense mutation (D99G)
	araE coding region	Transversion C→A	2979933	Missense mutation (S91I)
	pyrE upstream region	Deletion G→(-)	3813833	Unknown
	xylA CP25 promoter (57^{th} base)	Transversion G→T	None	Unknown

a) NCBI reference sequence NC_000913.2

D: Aspartic acid, G: Glycine, S: Serine, I: Isoleucine, L: Leucine

(1)*pyrE* upstream region

```
                                                 *
MG1655      TCATCTTACTTTTCTACAGACAAAAAAAAGGCGACTCATCAGTCGCCTTAAAAATCAGTT
AXcp        TCATCTTACTTTTCTACAGACAAAAAAAAGGCGACTCATCAGTCGCCTTAAAAATCAGTT
AXcpX50     TCATCTTACTTTTCTACAGACAAAAAAAAGGCGACTCATCAGTCGCCTTAAAAATCAGTT
AXcpAX50    TCATCTTACTTTTCTACAGACAAAAAAAAG-CGACTCATCAGTCGCCTTAAAAATCAGTT
Region      43 base upstream from start codon
```

(2)*ybjG* coding region

```
                                                 *
MG1655      CTT CCT GCA TCA TGC GGC GGA TGA CTC ATT CCC AAG CGA TCA CGG TAC
AXcp        CTT CCT GCA TCA TGC GGC GGA TGA CTC ATT CCC AAG CGA TCA CGG TAC
AXcpX50     CTT CCT GCA TCA TGC GGC GGA TGA CTC ATT CCC AAG CGA TCA CGG TAC
AXcpAX50    CTT CCT GCA TCA TGC GGC GGA TGG CTC ATT CCC AAG CGA TCA CGG TAC
Codon       92   93  94  95  96   97   98  99  100 101 102 103 104 105 106 107
```

(3)*araE* coding region

```
                                        *     *
MG1655      GTT CCG CCT GGG GCG TAA ATA CAG CCT GAT GGC GGG GGC CAT CCT GTT
AXcp        GTT CCG CCT GGG GCG TAA ATA CAG CCT GAT GGC GGG GGC CAT CCT GTT
AXcpX50     GTT CCG CCT GGG GCG TAA -TA CAG CCT GAT GGC GGG GGC CAT CCT GTT
AXcpAX50    GTT CCG CCT GGG GCG TAA ATA CAT CCT GAT GGC GGG GGC CAT CCT GTT
Codon       84   85  86  87   88  89   90  91   92  93  94   95  96  97   98   99
```

(4)*thiC* upstream region

```
                                                 *
MG1655      ATCAGGTTCCGCGGATCCCGAATAAACGGTCTCAGCCAGTTAAGGCACTCCGACAAGAAA
AXcp        ATCAGGTTCCGCGGATCCCGAATAAACGGTCTCAGCCAGTTAAGGCACTCCGACAAGAAA
AXcpX50     ATCAGGTTCCGCGGATCCCGAATAAACGGTATCAGCCAGTTAAGGCACTCCGACAAGAAA
AXcpAX50    ATCAGGTTCCGCGGATCCCGAATAAACGGTCTCAGCCAGTTAAGGCACTCCGACAAGAAA
Region      151base upstream from start codon
```

(5)*araF* coding region

```
                            *** *** *** *** ***
MG1655      TTG TGA CAT AAC GGC TGC CAG ACC AAT GGC TGC CAG GGC TTT AGT
AXcp        TTG TGA CAT AAC GGC TGC CAG ACC AAT GGC TGC CAG GGC TTT AGT
AXcpX50     TTG TGA CAT AAC --- --- --- --- --- GGC TGC CAG GGC TTT AGT
AXcpAX50    TTG TGA CAT AAC GGC TGC CAG ACC AAT GGC TGC CAG GGC TTT AGT
Codon        7   8   9   10  11  12  13  14  15  16  17  18  19  20  21
```

(6)*xylAB* CP25 promoter region

```
                                                                      *
Reference   CTTTGGCAGTTTATTCTTGACATGTAGTGAGGGGGCTGGTATAATCACATAGTACTGTTCACACAGGAA
AXcp        CTTTGGCAGTTTATTCTTGACATGTAGTGAGGGGGCTGGTATAATCACATAGTACTGTTCACACAGGAA
AXcpX50     CTTTGGCAGTTTATTCTTGACATGTAGTGAGGGGGCTGGTATAATCACATAGTACTTTTCACACAGGAA
AXcpAX50    CTTTGGCAGTTTATTCTTGACATGTAGTGAGGGGGCTGGTATAATCACATAGTACTTTTCACACAGGAA
Region              -10 region                        -35 region
```

Figure 8. Comparison of the mutated regions in wild type, AXcp, AXcpX50, and AXcpAX50. Mutated sites are marked as star.

(A) AraE

```
                                                                 *********
MG1655      LDIGVIAGALPFITDHFVLTSRLQEWVVSSMMLGAAIGALFNGWLSFRLGRKYSLMAGAI     132
AXcpX50     LDIGVIAGALPFITDHFVLTSRLQEWVVSSMMLGAAIGALFNGWLSFRLGRNTA------     126
```

(B)AraF

```
                  *****
MG1655      MHKFTKALAAIGLAAVMSQSAMAENLKLGFLVKQPEEPWFQTEWKFADKAGKDLGFEVIK     60
AXcpX50     MHKFTKALAA-----VMSQSAMAENLKLGFLVKQPEEPWFQTEWKFADKAGKDLGFEVIK     55
```

(C)AraE

```
       *
MG1655      LGRKYSLMAGAILFVLGSIGSAFATSVEMLIAARVVLGIAVGIASYTAPLYLSEMASENV     180
AXcpAX50    LGRKYILMAGAILFVLGSIGSAFATSVEMLIAARVVLGIAVGIASYTAPLYLSEMASENV     180
```

(D) YbjG

```
                                                *
MG1655      AIALAVSLFVSWTMGHLFPHDRPFVENIGYNFLHHAADDSFPSDHGTVIFTFALAFLCWH     120
AXcpAX50    AIALAVSLFVSWTMGHLFPHDRPFVENIGYNFLHHAADGSFPSDHGTVIFTFALAFLCWH     120
```

Figure 9. Partial alignments of amino acid sequences between the coding genes. The mutated sites are marked as star.

3.5 *in vivo* analysis of the effect of mutations using genome engineering

To investigate which mutations have significant effect on the improved xylose metabolism in the adapted strains, seven MAGE-primers were simultaneously transformed to AXcpM strain. After four MAGE cycles, cells were recovered during 16 hours, spread on M9 minimal plates supplemented with 4 g/L xylose , and then incubated in 30℃ incubator for 2 days. Twenty-eight of the big colonies were selected for further validation experiments. Selected cells were inoculated into 5 mL LB supplemented with chloramphenicol and grown overnight. The grown cells were transferred (100:1) to 4 g/L xylose M9 minimal media supplemented with chloramphenicol and grown overnight. Among them, 9 colonies showed a significant increase in the cell growth in M9 minimal media with xylose (Fig 10). This indicates that the introduced mutations by MAGE caused the improvement of xylose metabolism.

The genomic DNA was isolated from the nine colonies and the mutation region was PCR amplified and sequenced for mutation validation. The colony No. 9 and 22 have only CP25 promoter mutation for *xylAB*. These two strain showed approximately 3~4 fold higher OD_{600} than AXcpM strain. The colony No. 1, 4, 14, 26, and 28 have two mutations in *xylA* CP25 promoter and *araE* coding region (S99I). These two mutations caused almost 8-fold higher OD_{600} than AXcpM strain. The colony No. 24 has three mutations in *xylA* CP25 promoter, *araE* region (S99I), and *araF* region (partial signal peptide deletion) (Table 6.). This colony showed almost 9-fold higher OD_{600} than AXcpM strain. Interestingly, colony No.15 that possesses *xylA* CP25mutation and *ybjG* region mutation showed about 6-fold higher OD_{600} than AXcpM strain. The minor functions of YbjG are not clearly identified, but there are no predictions relating sugar metabolism yet.

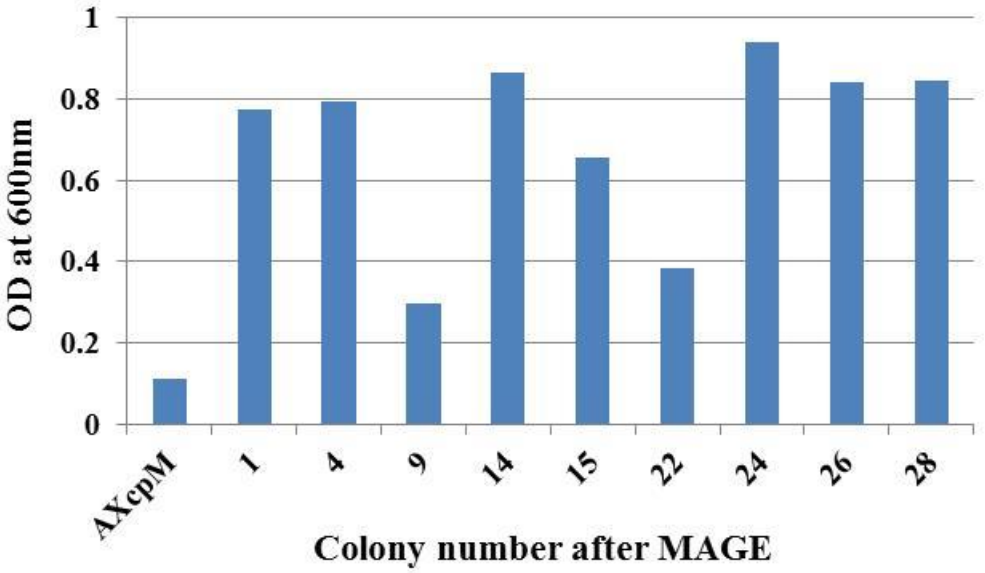

Figure 10. Precise genome engineering to identify critical mutations of different AXcp strains. Colony number 1, 4, 9, 14, 15, 22, 24, 26, and 28 showed improved growth, so sequencing validation was proceeded. X-axis indicates the colony number that shows improved growth in xylose minimal media. Y-axis indicates OD_{600} value after 16 hours in xylose minimal media.

Table 7. MAGE results after sequencing validation

Origin of mutation	Cell / Target	AXcpM	1	4	9	14	15	22	24	26	28
AXcpX50	*araE* (L126Stop)	-[a)]	-	-	-	-	-	-	-	-	-
	araF (Signal del.)	-	-	-	-	-	-	-	O[b)]	-	-
	thiC upstream	-	-	-	-	-	-	-	-	-	-
AXcpAX50	*araE* (S91I)	-	O	O	-	O	-	-	O	O	O
	pyrE upstream	-	-	-	-	-	-	-	-	-	-
	ybjG (D99G)	-	-	-	-	-	O	-	-	-	-
Both strain	*xylA* CP25	-	O	O	O	O	O	O	O	O	O

a) Mutation is not inserted; b) Mutation is inserted.

4. Discussion

In this work, we have engineered an *Escherichia coli* capable of effective co-fermentation of glucose and xylose by eliminating CCR. Previous studies using modulation of *crp*, *ptsG* or *crr* have significant problems such as decreased glucose consumption and arabinose effect on xylose metabolism[30, 31]. To avoid these, we constructed strains with deregulated sugar metabolism for arabinose and xylose by replacing *lac* and *ara* promoters controlled by glucose with synthetic constitutive promoters. Unfortunately, the resulting strain AXcp could not grow well on xylose or arabinose. The reasons for impaired growth of the AXcp strain on arabinose or xylose as sole carbon source could be 1) metabolic burden 2) energy imbalance 3) intermediate toxicity and 4) low expression of transporter genes. First, total five promoters are exchanged with constitutive promoters so that least 12 proteins would be expressed constitutively. Overexpressed mRNA and proteins can cause impaired growth. However, AXcp strain could not show significant impairment of cell growth on LB or M9-glucose medium. This result suggests that overproduced proteins from CP promoters would not affect cell growth. Therefore, metabolic burden seems to have little or no effect on cell growth of AXcp strain. A single mutation on CP25 promoter following adaptation might be because the original CP promoter may not be sufficient to support cell growth with xylose as a sole carbon source. Second, overexpression of arabinose or xylose transporters can affect to cell growth because both ABC sugar transporters encoded by *araBFE* and *xylFGH* use ATP for transporting energy source[13, 55, 56]. The high consumption of ATP for sugar transport may cause negative effect on cell growth. However, the amount of ATP used for sugar transport also would be similar with adapted strain in that there is no significant difference in protein expression. Third, excessive accumulation of intermediate in non-ideal metabolic pathways can affect to growth of AXcp strain on arabinose or xylose. The facilitated influx of arabinose or xylose causes accumulation of high energy and toxic intermediates[57], and up-regulation of glyoxylate pathway caused by phosphate shortage would possibly affect cell growth[58]. In particular, the intermediates of arabinose metabolism, L-ribulose-5-phosphate is toxic to the cell[57]. The excess amount of the intermediate that could not enter PPP pathway may negatively affect cell growth. Moreover, excess arabinose or xylose in high cAMP state causes overproduction of metylglyoxal that can inhibit cell growth[59-62]. In adapted strains, two mutations in *araE* region and *araF* were observed for balanced sugar metabolism by reducing the influx of pentose sugar. Especially, *araE* can also transport xylose in specific media condition[63]. Therefore, possibly mutated transporters lose their function and reduce sugar influx into the cell. Our genome engineering also identifies that premature truncation of AraE could help enhance xylose metabolism. Fourth, even high transcription levels of transporter genes possibly could not guarantee high expression levels of transporter proteins because translation could be down-regulated. Therefore,

the mutations related in arabinose transporters of two adapted strains could increase the activity of transporter for xylose transportation into the cell, especially AraE in AXcpAX50. Actually, AraE can also transport xylose in a specific condition[63].

4.1 Sugar metabolism in strains AXcpX50 and AXcpAX50

Even when AXcpX50 strain showed improved xylose metabolism in the presence of glucose, it could not overcome CCR. The regulation of *araC* is excluded and artificial promoter sequence is introduced with partial deletion of native transcriptional regulatory region. The existing CCR system cannot regulate the transcription of CP promoters in DNA level. So we should consider the possibility of mRNA level and protein level regulation for CCR[64]. Glucose metabolic intermediate would inhibit xylose metabolic enzyme as feedback inhibition or inhibit translation of *xylAB* mRNA[65, 66].

Interestingly, glucose metabolism in AXcp strain is interrupted in the presence of xylose even when there is no engineering to PTS system. It is hard to say exchanged promoters and *araC* deletion cause reduction of glucose metabolism. However, AXcpX50 strain's glucose metabolism is recovered after evolutionary adaptation but not in AXcpAX50. Further studies are needed for identifying mutations relating to glucose metabolism interruption or re-improvement.

By the result of pH measurement, AXcpX50 strain shows low pH fluctuation than other strains. Moreover, biomass formation of AXcpX50 strain is always best. It seems that AXcpX50 strain does not produce acidic metabolites and concentrate on producing biomass. For example, succinate pathway knock out caused improved sugar metabolism and deletion of side pathway improved the cell growth[67]. Therefore, AXcpX50 strain will be used pivotally to produce lactate or ethanol. If we can find pH related mutation not among metabolic genes like AXcpX50, alternative method can be chosen to construct *E. coli* that producing valuable products. Moreover, AXcpAX50 showed lowest pH among four strains. If AXcpAX50 strain can produce acetate or other acids for biofuels more than other strains, the AXcpAX50 also can be used as a platform strain for industry.

4.2 The effect of mutations on co-metabolism of sugars

Four mutations were identified for improvement of xylose metabolism in AXcpM strain such as *xylA* CP25, *araE* in AXcpAX50, *ybjG* mutation, and *araF* mutation. *araE* mutation in AXcpAX50 caused exchange of 91st amino acid residue from serine to isoleucine at end of cytoplasmic loop between transmembrane helix[50]. This mutation might decrease substrate affinity of arabinose transport. *ybjG* mutation in AXcpAX50 strain caused exchange of 99th amino acid residue from aspartic acid to

glycine. There are no academic announcements about *ybjG* in sugar metabolism. It might regulate peptidoglycan layer structure using undecaprenyl pyrophosphate, possibly inducing or increasing the activity of the xylose transporters by altered cell membrane structure[54]. Further studies are needed to understand the effect of *ybjG* on xylose metabolism. *araE* mutation in AXcpX50 caused non-sense mutation, so this protein might lose all their function. *araF* mutation in AXcpX50 strain caused partial deletion of signal peptide, so this protein seems improperly positioned or degraded by error in signal peptide[50]. In addition, the mechanism of mutation that promotes xylose metabolism should be studied in depth.

Unfortunately, Next generation sequencing method (Macrogen, korea) for whole genome re-sequencing could not detect long deletion and long insertion. For more accurate prediction, other methods are needed to find long deletion or insertion. Moreover, the region that could not be analyzed in genome (coverage ~99%) may also need to be identified. Actually, repeated FRT site insertion and recombination using Flp protein possibly cause long deletion or inversion[40].

There are some possibilities to causing mutation from MAGE process. The xylose metabolism of AXcpM strain after MAGE can be improved by spontaneous mutations. The probability seems low, but we should increase MAGE efficiency to reduce mutation from MAGE[47]. Otherwise, to figure out the function of each mutation separately, traditional mutation introducing method also can be used[68].

4.3 Future research plan

We expect that this study will be helpful to design an efficient platform strain for valuable bioproducts from lignocellulosic biomass containing large amount of glucose and xylose. On the basis of the results in the current study, several procedures will be going as followings: (1) For more accurate demonstration of each mutation, all mutations should be introduced into the AXcpM strain. (2) Combinations of mutations are needed to construct a glucose/xylose co-utilizing and fast fermenting strain. (3) The protein and mRNA expression levels should be analyzed to evaluate the expression level by synthetic constitutive promoter and mutations. (4) The improvement of sugar utilization of glucose and xylose in adapted strains using PTS-free glucose transporter or other transporters. (5) Application to ferment of lignocellose hydrolysates but not artificial sugar mixtures, for specific products such as lactate, ethanol, xylitol, and so on.

5. Conclusions

In this study, we constructed improved xylose metabolizing strains such as AXcpX50 and AXcpAX50 using synthetic constitutive promoter and evolutionary adaptation. Through exchange of native promoters with synthetic constitutive promoters and adaptation, we confirmed glucose and xylose co-metabolism and improved xylose metabolism. Whole genome re-sequencing and MAGE results showed sugar transporter and promoter engineering can be used to improve xylose metabolism in *E. coli* MG1655. Through MAGE of 7 mutations, 4 are most important for xylose metabolism such as *araE*, *araF*, CP25 in *xylA*, and *ybjG*. In particular, we discovered that *ybjG* is a new protein involved in xylose metabolism. This is a discovery that probably is worth to further study. Moreover, findings in this work will be applicable to other *E. coli* strains for improvement of xylose metabolism. Furthermore, intensive studies are needed to construct *E. coli* strain that can co-utilize multiple sugars simultaneously.

References

1. **Atsumi, S. & Liao, J.C. Metabolic engineering for advanced biofuels production from *Escherichia coli*. *Current Opinion in Biotechnology* 19, 414-419 (2008).**
2. **Peralta-Yahya, P.P. & Keasling, J.D. Advanced biofuel production in microbes. *Biotechnology Journal* 5, 147-162 (2010).**
3. **Schmidt, L.D. & Dauenhauer, P.J. Chemical engineering: hybrid routes to biofuels. *Nature* 447, 914-915 (2007).**
4. **Zhang, Y.H.P. What is vital (and not vital) to advance economically-competitive biofuels production. *Process Biochemistry* 46, 2091-2110 (2011).**
5. **Lynd, L.R., Weimer, P.J., van Zyl, W.H. & Pretorius, I.S. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews* 66, 506-577 (2002).**
6. **Asghari, A., Bothast, R.J., Doran, J.B. & Ingram, L.O. Ethanol production from hemicellulose hydrolysates of agricultural residues using genetically engineered *Escherichia coli* strain KO11. *Journal of Industrial Microbiology* 16, 42-47 (1996).**
7. **Bothast, R.J., Nichols, N.N. & Dien, B.S. Fermentations with new recombinant organisms. *Biotechnology Progress* 15, 867-875 (1999).**
8. **Hahn-Hägerdal, B., Galbe, M., Gorwa-Grauslund, M.F., Lidén, G. & Zacchi, G. Bioethanol - the fuel of tomorrow from the residues of today. *Trends in Biotechnology* 24, 549-556 (2006).**
9. **Gabor, E. et al. The phosphoenolpyruvate dependent glucose phosphotransferase system from *Escherichia coli* K-12 as the center of a network regulating carbohydrate flux in the cell. *European Journal of Cell Biology* 90, 711-720 (2011).**
10. **Scheller, H.V. & Ulvskov, P. Hemicelluloses. *Annual Review of Plant Biology* 61, 263-289 (2010).**
11. **Desai, T.A. & Rao, C.V. Regulation of arabinose and xylose metabolism in *Escherichia coli*. *Applied and Environmental Microbiology* 76, 1524-1532 (2010).**
12. **Cirino, P.C., Chin, J.W. & Ingram, L.O. Engineering *Escherichia coli* for xylitol production from glucose-xylose mixtures. *Biotechnology and Bioengineering* 95, 1167-1176 (2006).**
13. **Sumiya, M., Davis, E.O., Packman, L.C., McDonald, T.P. & Henderson, P.J. Molecular genetics of a receptor protein for D-xylose, encoded by the gene xylF, in *Escherichia coli*. *Receptors & channels* 3, 117-128 (1995).**
14. **Shamanna, D.K. & Sanderson, K.E. Uptake and catabolism of D-xylose in *Salmonella***

typhimurium **LT2.** ***Journal of Bacteriology*** **139, 64-70 (1979).**

15. **Groff, D. et al. Supplementation of Intracellular XylR Leads to Coutilization of Hemicellulose Sugars.** ***Applied and Environmental Microbiology*** **78, 2221-2229 (2012).**
16. **Song, S. & Park, C. Organization and regulation of the D-xylose operons in** ***Escherichia coli*** **K-12: XylR acts as a transcriptional activator.** ***Journal of Bacteriology*** **179, 7025-7032 (1997).**
17. **Fau., D.J. & Wiesmeyer, H. Control of xylose metabolism in** ***Escherichia coli*****. 497-499 (1970).**
18. **Gorke, B. & Stulke, J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients.** ***Nat Rev Micro*** **6, 613-624 (2008).**
19. **Deutscher, J. The mechanisms of carbon catabolite repression in bacteria.** ***Current Opinion in Microbiology*** **11, 87-93 (2008).**
20. **Pennetier, C., Oberto, J. & Plumbridge, J. An antisense transcript from within the** ***ptsG*** **promoter region in** ***Escherichia coli*****.** ***Journal of Molecular Microbiology and Biotechnology*** **18, 230-240 (2010).**
21. **Karimova, G., Ladant, D. & Ullmann, A. Relief of catabolite repression in a cAMP-independent catabolite gene activator mutant of** ***Escherichia coli*****.** ***Research in Microbiology*** **155, 76-79 (2004).**
22. **Peterkofsky, A. Redistribution of phosphate pools and the regulation of** ***Escherichia coli*** **adenylate cyclase activity.** ***Archives of Biochemistry and Biophysics*** **265, 227-233 (1988).**
23. **Jojima, T., Omumasaba, C., Inui, M. & Yukawa, H. Sugar transporters in efficient utilization of mixed sugar substrates: current knowledge and outlook.** ***Appl Microbiol Biotechnol*** **85, 471-480 (2010).**
24. **Shimada, T., Fujita, N., Yamamoto, K. & Ishihama, A. Novel roles of cAMP receptor protein (CRP) in regulation of transport and metabolism of carbon sources.** ***PLoS ONE*** **6, e20081 (2011).**
25. **Gulati, A. & Mahadevan, S. Mechanism of catabolite repression in the bgl operon of** ***Escherichia coli*****: involvement of the anti-terminator BglG, CRP-cAMP and EIIAGlc in mediating glucose effect downstream of transcription initiation.** ***Genes to Cells*** **5, 239-250 (2000).**
26. **Gosset, G. Improvement of** ***Escherichia coli*** **production strains by modification of the phosphoenolpyruvate:sugar phosphotransferase system.** ***Microbial Cell Factories*** **4, 14 (2005).**
27. **Yao, R. et al. Catabolic regulation analysis of** ***Escherichia coli*** **and its** ***crp, mlc, mgsA, pgi*** **and** ***ptsG*** **mutants.** ***Microbial Cell Factories*** **10, 67 (2011).**
28. **Nichols, N., Dien, B. & Bothast, R. Use of catabolite repression mutants for fermentation**

of sugar mixtures to ethanol. *Appl Microbiol Biotechnol* 56, 120-125 (2001).

29. **Lindsay, S.E., Bothast, R.J. & Ingram, L.O. Improved strains of recombinant *Escherichia coli* for ethanol production from sugar mixtures. *Appl Microbiol Biotechnol* 43, 70-75 (1995).**
30. **Hernández-Montalvo, V., Valle, F., Bolivar, F. & Gosset, G. Characterization of sugar mixtures utilization by an *Escherichia coli* mutant devoid of the phosphotransferase system. *Appl Microbiol Biotechnol* 57, 186-191 (2001).**
31. **Hernández-Montalvo, V. et al. Expression of galP and glk in a *Escherichia coli* PTS mutant restores glucose transport and increases glycolytic flux to fermentation products. *Biotechnology and Bioengineering* 83, 687-694 (2003).**
32. **Zhang, J. & Greasham, R. Chemically defined media for commercial fermentations. *Appl Microbiol Biotechnol* 51, 407-421 (1999).**
33. **Hollands, K., Busby, S.J. & Lloyd, G.S. New targets for the cyclic AMP receptor protein in the *Escherichia coli* K-12 genome. *FEMS Microbiol Lett* 274, 89-94 (2007).**
34. **Khankal, R., Chin, J., Ghosh, D. & Cirino, P. Transcriptional effects of CRP* expression in *Escherichia coli*. *Journal of Biological Engineering* 3, 13 (2009).**
35. **Balderas-Hernández, V., Hernández-Montalvo, V., Bolívar, F., Gosset, G. & Martínez, A. Adaptive evolution of *Escherichia coli* inactivated in the phosphotransferase system operon improves co-utilization of xylose and glucose under anaerobic conditions. *Appl Biochem Biotechnol* 163, 485-496 (2011).**
36. **Vinuselvi, P. et al. Engineering microorganisms for biofuel production. *Biofuels* 2, 153-166 (2011).**
37. **Garges, S. & Adhya, S. Sites of allosteric shift in the structure of the cyclic AMP receptor protein. *Cell* 41, 745-751 (1985).**
38. **Blattner, F.R. et al. The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453-1462 (1997).**
39. **Datta, S., Costantino, N. & Court, D.L. A set of recombineering plasmids for gram-negative bacteria. *Gene* 379, 109-115 (2006).**
40. **Cherepanov, P.P. & Wackernagel, W. Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* 158, 9-14 (1995).**
41. **Datsenko, K.A. & Wanner, B.L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences* 97, 6640-6645 (2000).**
42. **Warrens, A.N., Jones, M.D. & Lechler, R.I. Splicing by overlap extension by PCR using asymmetric amplification: an improved technique for the generation of hybrid proteins of**

immunological interest. *Gene* 186, 29-35 (1997).

43. Jensen, P.R. & Hammer, K. The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. *Applied and Environmental Microbiology* 64, 82-87 (1998).

44. Park, J.M., Vinuselvi, P. & Lee, S.K. The mechanism of sugar-mediated catabolite repression of the propionate catabolic genes in *Escherichia coli*. *Gene* 504, 116-121 (2012).

45. Wang, H.H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460, 894-898 (2009).

46. Sawitzke, J.A. et al. Probing cellular processes with oligo-mediated recombination and using the knowledge gained to optimize recombineering. *Journal of Molecular Biology* 407, 45-59 (2011).

47. Wang, H.H. & Church, G.M. in Methods in Enzymology, Vol. Volume 498. (ed. V. Christopher) 409-426 (Academic Press, 2011).

48. Ling, T., Hassan, O., Badri, K., Maskat, M. & Mustapha, W. Sugar recovery of enzymatic hydrolysed oil palm empty fruit bunch fiber by chemical pretreatment. *Cellulose*, 1-13 (2013).

49. Holt J.G., K.N.R., Sneath P., Staly J.T., Williams S.T. Bergey's manual of determinative bacteriology. *Williams & Wilkins* 9 (1994).

50. Hogg, R.W. & Hermodson, M.A. Amino acid sequence of the L-arabinose-binding protein from *Escherichia coli* B/r. *Journal of Biological Chemistry* 252, 5135-5141 (1977).

51. Link, A.J., Robison, K. & Church, G.M. Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12. *ELECTROPHORESIS* 18, 1259-1313 (1997).

52. Stoner, C. & Schleif, R. The araE low affinity L-arabinose transport promoter: Cloning, sequence, transcription start site and DNA binding sites of regulatory proteins. *Journal of Molecular Biology* 171, 369-381 (1983).

53. Zheng, J., Singh, V.K. & Jia, Z. Identification of an ITPase/XTPase in *Escherichia coli* by structural and biochemical analysis. *Structure* 13, 1511-1520 (2005).

54. Ghachi, M.E., Derbise, A., Bouhss, A. & Mengin-Lecreulx, D. Identification of multiple genes encoding membrane proteins with undecaprenyl pyrophosphate phosphatase (UppP) activity in *Escherichia coli*. *Journal of Biological Chemistry* 280, 18689-18695 (2005).

55. Kolodrubetz, D. & Schleif, R. L-arabinose transport systems in *Escherichia coli* K-12. *Journal of Bacteriology* 148, 472-479 (1981).

56. Horazdovsky, B.F. & Hogg, R.W. Genetic reconstitution of the high-affinity L-arabinose transport system. *Journal of Bacteriology* 171, 3053-3059 (1989).

57. Khushnuma, K. & Christopher, V.R. Identification and analysis of the putative pentose

sugar efflux transporters in *Escherichia coli*. *PLoS ONE* 7 (2012).

58. Ackerman, R.S., Cozzarelli, N.R. & Epstein, W. Accumulation of toxic concentrations of methylglyoxal by wild-type *Escherichia coli* K-12. *Journal of Bacteriology* 119, 357-362 (1974).

59. Yomano, L.P., York, S.W., Shanmugam, K.T. & Ingram, L.O. Deletion of methylglyoxal synthase gene (mgsA) increased sugar co-metabolism in ethanol-producing *Escherichia coli*. *Biotechnol Lett* 31, 1389-1398 (2009).

60. Tötemeyer, S., Booth, N.A., Nichols, W.W., Dunbar, B. & Booth, I.R. From famine to feast: the role of methylglyoxal production in *Escherichia coli*. *Molecular Microbiology* 27, 553-562 (1998).

61. Hopper, D.J. & Cooper, R.A. The regulation of *Escherichia coli* methylglyoxal synthase; a new control site in glycolysis? *FEBS Letters* 13, 213-216 (1971).

62. Ferguson, G.P., Tötemeyer, S., MacLean, M.J. & Booth, I.R. Methylglyoxal production in bacteria: suicide or survival? *Arch Microbiol* 170, 209-218 (1998).

63. Hasona, A., Kim, Y., Healy, F.G., Ingram, L.O. & Shanmugam, K.T. Pyruvate formate lyase and acetate kinase are essential for anaerobic growth of *Escherichia coli* on xylose. *Journal of Bacteriology* 186, 7593-7600 (2004).

64. Kaplan, S., Bren, A., Zaslaver, A., Dekel, E. & Alon, U. Diverse two-dimensional input functions control bacterial sugar genes. *Molecular Cell* 29, 786-792 (2008).

65. Aiba, H. Mechanism of RNA silencing by Hfq-binding small RNAs. *Current Opinion in Microbiology* 10, 134-139 (2007).

66. Beisel, C.L. & Storz, G. The base-pairing RNA spot 42 participates in a multioutput feedforward loop to help enact catabolite repression in *Escherichia coli*. *Molecular Cell* 41, 286-297 (2011).

67. Chiang, C.J. et al. Systematic approach to engineer *Escherichia coli* pathways for co-utilization of a glucose–xylose mixture. *Journal of Agricultural and Food Chemistry* 61, 7583-7590 (2013).

68. Pósfai, G., Kolisnychenko, V., Bereczki, Z. & Blattner, F.R. Markerless gene replacement in *Escherichia coli* stimulated by a double-strand break in the chromosome. *Nucleic Acids Research* 27, 4409-4415 (1999).