ACS Synthetic Biology

pubs.acs.org/synthbio

Research Article

New Target Gene Screening Using Shortened and Random sgRNA Libraries in Microbial CRISPR Interference

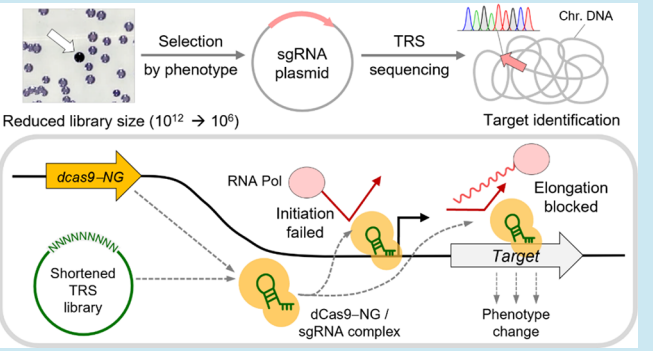
Song Hee Jeong, Hyun Ju Kim, and Sang Jun Lee*

Cite This: *ACS Synth. Biol.* 2023, 12, 800–808

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: CRISPR interference (CRISPRi) screening has been used for identification of target genes related to specific phenotypes using single-molecular guide RNA (sgRNA) libraries. In CRISPRi screening, the sizes of random sgRNA libraries contained with the original target recognition sequences are large ($\sim 10^{12}$). Here, we demonstrated that the length of the target recognition sequence (TRS) can be shortened in sgRNAs from the original 20 nucleotides (N_{20}) to 9 nucleotides (N_9) that is still sufficient for dCas9 to repress target genes in the xylose operon of *Escherichia coli*, regardless of binding to a promoter or open reading frame region. Based on the results, we constructed random sgRNA plasmid libraries with 5′-shortened TRS lengths, and identified xylose metabolic target genes by Sanger sequencing of sgRNA plasmids purified from Xyl^- phenotypic cells. Next, the random sgRNA libraries were harnessed to screen for target genes to enhance violacein pigment production in synthetic *E. coli* cells. Seventeen target genes were selected by analyzing the redundancy of the TRS in sgRNA plasmids in dark purple colonies. Among them, seven genes (*tyrR*, *pykF*, *cra*, *ptsG*, *pykA*, *sdaA*, and *tnaA*) have been known to increase the intracellular L-tryptophan pool, the precursor of a violacein. Seventeen cells with a single deletion of each target gene exhibited a significant increase in violacein production. These results indicate that using shortened random TRS libraries for CRISPRi can be simple and cost-effective for phenotype-based target gene screening.



KEYWORDS: *CRISPR interference, deactivated Cas9, random library, shortened sgRNA, target recognition sequence, violacein, phenotype-based target screening*

INTRODUCTION

The CRISPR-Cas system, which was first discovered as an adaptive immune system in prokaryotes,[1,2] has evolved as a genome-editing tool in a variety of living organisms, from microorganisms to humans.[3−8] The CRISPR-Cas9 system consists of a target-recognizing single-molecular guide RNA (sgRNA) and a Cas9 proteinaceous nuclease that cleaves target DNA.[9] By removing the nucleolytic activity of Cas9 to make the deactivated Cas9 (dCas9) protein and leaving sgRNA capable of recognizing specific sequences, an artificial transcriptional repressor can be created for specific control over the expression of the desired gene. This is referred to as CRISPR interference (CRISPRi).[10]

CRISPRi can simultaneously regulate the expression of several specific genes by introducing multiple sgRNAs and reversibly controlling the degree of gene expression and repression.[11] Recently, CRISPR screening studies have revealed new target genes, pathways, and mechanisms at the genomic level through genotype−phenotype mapping using sgRNA libraries. CRISPR screening is generally performed in a pooled and arrayed manner.[12]

First, the pooled CRISPR screen identifies target genes by sequencing sgRNAs extracted from cells showing specific phenotypic changes by introducing a mixture of sgRNA libraries into the cells.[13,14] For example, an L-proline exporter was successfully discovered in *Corynebacterium glutamicum* using an sgRNA library targeting 397 potential transporter genes.[15] Target genes were identified by screening both cell growth and L-lactate productivity using an sgRNA library targeting specific genes in the cyanobacterium *Synechocystis* sp. PCC 6803.[16]

Second, the arrayed CRISPR screening method introduces an sgRNA that targets one gene per well in a multiwell plate and observes the phenotype occurring in each cell.[17] For instance, a recent study showed that the yield and productivity of fatty acids were increased using an sgRNA library targeting

Received: November 6, 2022
Published: February 14, 2023

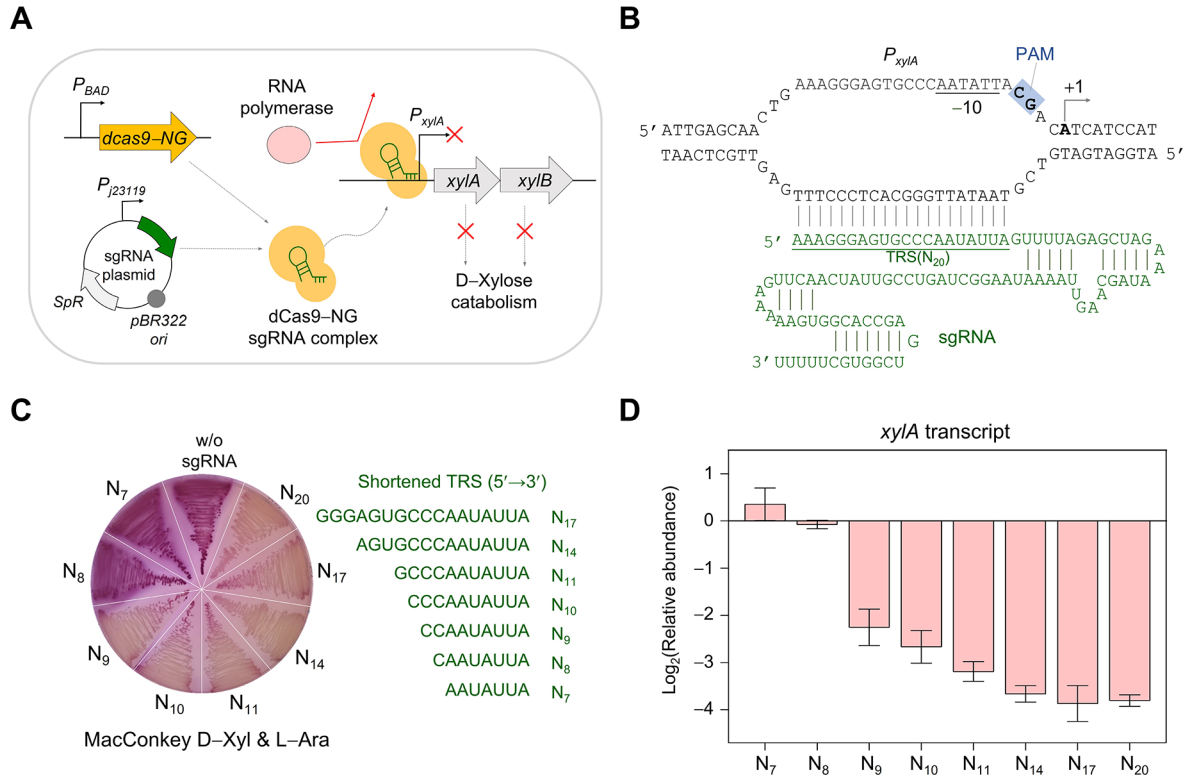


Figure 1. Minimum target recognition sequence (TRS) length of sgRNA. (A) CRISPRi system with L-arabinose-inducible genomic *dcas9-NG* and sgRNA plasmid. The dCas9-NG/sgRNA complex can repress the transcription of *xylAB* genes. (B) sgRNA with the full length (N_{20}) of TRS hybridizing to the −10 region of the *xylA* promoter. (C) D-Xylose utilization assay of cells with dCas9-NG/sgRNAs with various shortened TRS (N_7, N_8, N_9, N_{10}, N_{11}, N_{14}, and N_{17}). (D) Quantitative transcript analysis of D-xylose catabolic gene (*xylA*) expression by sgRNAs with shortened TRS.

genes selected through omics analyses. This had no direct relationship with free fatty acid metabolism.[18] It has also been shown that phosphatase genes that the increase in terpenoid production in *Escherichia coli* can be identified using specific sgRNA libraries that repress phosphatase-encoding genes.[19]

In most CRISPRi screening studies, large numbers of oligonucleotides are individually synthesized using microarrays when preparing sgRNA libraries. However, building sgRNA libraries using array-synthesized oligonucleotides is expensive, and the use of constructed libraries is limited to cells. In this study, we found that the short target recognition sequence (TRS) lengths of sgRNAs can effectively repress gene expression. We confirmed that genes related to a specific phenotype could be identified using the shortened random sgRNA library method with *E. coli* as a model, and it was effective when applied to the discovery of target genes that increase microbial violacein production. We have further discussed the effectiveness of the 5′-shortened random TRS library method in comparison with other CRISPR screening techniques.

RESULTS

Length of TRS in sgRNA Required for CRISPRi. We investigated the length of the TRS in sgRNAs required for CRISPRi in the *E. coli* HK1160 strain, where engineered dCas9 proteins with 5′-NG as the PAM sequence (dCas9-NG) were inducible by the addition of L-arabinose. *E. coli* HK1160 cells were transformed with sgRNA plasmids targeting the *xylA* promoter. When the dCas9-NG/sgRNA complex bound to the *xylA* promoter, the transcription of D-xylose metabolic *xylAB* genes (encoding xylose isomerase and xylulokinase) were repressed (Figure 1A). The original length (20 nt; N_{20}) of the TRS in the sgRNA was designed to hybridize with the −10 region of the *xylA* promoter (Figure 1B). Various sgRNA plasmids with different lengths of TRS (N_7, N_8, N_9, N_{10}, N_{11}, N_{14}, N_{17}, and N_{20}) were transformed into HK1160 cells, and D-xylose utilization was checked on MacConkey agar containing both D-xylose and L-arabinose. When sgRNA plasmids carrying the original length of the TRS (N_{20}) and 5′-end-shortened TRS (N_9, N_{10}, N_{11}, N_{14}, and N_{17}) were transformed, white colonies were observed, signifying the repression of D-xylose metabolic enzyme genes. However, HK1160 cells harboring either sgRNA-deleted or sgRNA plasmids carrying a relatively shorter TRS (N_7 and N_8) formed red colonies, indicating that dCas9-NG/shortened sgRNAs could not repress the *xylA* promoter when the length of the TRS was less than 9 nt (N_9) (Figure 1C). The repression of the *xylA* promoter by the dCas9-NG/sgRNA complex was verified by quantitative reverse transcription PCR (RT-qPCR). The relative abundance of *xylA* transcripts in cells transformed with various sgRNA plasmids was compared to that in cells transformed with the sgRNA-deleted plasmid. The degree of *xylA* transcriptional repression did not change from N_{14} to N_{20} of the TRS in sgRNAs. The degree of *xylA* repression gradually increased from N_9 to N_{11} in the TRS. *xylA* repression was not observed in the case of shorter TRS (N_7 and N_8) (Figure 1D). The RT-qPCR results showed that the shorter the length of the TRS in the sgRNA, the weaker the binding activity of the dCas9-NG/sgRNA complex to the DNA target. Collectively, these results demonstrated that the minimum length of the

https://doi.org/10.1021/acssynbio.2c00595
ACS Synth. Biol. 2023, 12, 800−808

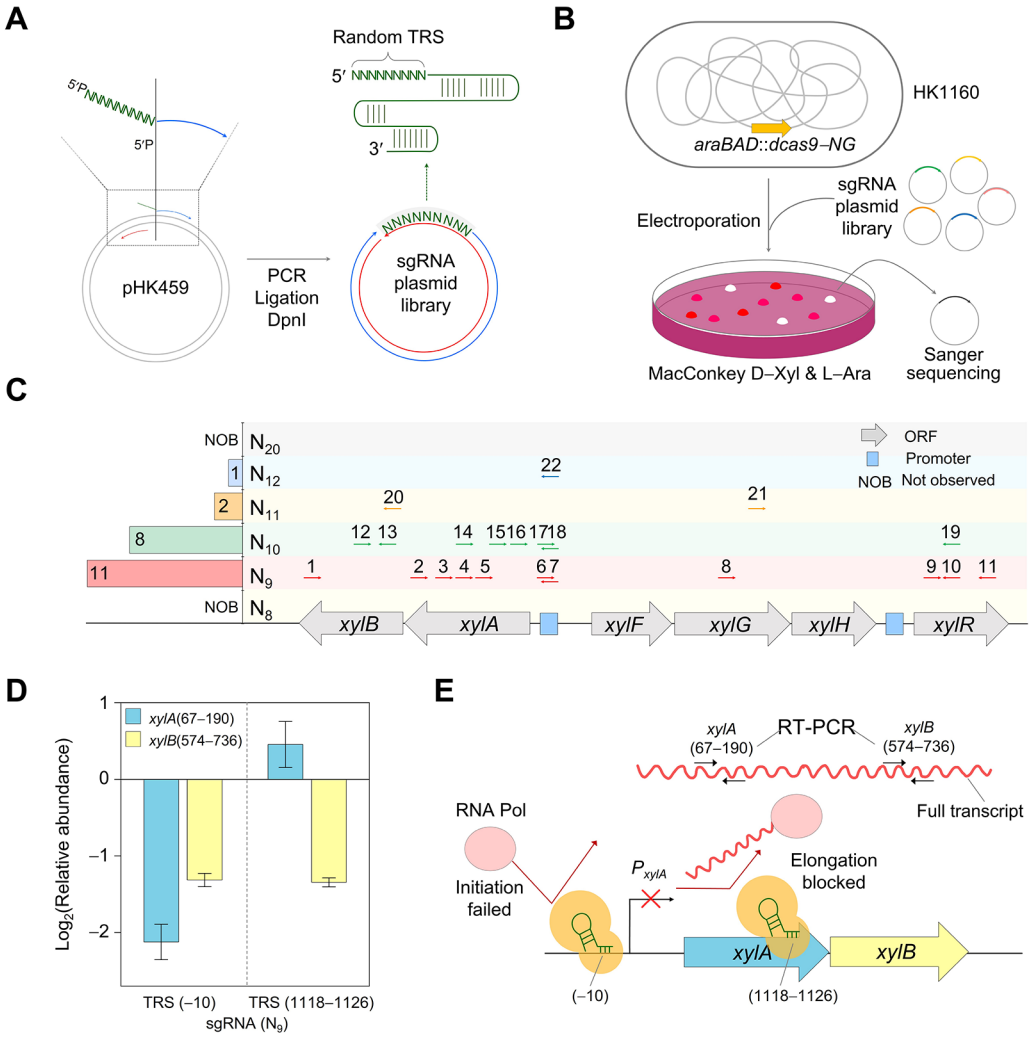


Figure 2. Use of sgRNA libraries with random and shortened TRS for CRISPRi screening of target genes responsible for Xyl$^-$ phenotypes in the D-xylose operon. (A) Construction of 5′-shortened random sgRNA library using 5′-phosphorylated random primers. (B) Selection of cells with the Xyl$^-$ phenotype among transformants with the shortened random sgRNA library. (C) Positions of target sequences and length of TRS found in the xylose operon using six random and shortened sgRNA libraries in the dCas9-NG system. (D) RT-qPCR analysis of *xylA* and *xylB* transcripts according to sgRNA binding sites (promoter and open reading frame). (E) Binding sites of the dCas9-NG/sgRNA complex in the *xylA* promoter (−10) and the middle of the open reading frame of *xylA*. The positions of primers for RT-qPCR were designed before and after the binding site (1118−1126) of the dCas9-NG/sgRNA complex.

TRS in sgRNAs required for transcriptional repression and phenotype change was confirmed to be 9 nt.

Shortened and Random TRS in sgRNA Libraries for Target Screening. Based on the above results showing that markedly shortened sgRNAs can control gene expression and change cellular phenotypes in CRISPRi, we explored whether target genes could be screened using random shortened TRS libraries. Primers with shortened and random TRS were designed, and sgRNA plasmid libraries with various lengths of TRS (N_8, N_9, N_{10}, N_{11}, N_{12}, and N_{20}) were constructed (Figure 2A).

We then obtained random sgRNA libraries in *E. coli* DH5α, the sizes of which were approximately 10^5 CFU/ligated DNA (μg). Next, each random sgRNA plasmid library was transformed into HK1160 cells carrying L-arabinose-inducible dCas9-NG, an engineered dCas9 that recognizes 5′-NG sequence as PAM (Figure 2B). Transformants ($\sim 10^4$) were grown on MacConkey agar in a square dish containing D-xylose and L-arabinose (Table S1). In the HK1160 transformant cells, 1, 2, 8, and 11 white colonies were observed in the N_{12}, N_{11}, N_{10}, and N_9 sgRNA plasmid libraries, respectively (Figure 2C). White colonies were not observed among transformant cells ($\sim 10^4$) on square dishes when the N_8 and N_{20} sgRNA plasmid libraries were transformed (Figure S1). For N_8, it seems that it cannot be repressed because it cannot be tightly attached to the target, as shown in Figure 1D. When the length of TRS in sgRNAs is 20 nt, the size of random TRS with all possible nucleotide sequences was 4^{20} ($\sim 10^{12}$). We failed to obtain white colonies with the N_{20} sgRNA plasmid library, probably due to the limited transformant cells. The size ($\sim 10^{12}$) of N_{20} sgRNA plasmid library is much larger than the number ($\sim 10^4$) of transformants.

Twenty-two different sgRNA plasmids were obtained from 1, 2, 8, and 11 white colony-forming transformant cells. Since the target recognition of dCas9-NG was determined by the TRS and PAM sequences, the TRS and 5′-NG PAM were searched in the xylose operon in the *E. coli* genome to identify target nucleotide sequences. For example, a specific gene target can be identified in the xylose operon by finding 5′-CGTAATATTNG, which was obtained by combining 5′-

https://doi.org/10.1021/acssynbio.2c00595
ACS Synth. Biol. 2023, 12, 800−808

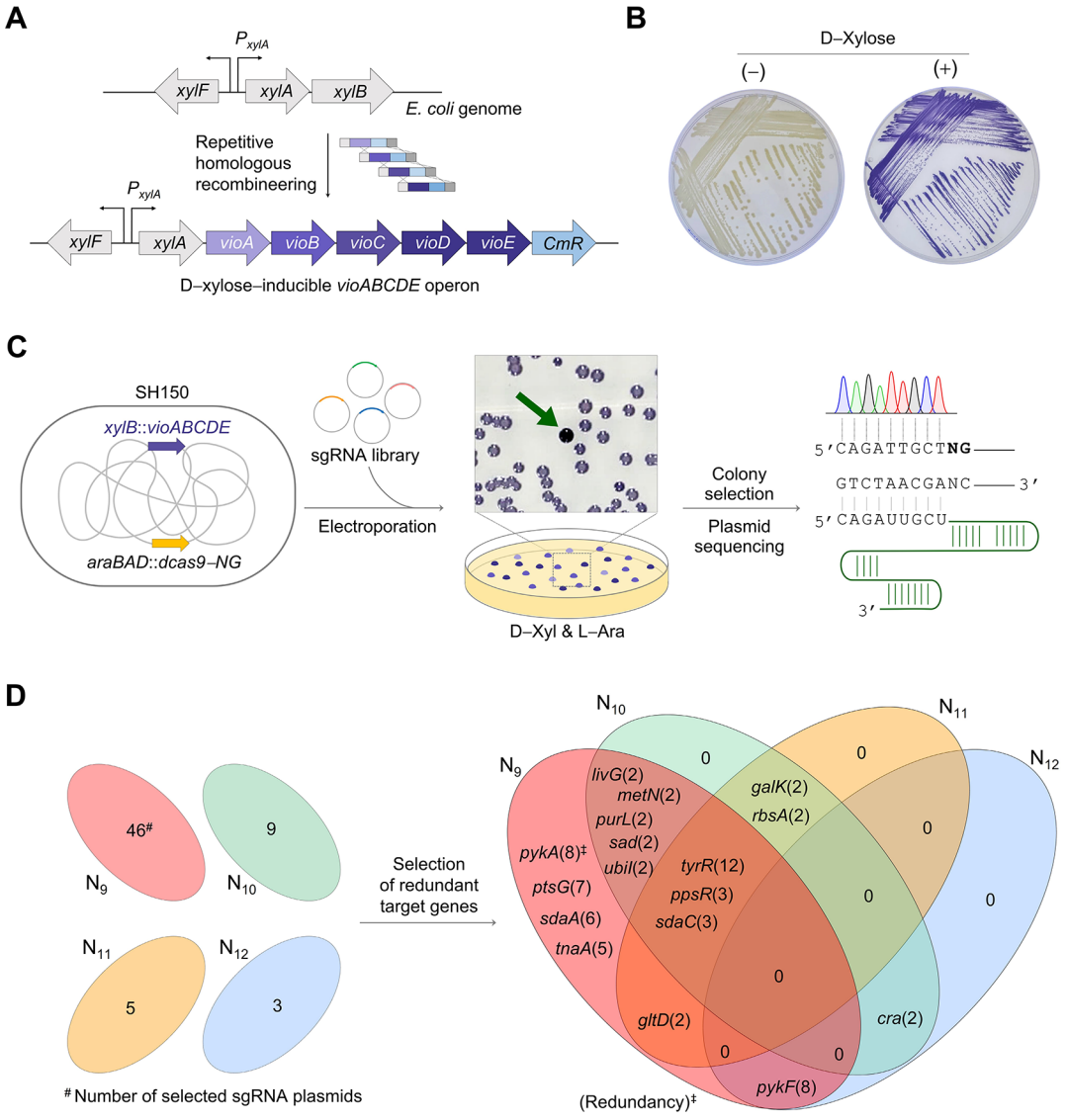


Figure 3. Screening of novel target genes for violacein production. (A) Genomic integration of violacein biosynthetic genes (*vioABCDE*, 7.8 kb) in *Escherichia coli* using repetitive homologous recombineering. (B) D-Xylose-inducible heterologous violacein production. (C) Phenotype-based CRISPRi screening of dCas9-NG-expressing and violacein-producing *E. coli* cells using 5′-shortened random sgRNA libraries. (D) Selection of redundantly found target genes by overlapping of the target candidate genes obtained from 5′-shortened sgRNA libraries.

CGTAATATT from sequencing, and 5′-NG, which was the PAM sequence (Table S2). These results indicate that xylose metabolism-related genes can be targeted using 5′-shortened sgRNA libraries.

We visualized the location of the TRS in the xylose operon. TRS can be located in regulatory or structural gene regions (Figure 2C and Table S2). Five TRSs in the promoter region and 17 in the structural gene were identified. We then tested whether sgRNA could effectively repress xylose operon gene expression, even when bound to a structural gene, using RT-qPCR (Figure 2D). When the sgRNA was bound to the *xylA* promoter, the expression of *xylA* and *xylB* was reduced by 2.1- and 1.3-fold, respectively, by CRISPRi. When the sgRNA was bound to the end (1118−1126 nt) of the *xylA* structural gene, the expression of *xylA* was not repressed but instead increased by 0.5-fold, and that of *xylB* was decreased by 1.3-fold by CRISPRi. In the case of dCas9-NG/sgRNA binding to the *xylA* promoter, the transcripts of *xylA* and *xylB* did not form. When the dCas9-NG/sgRNA complex bound to the *xylA* structural gene, the front transcript of the *xylA* gene formed, but the transcript of the *xylB* gene did not form (Figure 2E). Using the 5′-shortened random TRS library, it was possible to identify the target sequences in the genomes of selected cells with altered phenotypes. In addition, it was confirmed that these phenotypic changes resulted from transcriptional repression, regardless of the structural or regulatory regions of the target genes.

CRISPR Screening of Target Genes for Violacein Biosynthesis. To expand the 5′-shortened and random TRS library method to screen for new target genes in the genome, we used a violacein biosynthetic phenotype as a model. We prepared an *E. coli* strain in which the violacein biosynthesis gene *vioABCDE* was introduced into the xylose operon of the MG1655 genome (Figure 3A). In the presence of D-xylose in the culture medium, the recombinant cells formed purple colonies due to the biosynthesis and intracellular accumulation of violacein (Figure 3B). We transformed 5′-shortened random sgRNA libraries (N_8, N_9, N_{10}, N_{11}, N_{12}, and N_{20}) into the

https://doi.org/10.1021/acssynbio.2c00595
ACS Synth. Biol. 2023, 12, 800−808

SH150 strain carrying the D-xylose-inducible *vioABCDE* operon and L-arabinose-inducible *dcas9-NG* in the genome (Figure 3C). Colonies with increased violacein production were selected among the colonies (~10^4) on the LB agar containing L-arabinose and D-xylose (Figure S2). Colored images of the colonies were converted to red, green, and blue (RGB) parameters. In general, the darker the purple color, the lower the RGB value. In addition, the value of green (G), which is a complementary color to purple, decreased. Light and dark-purple colonies were distinguished using the RGB and G values (Figure S3 and Table S3). We determined the TRS in 68 sgRNAs purified from dark purple colonies (49 colonies from the N_9 sgRNA library, 9 from N_{10}, 6 from N_{11}, and 4 from N_{12}) by Sanger sequencing. With the five redundant sgRNA plasmids, 63 TRS were selected (N_9 = 46, N_{10} = 9, N_{11} = 5, and N_{12} = 3).

The following criteria were used for the selection of target genes responsible for enhanced violacein production: (1) target genes that were repeatedly found in the same library and (2) target genes that were repeatedly found in different libraries. For the N_9 sgRNA library, 460 different sites were identified in the *E. coli* genome by querying 46 TRS (Figure 3D and Table S4). In particular, six genes (*ptsG*, *pykA*, *pykF*, *sdaA*, *tnaA*, and *tyrR*) were identified more than five times (Figure 3D and Figure S4). However, no repeated target genes were found in the N_{10}, N_{11}, and N_{12} libraries. Next, we selected overlapping target genes among different sgRNA libraries. Although *secA* and *ligA* were found as target genes four and three times, respectively (Table S4), they were excluded from target selection for further gene knockout because they are known to be essential genes in cell growth in complex medium.[20] Finally, 17 target genes were selected based on the above criteria, and each target gene was not biased in the genome (Figure S5 and Table S5). In five genes (*cra*, *ptsG*, *pykA*, *pykF*, and *tnaA*), target sequences were identified only in the promoters. In seven genes (*galK*, *gltD*, *livG*, *metN*, *purL*, *rbsA*, and *ubiI*), target sequences were identified only in the middle of the structural region. In five genes (*ppsR*, *sad*, *sdaA*, *sdaC*, and *tyrR*), target sequences were identified in both the promoter and structural regions (Figure S6).

Engineering of Target Genes for Enhanced Violacein Production. We tested whether violacein production increased with the repression of the target gene expression. Seventeen cells with a single deletion of each target gene were grown in culture medium to compare the level of violacein production with that of wild-type SH148. All single-gene-deleted strains formed dark purple colonies compared to the wild-type SH148 strain on LB agar supplemented with D-xylose (Figure 4A). In the presence of D-xylose, the broth culture of wild-type SH148 produced 24.2 μg/mL violacein, and all 17 strains showed a higher violacein concentration than the SH148 strain (Figure 4B). Of the 17 target genes, 7 genes (*tyrR*, *pykF*, *cra*, *ptsG*, *pykA*, *sdaA*, and *tnaA*) are known to increase the intracellular concentration of L-tryptophan, a precursor of violacein. All 17 strains exhibited a significant increase in violacein production. Notably, single deletion of *livG* enhanced violacein production to 82.6 μg/mL, which was a 3.4-fold increase compared to that of the wild-type cells (Figure 4B). These results show that violacein production can be greatly increased by deleting novel target genes.

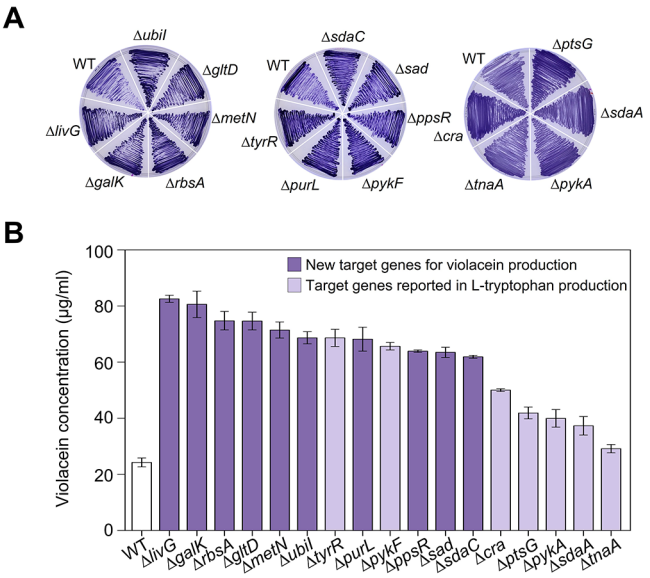


Figure 4. Enhanced violacein production in target-gene knockout strains. (A) Purple violacein accumulation in target-gene single-knockout strains on LB agar containing L-arabinose and D-xylose. (B) Quantitative analysis of violacein production in target-gene single-knockout cells grown in liquid culture.

DISCUSSION

The regulation of gene expression or deletion of a gene from the genome can help to understand the function of a specific gene in a cell. Among the various tools for functional genomics, widely used transposon mutagenesis[21] has the disadvantage of potential bias toward a specific nucleotide sequence, and essential genes cannot be disrupted. However, CRISPRi can modulate the expression of target genes using specific sgRNAs without damaging the genome.[22] CRISPRi can not only completely repress the expression of target genes but also conditionally regulate essential genes for functional analysis. CRISPR screening was developed as a screening method to identify the target genes responsible for various phenotypes.[23] In this study, we proposed a new CRISPR screening method for identifying microbial target genes by reducing the library size using a 5′-shortened random TRS library.

Generally, the size of a library is as important as its quality in screening methods. In the case of CRISPR-Cas9, the lengths of the TRS of sgRNAs were 20 nt, and the sizes of random TRS with all possible nucleotide sequences were 4^{20} (~10^{12}). Therefore, most TRS constituting random sgRNA libraries do not exist in the genomic sequences (~10^6) of certain microorganisms. We revisited and determined the minimum length of TRS required for transcriptional repression to prepare a library of appropriate size for screening in microorganisms. A previous study reported that the length of TRS of sgRNAs required for the repression of gene expression in dCas9 was 12 nt.[10] Cells with the Xyl$^-$ phenotype (white colonies on MacConkey agar) were observed, even with a 9 nt-long TRS of the dCas9-NG/shortened sgRNA complex targeting the *xylA* promoter (Figure 1C). In addition, RT-qPCR demonstrated that the transcription of *xylA* could be effectively repressed by dCas9-NG/sgRNA with a 9 nt-long TRS (Figure 1D).

Recent pooled or arrayed CRISPRi screening methods have had small library sizes (10^2–10^4) because oligonucleotides are

https://doi.org/10.1021/acssynbio.2c00595
ACS Synth. Biol. 2023, 12, 800−808

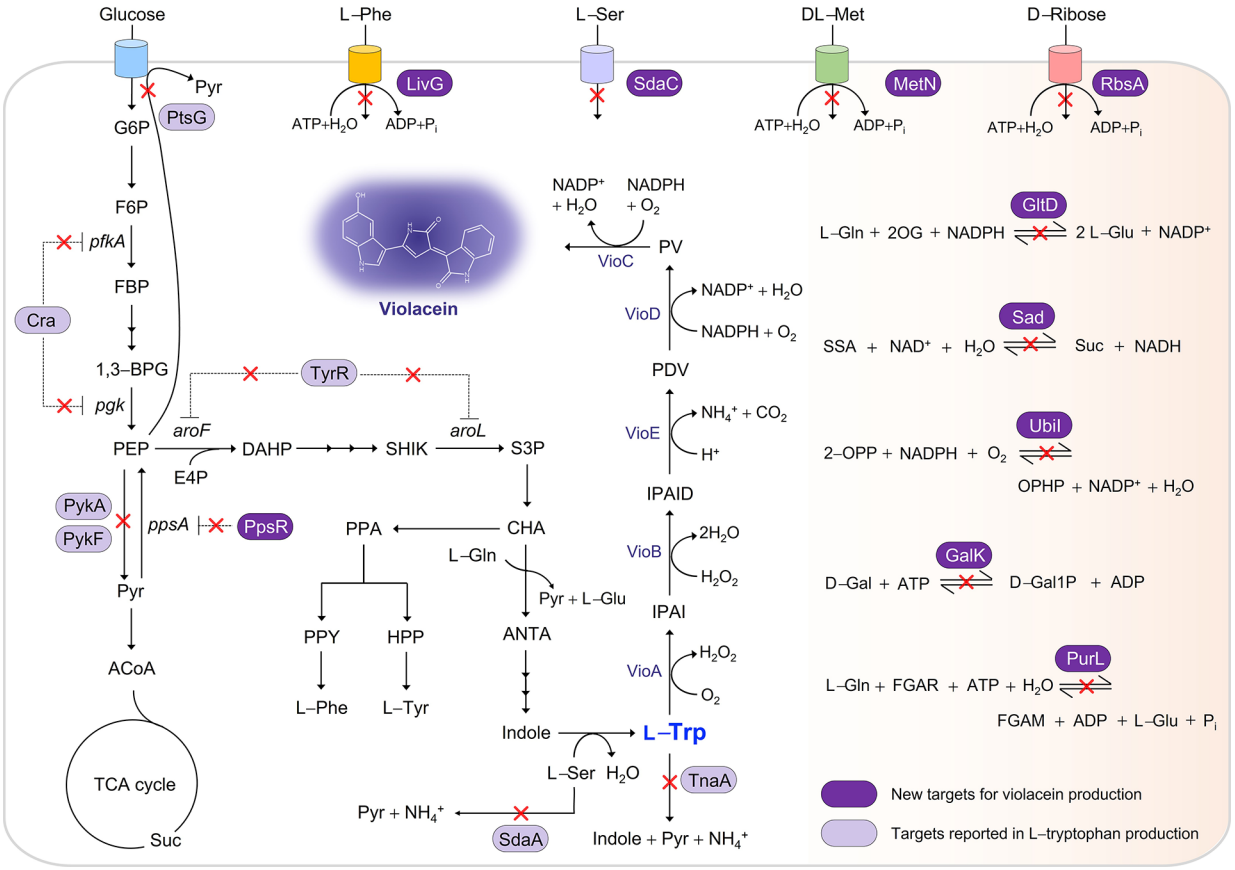


Figure 5. Metabolic pathways of target genes for violacein production selected by CRISPRi screening. Gray oval: targets reported in the production of L-tryptophan, a precursor of violacein; purple oval: target genes newly identified for violacein production. A red cross mark indicates the inhibition of molecular and biochemical reactions. ACoA, acetyl-CoA; ANTA, anthranilate; 1,3-BPG, 1,3-bisphosphoglycerate; CHA, chorismite; DAHP, 3-deoxy-D-arabino-heptulosonate-7-phosphate; E4P, erythrose 4-phosphate; FBP, fructose-1,6-bisphosphate; FGAM, 2-(formamido)-N^1-(5-phospho-β-D-ribosyl)acetamidine; FGAR, N^2-formyl-N^1- (5-phospho-β-D-ribosyl)glycinamide; F6P, fructose-6-phosphate; D-Gal, D-galactose; D-Gal1P, D-galactose-1-phosphate; G6P, glucose-6-phosphate; L-Glu, L-glutamate; L-Gln, L-glutamine; HPP, 4-hydroxyphenylpyruvate; IPAI, indole-3-pyruvic acid imine; IPAID, indole-3-pyruvic acid imine dimer; DL-Met, DL-methionine; 2OG, 2-oxoglutarate; OPHP, 2-octaprenyl-6-hydroxyphenol; 2-OPP, 2-octaprenylphenol; PDV, prodeoxyviolacein; PEP, phosphoenolpyruvate; PPA, prephenate; PPY, phenylpyruvate; L-Phe, L-phenylalanine; PV, proviolacein; Pyr, pyruvate; L-Ser, L-serine; S3P, Shikimate 3-phosphate; SHIK, Shikimate; SSA, succinate semialdehyde; Suc, succinate; L-Trp, L-tryptophan; L-Tyr, L-tyrosine.

synthesized for targeted genes using microarrays.[13,14] However, the design and synthesis of many oligonucleotides are required based on genome information.[24,25] For random sgRNA library generation, genome-wide screening is possible at a low cost and in a relatively shorter amount of time. Our random sgRNA library with shortened TRS was prepared through PCR using two oligonucleotides with phosphorylated 5′ ends (Figure 2A). Theoretically, 4^9 ($\sim 2.5 \times 10^5$) different types of random sgRNA can be generated for N_9 sgRNA production. We obtained 6.0×10^4 transformants through transformation with 1 μg of ligated DNA (Table S1). Assuming that all transformants have different TRS sequences, and considering the targetable genome size (1/4 of the *E. coli* genome) with 5′-NG PAM restriction, the library coverage is calculated as 0.46, which means that 46% of the targetable sequences in the *E. coli* genome can be covered by the library of 4^9 different types of random sgRNAs used in the CRISPRi screening.

In order to evaluate the library quality, we selected 10 different colonies and confirmed different TRSs by Sanger sequencing (Figure 2B) and also verified whether the shortened sgRNA library could be used to identify the genes involved in xylose metabolism (Figure 2B). Of the 22 TRS with different nucleotide sequences, all 22 were targeted to the regulatory region and structural gene of genes involved in xylose catabolism. If an appropriate phenotype such as Xyl is not available, NGS can be performed to confirm the library diversity for target gene screening. Five of the 22 TRS were targeted to the promoter and 17 to the open reading frames (ORFs) in the xylose operon (Figure 2C and Table S2). RT-qPCR confirmed that a 9 nt-long TRS was of sufficient length to effectively inhibit gene expression by binding to the ORF (Figure 2D,E). These results demonstrate that the Cas9-NG/shortened sgRNA complex can change the phenotype when targeting the ORF and promoter. Therefore, we confirmed that if there is an observable phenotype, the target gene can be screened using the 5′-shortened random sgRNA library.

To determine whether the sgRNA library method can identify genes related to the phenotype, we screened genes related to the production of violacein in the recombinant strain. Since violacein has various physiologically active functions, such as anticancer, antifungal, and antibacterial activities, much attention has been paid to its production.[26−29] Target genes that increase violacein production were screened using dCas9-NG. The precise gene that produces a phenotype

https://doi.org/10.1021/acssynbio.2c00595
ACS Synth. Biol. 2023, 12, 800−808

among the genes selected by library screening may not be determined because a single shortened sgRNA can bind to several target sequences in the genome. Therefore, we considered a method that uses the redundancy of genes found in various libraries when selecting target genes. If a target candidate gene is repeatedly seen in the list of target genes of different TRS, it is thought to be a target gene related to the phenotype.

Target genes were selected by analyzing the redundancy of the targets found in different TRS and subsequently excluding two essential genes (*secA* and *ligA*) for further knockout experiments (Figure 3D and Table S4). We identified 17 target genes, including three NAD(*P*)H-dependent oxidoreductases (*gltD*, *ubiI*, and *sad*), three transcription factors (*cra*, *ppsR*, and *tyrR*), and five transporters (*ptsG*, *sdaC*, *metN*, *livG*, and *rbsA*) (Figure 5 and Table S5). Previous studies have reported that 7 genes (*tnaA*, *sdaA*, *pykA*, *pykF*, *ptsG*, *tyrR*, and *cra*) of the selected 17 genes are target genes that increase the concentration of L-tryptophan.[30−32] Therefore, it was thought that the target gene could be effectively identified based on the results obtained using the shortened sgRNA library method. We confirmed that all strains in which 17 target genes were deleted increased violacein production in liquid culture (Figure 4B). Some of the target genes obtained in this study are directly related to violacein biosynthesis, but it is unclear how the other target genes are involved. It may be hypothesized that it is helpful to maintain high intracellular concentrations of L-tryptophan. For example, chorismate is a common metabolic intermediate in the biosynthesis of aromatic amino acids, including L-tryptophan, L-tyrosine, and L-phenylalanine.[33] Presumably, when the L-phenylalanine transporter LivG is removed, an increase in the intracellular chorismate pool leads to an increase in L-tryptophan biosynthesis and subsequent violacein biosynthesis.[34]

In summary, we identified genes associated with phenotypic changes in ∼10^4 cells by reducing the library size using the 5′-shortened TRS library method. Although several target candidates for each short TRS may be found, target genes can be specified by analyzing target redundancy, as shown in the violacein model. Our method does not require the design and synthesis of a large number of oligonucleotides as in recent pooled or arrayed CRISPR screening studies. If the CRISPR system is established and a replicating plasmid is available, the principle of library design can be used in other bacteria. The 5′-shortened sgRNA library method can identify genes associated with a specific phenotype in the genome under certain conditions or environments, determine the unknown roles of the genes in microbial cells, and direct the cellular system in a beneficial direction.

METHODS

Bacterial Strains and Culture Conditions. The bacterial strains used in this study are listed in Table S6. The *Massilia* sp. NR4-1 strain (KCTC 42724) was obtained from the Korean Culture Type Collection (KCTC). The strain was streaked and grown on R2A medium (MBcell, MB-R2230) at 25 °C under aerobic conditions. *E. coli* strains were streaked and grown on LB medium (LPS Solution, LB-05). The *E. coli* DH5*α* strain was used as the cloning host for sgRNA construction. The *E. coli* MG1655 strain was used for sgRNA library screening using a dCas9 variant with the 5′-NG sequence as PAM (dCas9-NG) and violacein production. MacConkey agar supplemented with 0.3% D-xylose was used to confirm whether D-xylose had been metabolized. When necessary, the antibiotic concentrations used included 50 *μ*g/mL ampicillin, 10 *μ*g/mL chloramphenicol, 25 *μ*g/mL kanamycin, and 75 *μ*g/mL spectinomycin.

sgRNA Construction. All the constructs used in this study are listed in Table S7. The sequences of the primers used in this study are listed in Table S8. PCR was performed using KOD FX (ToYoBo, KFX-101). sgRNA plasmids (5′-shortened; N_7, N_8, N_9, N_{11}, N_{14}, N_{17}, and N_{20}) for the *xylA* promoter were cloned by isothermal assembly using the Gibson Assembly Master Mix (NEB, E2611) using two PCR fragments: to construct the pSH035 plasmid, the TRS and partial spectinomycin resistance gene were amplified using pHK459[35] as a template with the Pxyl-20NG-F and Sm-ATGout primer sets. The pBR322 origin and partial spectinomycin resistance gene were amplified using pHK459 as a template with the Sm-TAAout and Pxyl-20NG-R primer sets. The pSH086, pSH087, pSH088, pSH036, pSH037, pSH038, and pSH039 plasmids were constructed in a similar manner. The TRS and lengths were included in the primers.

Random sgRNA libraries (N_8, N_9, N_{10}, N_{11}, N_{12}, and N_{20}) were constructed using random primers. Random primers (5′-phosphorylated) were synthesized with different lengths of random regions depending on the length of the TRS. Two fragments amplified using random primers were treated with the DpnI restriction enzyme to remove the template and be purified. The two purified fragments were then ligated using T4 ligase (Invitrogen, 15224-017) and transformed into DH5*α* competent cells. All transformants were harvested, and the plasmids were purified to obtain the random sgRNA library. To verify the library quality, we randomly selected colonies in constructed random sgRNA libraries (TRS = N_8, N_9, N_{10}, N_{11}, N_{12}, and N_{20}) and confirmed the TRS sequence of the sgRNA. All ten TRSs in the random libraries were identified as different sequences.

Genomic Construction. Violacein is a water-insoluble purple pigment found in various Gram-negative bacteria.[36,37] The violacein biosynthetic gene cluster (*vioABCDE*, 7.8 kb) was amplified from the genome of *Massilia* sp. NR4-1 strain[38] and designed to be induced by the presence of D-xylose and integrated into the *xylB* region of the *E. coli* MG1655 genome by repetitive homologous recombineering. DNA fragments (∼1 kb) containing a partial fragment of the *vioA* gene were PCR-amplified using *Massilia* sp. NR4-1 genomic DNA as a template and fused with a kanamycin resistance gene (KmR) using overlap PCR to generate a partial *vioA*-KmR cassette (∼2.5 kb). The purified PCR products were then electroporated into L-arabinose-induced *E. coli* MG1655 harboring the plasmid pKD46[39] to generate the SH145 strain. Next, DNA fragments (∼3.6 kb) containing a partial *vioAB* fragment (∼2.5 kb) and a chloramphenicol-resistant gene (∼1.1 kb) were amplified and electroporated into the L-arabinose-induced SH145 strain harboring the pKD46 plasmid to generate the SH146 strain. In the same manner, we integrated other violacein biosynthetic DNA fragments into the genome of *E. coli* MG1655 and constructed the SH148 strain, in which the structural *vioABCDE* genes of violacein biosynthesis were transcriptionally fused to the *xylA* gene (*xylA*∼*vioABCDE*-CmR).

The P1*vir* phage was used to transfer mutations to other strains via standard P1 transduction. The P1 lysates of chloramphenicol-resistant *E. coli* SH148 cells were used to transduce the HK1160 (*araBAD*::*dcas9-NG*-KmR)[40] strain to

https://doi.org/10.1021/acssynbio.2c00595
ACS Synth. Biol. 2023, 12, 800−808

generate the SH150 strain. Target genes identified from the sgRNA library were knocked out using the following method. The P1 lysates of kanamycin-resistant strains from the Keio collection[41] were used to transduce *E. coli* MG1655 as the recipient strain.

Library Screening. Six random sgRNA libraries (N_8, N_9, N_{10}, N_{11}, N_{12}, and N_{20}) (~1 μg) were electroporated into HK1160 competent cells. Electroporation was performed at 25 μF, 200 Ω, and 1.8 kV, and a 0.1 cm electroporation cuvette was used. Immediately after that, the cell was transferred to 950 mL of SOC medium and incubated at 37 °C and 180 rpm for a 1 h recovery period. The cell was then spread on MacConkey agar (BD Difco, 281,810) containing D-xylose (0.3%) (Junsei, 64220S0650), L-arabinose (0.3%) (TCI, A0515), and spectinomycin (75 μg/mL) in a square dish (245 × 245 × 20 mm; SPL, 11245) and incubated at 37 °C for 12 h. White colonies in which xylose metabolism was suppressed by CRISPRi were selected, isolated, and cultured among transformants (~10^4 colonies). The sgRNA plasmids of the candidates were extracted and analyzed by Sanger sequencing to confirm the TRS. The target genes were determined by comparing the TRS of the *E. coli* MG1655 genome sequence (NC_000913). The library coverage was defined as: the library coverage = (the length of random TRS × the number of transformants)/targetable genome size.

The six random sgRNA libraries (N_8, N_9, N_{10}, N_{11}, N_{12}, and N_{20}) (~1 μg) were then electroporated into SH150 competent cells. Electroporation was performed as described above. The cells were spread on LB agar containing D-xylose (0.3%), L-arabinose (0.3%), and spectinomycin (75 μg/mL) in a square dish and incubated at 25 °C for 48 h. Dark purple colonies were selected, isolated, and cultured using the RGB values in the JPEG image files. The RGB values were statistically analyzed using Origin 2020 (OriginLab Corp., USA).

Transcript Analysis. The expression levels of *xylA* were measured using RT-qPCR. The HK1160 cells carrying the 5′-shortened sgRNAs (pHL308, pSH035, pSH036, pSH037, pSH038, pSH039, pSH086, pSH087, and pSH088) were grown at 37 °C in LB broth containing spectinomycin. At an OD 600 nm of 0.3, L-arabinose (0.3%) was added for dCas9-NG expression. The cells were harvested after 2 h, and total RNA was isolated using the RNeasy Mini Kit (Qiagen, 74,104). Primers for the target *xylA* gene were designed by Eurofins Genomics (https://eurofinsgenomics.eu/en/ecom/tools/qpcr-assay-design/). RT-qPCR was carried out on a CFX Connect system (Bio-Rad, 1,855,201) using the RealHelix qPCR Kit (NanoHelix, QRT-S500). Five nanograms of total RNA was used in the RT-qPCR reactions under the following conditions: cDNA synthesis (50 °C, 40 min), denaturation (95 °C, 12 min), and amplification for 40 cycles (95 °C, 20 s; 60 °C, 1 min). Raw fluorescence data were normalized to 16S ribosomal RNA expression levels. To calculate the relative abundance of the *xylA* gene transcript, the mRNA levels of the gene with various sgRNA plasmids were divided by the mRNA levels of the corresponding gene with sgRNA-deleted plasmids.

Metabolite Analysis. For violacein extraction, cells were harvested by centrifugation (12,000 rpm, 4 °C, 10 min), and the supernatant was discarded. Methanol (100%; the same volume as the culture broth) was added to the collected cell pellet, mixed vigorously by vortex, and incubated at 25 °C for 20 min. After extracting the methanol, the cell pellets were harvested and discarded after centrifugation (12,000 rpm, 4 °C, 10 min). The concentration of violacein in the supernatant was measured by high-performance liquid chromatography (HPLC) on an Agilent 1100 Series HPLC system using a C18 column (100 Å, 5 μm, 4.6 × 250 mm; Waters Corp., WAT05427). Ten microliters of the reaction sample was injected for the HPLC analysis. The ternary mobile phase was composed of methanol, acetonitrile, and distilled water (1:1:2, v/v). The pH was adjusted to 3.6 with acetic acid. Isocratic elution was performed at 30 °C with a flow rate of 0.5 mL/min, and violacein was detected using ultraviolet light at 575 nm. Violacein standards were purchased from Sigma-Aldrich (V9389).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acssynbio.2c00595.

Additional tables and figures including strains, primers, plasmids, phenotype-based CRISPRi screening results, and Sanger sequences (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Sang Jun Lee − *Department of Systems Biotechnology, and Institute of Microbiomics, Chung-Ang University, Anseong 17546, Republic of Korea;* orcid.org/0000-0002-2803-753X; Email: sangjlee@cau.ac.kr

Authors

Song Hee Jeong − *Department of Systems Biotechnology, and Institute of Microbiomics, Chung-Ang University, Anseong 17546, Republic of Korea*

Hyun Ju Kim − *Department of Systems Biotechnology, and Institute of Microbiomics, Chung-Ang University, Anseong 17546, Republic of Korea;* orcid.org/0000-0002-5852-9578

Complete contact information is available at: https://pubs.acs.org/10.1021/acssynbio.2c00595

Author Contributions

S.H.J. and H.J.K. contributed to the study design, performed experiments and data analysis, and wrote the manuscript. S.J.L. contributed to the study design, data analysis, and manuscript writing. All authors read and approved the final manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (2021R1A2C1013606), Republic of Korea.

■ REFERENCES

(1) Hryhorowicz, M.; Lipiński, D.; Zeyland, J.; Słomski, R. CRISPR/Cas9 Immune system as a tool for genome engineering. *Arch. Immunol. Ther. Exp.* **2017**, *65*, 233−240.

(2) Barrangou, R.; Marraffini, L. A. CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Mol. Cell* **2014**, *54*, 234−244.

(3) Zhang, Y.; Bai, Y.; Wu, G.; Zou, S.; Chen, Y.; Gao, C.; Tang, D. Simultaneous modification of three homoeologs of *TaEDR1* by genome editing enhances powdery mildew resistance in wheat. *Plant J.* **2017**, *91*, 714−724.

(4) Fogarty, N. M. E.; McCarthy, A.; Snijders, K. E.; Powell, B. E.; Kubikova, N.; Blakeley, P.; Lea, R.; Elder, K.; Wamaitha, S. E.; Kim, D.; et al. Genome editing reveals a role for OCT4 in human embryogenesis. *Nature* **2017**, *550*, 67−73.
(5) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **2012**, *337*, 816−821.
(6) Oishi, I.; Yoshii, K.; Miyahara, D.; Tagami, T. Efficient production of human interferon beta in the white of eggs from ovalbumin gene-targeted hens. *Sci. Rep.* **2018**, *8*, 10203.
(7) Lee, H. J.; Kim, H. J.; Park, Y. J.; Lee, S. J. Efficient Single-Nucleotide Microbial Genome Editing Achieved Using CRISPR/Cpf1 with Maximally 3′-End-Truncated crRNAs. *ACS Synth. Biol.* **2022**, *11*, 2134−2143.
(8) Lee, H. J.; Kim, H. J.; Lee, S. J. CRISPR-Cas9-mediated pinpoint microbial genome editing aided by target-mismatched sgRNAs. *Genome Res.* **2020**, *30*, 768−775.
(9) Makarova, K. S.; Wolf, Y. I.; Alkhnbashi, O. S.; Costa, F.; Shah, S. A.; Saunders, S. J.; Barrangou, R.; Brouns, S. J. J.; Charpentier, E.; Haft, D. H.; et al. An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **2015**, *13*, 722−736.
(10) Qi, L. S.; Larson, M. H.; Gilbert, L. A.; Doudna, J. A.; Weissman, J. S.; Arkin, A. P.; Lim, W. A. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **2013**, *152*, 1173−1183.
(11) Zhang, R.; Xu, W.; Shao, S.; Wang, Q. Gene silencing through CRISPR interference in bacteria: current advances and future prospects. *Front. Microbiol.* **2021**, *12*, No. 635227.
(12) Bock, C.; Datlinger, P.; Chardon, F.; et al. High-content CRISPR screening. *Nat. Rev. Methods Primers* **2022**, *2*, 8.
(13) Bodapati, S.; Daley, T. P.; Lin, X.; Zou, J.; Qi, L. S. A benchmark of algorithms for the analysis of pooled CRISPR screens. *Genome Biol.* **2020**, *21*, 62.
(14) Imkeller, K.; Ambrosi, G.; Boutros, M.; Huber, W. gscreend: modelling asymmetric count ratios in CRISPR screens to decrease experiment size and improve phenotype detection. *Genome Biol.* **2020**, *21*, 53.
(15) Liu, J.; Liu, M.; Shi, T.; Sun, G.; Gao, N.; Zhao, X.; Guo, X.; Ni, X. M.; Yuan, Q. Q.; Feng, J. H.; et al. CRISPR-assisted rational flux-tuning and arrayed CRISPRi screening of an L-proline exporter for L-proline hyperproduction. *Nat. Commun.* **2022**, *13*, 891.
(16) Yao, L.; Shabestary, K.; Björk, S. M.; Asplund-Samuelsson, J.; Joensson, H. N.; Jahn, M.; Hudson, E. P. Pooled CRISPRi screening of the cyanobacterium *Synechocystis* sp. PCC 6803 for enhanced industrial phenotypes. *Nat. Commun.* **2020**, *11*, 1666.
(17) Agrotis, A.; Ketteler, R. A new age in functional genomics using CRISPR/Cas9 in arrayed library screening. *Front. Genet.* **2015**, *6*, 300.
(18) Fang, L.; Fan, J.; Luo, S.; Chen, Y.; Wang, C.; Cao, Y.; Song, H. Genome-scale target identification in *Escherichia coli* for high-titer production of free fatty acids. *Nat. Commun.* **2021**, *12*, 4976.
(19) Wang, T.; Guo, J.; Liu, Y.; Xue, Z.; Zhang, C.; Xing, X. H. Genome-wide screening identifies promiscuous phosphatases impairing terpenoid biosynthesis in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **2018**, *102*, 9771−9780.
(20) Goodall, E. C. A.; Robinson, A.; Johnston, I. G.; Jabbari, S.; Turner, K. A.; Cunningham, A. F.; Lund, P. A.; Cole, J. A.; Henderson, I. R. The essential genome of *Escherichia coli* K-12. *mBio* **2018**, *9*, e02096-17.
(21) van Opijnen, T.; Bodi, K. L.; Camilli, A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat. Methods* **2009**, *6*, 767−772.
(22) Larson, M. H.; Gilbert, L. A.; Wang, X.; Lim, W. A.; Weissman, J. S.; Qi, L. S. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* **2013**, *8*, 2180−2196.
(23) Gilbert, L. A.; Horlbeck, M. A.; Adamson, B.; Villalta, J. E.; Chen, Y.; Whitehead, E. H.; Guimaraes, C.; Panning, B.; Ploegh, H. L.; Bassik, M. C.; et al. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **2014**, *159*, 647−661.
(24) Yates, J. D.; Russell, R. C.; Barton, N. J.; Yost, H. J.; Hill, J. T. A simple and rapid method for enzymatic synthesis of CRISPR-Cas9 sgRNA libraries. *Nucleic Acids Res.* **2021**, *49*, No. e131.
(25) Wang, T.; Guan, C.; Guo, J.; Liu, B.; Wu, Y.; Xie, Z.; Zhang, C.; Xing, X. H. Pooled CRISPR interference screening enables genome-scale functional genomics study in bacteria with superior performance. *Nat. Commun.* **2018**, *9*, 2475.
(26) Andrighetti-Fröhner, C. R.; Antonio, R. V.; Creczynski-Pasa, T. B.; Barardi, C. R. M.; Simões, C. M. O. Cytotoxicity and potential antiviral evaluation of violacein produced by *Chromobacterium violaceum*. *Mem. Inst. Oswaldo Cruz* **2003**, *98*, 843−848.
(27) Kanelli, M.; Mandic, M.; Kalakona, M.; Vasilakos, S.; Kekos, D.; Nikodinovic-Runic, J.; Topakas, E. Microbial production of violacein and process optimization for dyeing polyamide fabrics with acquired antimicrobial properties. *Front. Microbiol.* **2018**, *9*, 1495.
(28) Sasidharan, A.; Sasidharan, N. K.; Amma, D. B. N. S.; Vasu, R. K.; Nataraja, A. V.; Bhaskaran, K. Antifungal activity of violacein purified from a novel strain of *Chromobacterium* sp. NIIST (MTCC 5522). *J. Microbiol.* **2015**, *53*, 694−701.
(29) Fang, M. Y.; Zhang, C.; Yang, S.; Cui, J. Y.; Jiang, P. X.; Lou, K.; Wachi, M.; Xing, X. H. High crude violacein production from glucose by *Escherichia coli* engineered with interactive control of tryptophan pathway and violacein biosynthetic pathway. *Microb. Cell Fact.* **2015**, *14*, 8.
(30) Niu, H.; Li, R.; Liang, Q.; Qi, Q.; Li, Q.; Gu, P. Metabolic engineering for improving L-tryptophan production in *Escherichia coli*. *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 55−65.
(31) Liu, L.; Bilal, M.; Luo, H.; Zhao, Y.; Iqbal, H. M. N. Metabolic engineering and fermentation process strategies for L-tryptophan production by *Escherichia coli*. *Processes* **2019**, *7*, 213.
(32) Panichkin, V. B.; Livshits, V. A.; Biryukova, I. V.; Mashko, S. V. Metabolic engineering of *Escherichia coli* for L-tryptophan production. *Appl. Biochem. Microbiol.* **2016**, *52*, 783−809.
(33) Liu, X.; Niu, H.; Li, Q.; Gu, P. Metabolic engineering for the production of L-phenylalanine in *Escherichia coli*. *3 Biotechnol.* **2019**, *9*, 85.
(34) Nazos, P. M.; Mayo, M. M.; Su, T. Z.; Anderson, J. J.; Oxender, D. L. Identification of LivG, a membrane-associated component of the branched-chain amino-acid transport in *Escherichia coli*. *J. Bacteriol.* **1985**, *163*, 1196−1202.
(35) Kim, B.; Kim, H. J.; Lee, S. J. Regulation of microbial metabolic rates using CRISPR interference with expanded PAM sequences. *Front. Microbiol.* **2020**, *11*, 282.
(36) Blosser, R. S.; Gray, K. M. Extraction of violacein from *Chromobacterium violaceum* provides a new quantitative bioassay for *N*-acyl homoserine lactone autoinducers. *J. Microbiol. Methods* **2000**, *40*, 47−55.
(37) Choi, S. Y.; Yoon, K. H.; Lee, J. I.; Mitchell, R. J. Violacein: properties and production of a versatile bacterial pigment. *Biomed Res. Int.* **2015**, *2015*, No. 465056.
(38) Myeong, N. R.; Seong, H. J.; Kim, H. J.; Sul, W. J. Complete genome sequence of antibiotic and anticancer agent violacein producing *Massilia* sp. strain NR 4-1. *J. Biotechnol.* **2016**, *223*, 36−37.
(39) Datsenko, K. A.; Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 6640−6645.
(40) Kim, B.; Kim, H. J.; Lee, S. J. Effective blocking of microbial transcriptional initiation by dCas9-NG-mediated CRISPR interference. *J. Microbiol. Biotechnol.* **2020**, *30*, 1919−1926.
(41) Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K. A.; Tomita, M.; Wanner, B. L.; Mori, H. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2006**, *2*, 2006.0008.

https://doi.org/10.1021/acssynbio.2c00595
ACS Synth. Biol. 2023, 12, 800−808