Undergraduate Journal of Experimental Microbiology and Immunology (UJEMI) Vol. 27:1-12
Copyright © September 2022, M&I UBC

Cloning optimization for substrate-induced gene expression technology

Kyle MacPherson, Jaskirat Malhi, Nemat Haroon, Andi Musa

Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, Canada

SUMMARY Substrate-induced gene expression (SIGEX) is a genetic technique utilized for the isolation of novel substrate-inducible catabolic genes from environmental metagenomic samples. Previous SIGEX plasmids have relied on a single reporter gene for high-throughput detection of novel genes in metagenomic libraries. This limited its use for genetic inserts in the opposite direction of the reporter gene which theoretically occurs in half of all inserts. A recent study by Abrishamkar *et al.* described the creation of a new duo-directional SIGEX-based plasmid named pSPPH21 by including two fluorescent reporter genes in opposite orientations. This newly created plasmid has yet to be validated in a SIGEX-based experiment. Here we aimed to test the functionality of this plasmid using an inserted inducible *lac* operon promoter sequence. We were unable to clone the insert however, several strategies were researched and utilized for cloning. We have summarized these strategies and we hope this will serve as a resource for optimizing future implementation of SIGEX technology.

INTRODUCTION

Single-celled organisms have diverse metabolic capabilities allowing them to break down a wide range of substrates. This makes microbial catabolic pathways an attractive target for the degradation of environmental pollutants and other substrates. Additionally, the role of microorganisms has been elucidated in human health ranging from general development, digestive functions, mood disorders, and formation of a well-adapted immune system (1). Given this important role, understanding microbial metabolism and how it contributes to the well-being of other life forms is key. Just as important is the study of their ability to withstand harsh environmental conditions in distinct niches around the world. However, it is estimated that more than 99% of microbial species are difficult to culture in the lab by means of conventional microbiological techniques (2). This has led to the development of culture-independent methods including high throughput methods that integrate next-generation sequencing for the study of microorganisms that are unculturable (3). Culture-independent analysis has opened a new branch of biology termed metagenomics which is tasked with identifying previously unknown microorganisms from environmental samples, and the discovery and isolation of novel genes unique to these unculturable microorganisms (4).

Metagenomic methods are based on environmental samples that contain a mixture of genomes from various organisms, which are ultimately screened for genes of interest using a variety of methods. One category of genes that are of interest to microbiologists are the catabolic genes responsible for the breakdown of larger molecules for biological use in organisms. These genes can be identified using a sequence-based approach based on sequence similarity and annotation, however close to 40% of annotated genes in microbial genomes are of unknown function (5). Since the success of this method is based on the completeness of genomic databases and the corresponding accuracy of their genetic annotations, it is not well-suited for novel genes that have weak similarity to any previously described gene (5). These shortcomings can be addressed through a function-based metagenomics approach which is the preferred method when it comes to discovering genes with novel functions in microbial populations (7). This method entails in its simplest form the screening of colonies on agar plates for enzyme activity but has many variations including a high-throughput approach involving fluorescence-activated cell sorting (FACS) using reporter genes (7).

Substrate-induced gene expression (SIGEX) is one such method with a function-based approach. The foundation of this method is that catabolic gene expression is most likely induced through the target substrate that will be broken down and that expression is regulated

Published Online: September 2022

Citation: Kyle MacPherson, Jaskirat Malhi, Nemat Haroon, Andi Musa. 2022. Cloning optimization for substrate-induced gene expression technology. UJEMI 27:1-12

Editor: Andy An and Gara Dexter, University of British Columbia

Copyright: © 2022 Undergraduate Journal of Experimental Microbiology and Immunology. All Rights Reserved.

Address correspondence to: https://jemi.microbiology.ubc.ca/

by nearby regulatory elements (8). SIGEX is designed to select clones with active catabolic genes through reporter genes such as green fluorescent protein (GFP). After substrate introduction, FACS and Sanger sequencing are used to isolate and identify substrate-induced promoters (9). Promoter trap vectors allow for shotgun cloning across many host samples in liquid culture, thus permitting high-throughput screening (8). Combined with steps to remove self-ligated clones and clones expressing GFP constitutively, the SIGEX method provides an efficient method of high-throughput screening for catabolic gene expression (9). However, it is well known that some regulatory elements can be distantly located from the target gene at distances of up to 1000kb from the gene of interest – thus preventing gene identification through SIGEX (8, 10). Genes that are constitutively expressed and not affected by exogenous substrates would similarly not be considered within the applicability of SIGEX (8). Lastly, SIGEX is sensitive to the orientation of the inserted gene since the gene must be oriented in the same direction as the reporter gene in order to be detected (9).

Recently, a study described the creation of a new plasmid whose design addresses the uni-directional limitation of SIGEX by including a second fluorescent reporter gene in the reverse orientation to aid in the detection of clones where the gene is inserted in the opposite direction (11). This new plasmid – named pSPPH21 – contains both a red fluorescent protein (RFP) and a green fluorescent protein (GFP) reporter gene in opposite orientations. However, this newly-created plasmid has yet to be validated in a SIGEX-based experiment. Given the utility of this duo-directional plasmid in addressing one of the limitations of SIGEX, validating its use would increase its usefulness in future metagenomic studies by expanding the range of detection for SIGEX based experiments involving catabolic genes.

In this study we aimed to test the functionality of the pSPPH21 plasmid within a SIGEX experiment by inserting an IPTG-inducible *lac* promoter and operator sequence within the pSPPH21 plasmid. Induction of the lac operon with IPTG, which is a molecular mimic of allolactose, should lead to equal expression of both reporter genes across many samples given that a blunt ended DNA insertion should occur equally in both orientations. We hypothesized that the pSPPH21 SIGEX duo-directional reporter plasmid would ligate with a blunt ended DNA product encoding the lac promoter and operator sequences in equal frequency with respect to orientation. Transformed colonies of *Escherichia coli* would express a 1:1 ratio of green and red fluorescent proteins upon IPTG induction. However, our cloning attempts were not successful after several attempts suggesting that cloning metagenomic DNA into this vector may be more complicated than initially anticipated. We present a series of optimized procedures with the goal of increasing the yield of successful transformants in future experiments. While none of these modifications were successful in this study, previous authors have confirmed their utility in optimizing blunt-end cloning, thus making these optimizations a useful resource for future researchers looking to use SIGEX technology to screen metagenomic DNA samples.

METHODS AND MATERIALS

Bacterial strains, plasmids and gene fragments. The original pSPPH21 plasmid was isolated from *E. coli* DH5α left over from Abrishamkar *et al.* pSPPH21 was maintained in *Escherichia coli* DH5α grown on Luria-Bertani (LB) agar containing the antibiotic Chloramphenicol (25ug/mL) at 37°C. pSPPH21 plasmid DNA was extracted using the BioBasic™ EZ-10 Spin Column Plasmid DNA Miniprep Kit and stored in 10mM Tris-HCl pH 8 (EB buffer) at -20°C. For transformations, we used competent *E. coli* DH5α, Invitrogen™ Subcloning Efficiency™ DH5α Competent Cells, and OneShot™ TOP10 Chemically Competent *E. coli*. To validate the promoter trap activity of the pSPPH21 plasmid, we used a known substrate-induced promoter – the *lac* promoter (*lacp)*. The *lac* promoter was custom ordered as part of a gBlocks™ gene fragment from Integrated DNA Technologies™ (IDT) which contains the entire *lac* operon regulatory region – the *lac* promoter, operator and a CAP binding site. This gene fragment was resuspended in 25uL of EB buffer (final concentration 10ng/uL), vortexed thoroughly, incubated at 50°C for 20 minutes, and stored at -20°C – as recommended by the supplier (IDT).

Restriction digestion. We linearized the pSPPH21 plasmid using the NruI restriction enzyme (New England BioLabs). We also used Shrimp Alkaline Phosphatase to dephosphorylate the

5' ends of the newly digested pSPPH21 plasmid to prevent them from re-annealing without ligated insert (New England BioLabs). The 50uL digestion reaction was set up according to manufacturer instructions, as follows (New England BioLabs): input DNA (1ug), NruI Restriction Enzyme (1uL, 10 units), 10X NEBuffer™ (5uL), Shrimp Alkaline Phosphatase (2.5uL), 10X rCutSmart Buffer™ (5uL), Nuclease-free water (to a final volume of 50uL).

Input DNA was used in excess to ensure the final product had enough pure, high quality, digested plasmid DNA. The thoroughly mixed reaction was incubated at 37°C for one hour, and then heat-inactivated at 65°C for 5 minutes to deactivate the Shrimp Alkaline Phosphatase. Since NruI cannot be heat-inactivated, the digested mix was then purified using the GeneJet™ PCR Purification Kit (ThermoFisher). Digested and phosphatase-treated DNA was eluted in EB buffer and quantified using the NanoDrop™ Spectrophotometer (ThermoFisher).

Blunt-end ligation. We then ligated the digested pSPPH21 plasmid with the *lac* regulatory region using T4 DNA Ligase (ThermoFisher). The 20uL ligation reaction was set up according to manufacturer instructions, as follows (ThermoFisher): 5X T4 DNA Ligase buffer (4uL), pSPPH21 DNA (20 fmol), *lac* regulator DNA (60 fmol), Nuclease-free water (to a final volume of 20 uL), T4 DNA Ligase (1 uL at a concentration of 1U/uL).

The ligation reaction was incubated at 14°C overnight. The next morning, 2uL of the ligation reaction was used to transform 50uL of competent cells. Digested pSPPH21 plasmid without insert was used as a positive control for ligation.

Bacterial transformation. Chemically competent *Escherichia coli* DH5α were prepared previously with a transformation efficiency of 7.4 x 10^5 CFU/ug DNA. We also used Invitrogen™ Subcloning Efficiency™ DH5α Competent Cells and OneShot™ TOP10 Chemically Competent *E. coli,* which had reported transformation efficiencies of 1 x 10^6 CFU/ug DNA and 1 x 10^9 CFU/ug DNA, respectively (ThermoFisher). For transformations, 50uL of cell suspension was mixed with 2uL of DNA. Cells were transformed using the heat shock method (12). Undigested pSPPH21 plasmid was used as a positive control and digested pSPPH21 plasmid was used as a negative control.

Agarose gel electrophoresis. Fifteen ligated colonies were selected for insert screening using agarose gel electrophoresis. The controls for this procedure were undigested and digested pSPPH21 plasmid. We prepared 5X TBE buffer by mixing 27g Tris base, 13.75g boric acid, and 10mL of 0.5M EDTA (pH 8) into 500mL of distilled water. We ran 1% agarose gels using 1g of agarose in 100mL 1X TBE. 75ng of DNA was diluted in 6X Gel Purple Loading Dye to a final 1X concentration and 10uL was loaded per well (New England BioLabs). We also loaded 10uL of the Invitrogen™ 1Kb Plus DNA Ladder (ThermoFisher). Electrophoresis ran at 100V for 60 minutes. DNA gels were visualized on the BioRad ChemiDoc MP using the ethidium bromide setting (BioRad).

SIGEX induction. To test substrate induction in the ligated plasmid, we used IPTG to induce colonies that passed restriction enzyme analysis. To do this, we incubated transformed bacteria in 5mL of LB liquid medium supplemented with 25ug/uL Chloramphenicol and 1mM IPTG at 37°C overnight (13). The control for this procedure was unligated bacteria with pSPPH21 supplemented with IPTG. After incubation, 1mL of the culture was centrifuged at 1,000 x g for 5 minutes. The supernatant was removed, and the culture was washed twice with PBS and then re-suspended in 1mL PBS to remove optical interference from the culture medium (13). The cell suspension was then transferred to a 96-well plate in triplicate 100uL aliquots with an OD_{600} reading less than 1. To visualize GFP, we set the excitation wavelength to 485nm and the emission filter to 525nm. To visualize RFP, we set the excitation wavelength to 580nm and emission filter to 620nm.

Sanger sequencing. To verify the presence and direction of the ligated *lac* regulatory region, we submitted DNA from transformed clones 5 and 6 for Sanger sequencing. These colonies were selected based on the results of our agarose gel electrophoresis screen. Samples were submitted according to vendor specifications (GENEWIZ). For each sample, plasmid DNA

was submitted in duplicate with a total of 400ng of DNA per submission. Each sample was sequenced with RFP forward primer (5'- GGC GTA TCA CGA GGC AGA ATT TC - 3') and RFP reverse primer (5'- GGA AGC CTG CAT AAC GCG AAG -3') (11). Sequencing results were visualized using the SnapGene viewer and analyzed for the presence of the *lac* regulatory region using the NCBI alignment tool.

RESULTS

To validate and characterize the duo-directional promoter trap plasmid pSPPH21, we designed an insert DNA molecule encoding the *lac* operon regulatory region. This insert fragment contained the *lac* promoter, *lac* operator, and a CAP binding site and was 125bp in length. By ligating this well-characterized regulatory region into the pSPPH21 vector, our aim was to obtain fluorescent transformants and quantify the vector's insertion bias. To clone *lacp* into the pSPPH21 vector, we first performed NruI restriction digestion to linearize pSPPH21, generating a fragment with blunt ends. NruI functionality was confirmed by resolving digested fragments on an electrophoresis gel, producing a prominent band representing a linear product (Figure 1). This was followed by Shrimp Alkaline Phosphatase (SAP) treatment to dephosphorylate the 5' ends of the plasmid in order to prevent recircularization without insert. The ligation mixture was then transformed into *E. coli strains* TOP10 and DH5α using heat shock method. After ligation, we screened the resulting transformants using a restriction digestion screen with NruI. Since insertion into the vector should destroy the NruI site, plasmids with insert should not be linearized with NruI treatment. Clones that passed the restriction digest screen were further characterized by fluorescence quantification and Sanger sequencing to confirm the presence and direction of the insert.

Cells with higher transformation efficiency compensate for low-efficiency ligation. We observed that using competent *E. coli* with a higher transformation efficiency produced ligated transformants where less efficient *E. coli* strains were not able to. For transformation, digested (linear) pSPPH21 was the negative control, undigested (circular) pSPPH21 was the positive control and pSPPH21 from our *lacp* ligation reaction was used as our test condition. The linear negative control plasmid transformation yielded no colonies using both cell types. This was expected since a linearized plasmid cannot replicate and therefore cannot confer chloramphenicol resistance in the resulting daughter cells. The circular plasmid positive control transformation yielded many colonies for both cell types as expected, which also served to verify that our experimental conditions allowed for transformation. Notably, the higher efficiency TOP10 *E. coli* strain produced many more colonies in the positive control compared to *E. coli* DH5α. For the test ligation reaction mixture, colonies were not observed using lower efficiency *E. coli* strain DH5α, whereas 15 colonies were present using *E. coli* strain TOP10. These results suggest that the higher transformation efficiency of the TOP10 cells allowed us to obtain transformants.

Two transformed colonies were not linearized by NruI digestion, suggesting successful insertion. We examined whether the transformed colonies had inserted *lacp* by isolating the plasmid DNA and performing an NruI restriction digestion screen followed by an electrophoresis gel to resolve the fragments. The positive control contained NruI digested pSPPH21, resulting in one prominent band at ~3500 bp representing a linear product (Figure 1). The negative control contained undigested pSPPH21, resulting in three discernible bands corresponding to different plasmid conformations, presumably supercoiled, open circular, and linear (Figure 1). All clones, except 5 and 6, show a similar pattern to the pSPPH21 NruI digested positive control, but under the undigested condition appear to adopt the conformations seen in the negative control, suggesting they are susceptible to NruI linearization and that the NruI site is intact in the vector (Figure 1). NruI treatment of clone 6 produced a similar band pattern to the undigested condition and resembled the negative control, suggesting it is not susceptible to NruI digestion (Figure 1B). Clone 5 produced one equivalent band under both NruI conditions, smaller in size compared to both controls (Figure 1A). Therefore, since *lacp* insertion is expected to render the NruI site non-functional, these results indicate that clones 1-4 and 7-11 are empty vectors due to NruI susceptibility. Clones

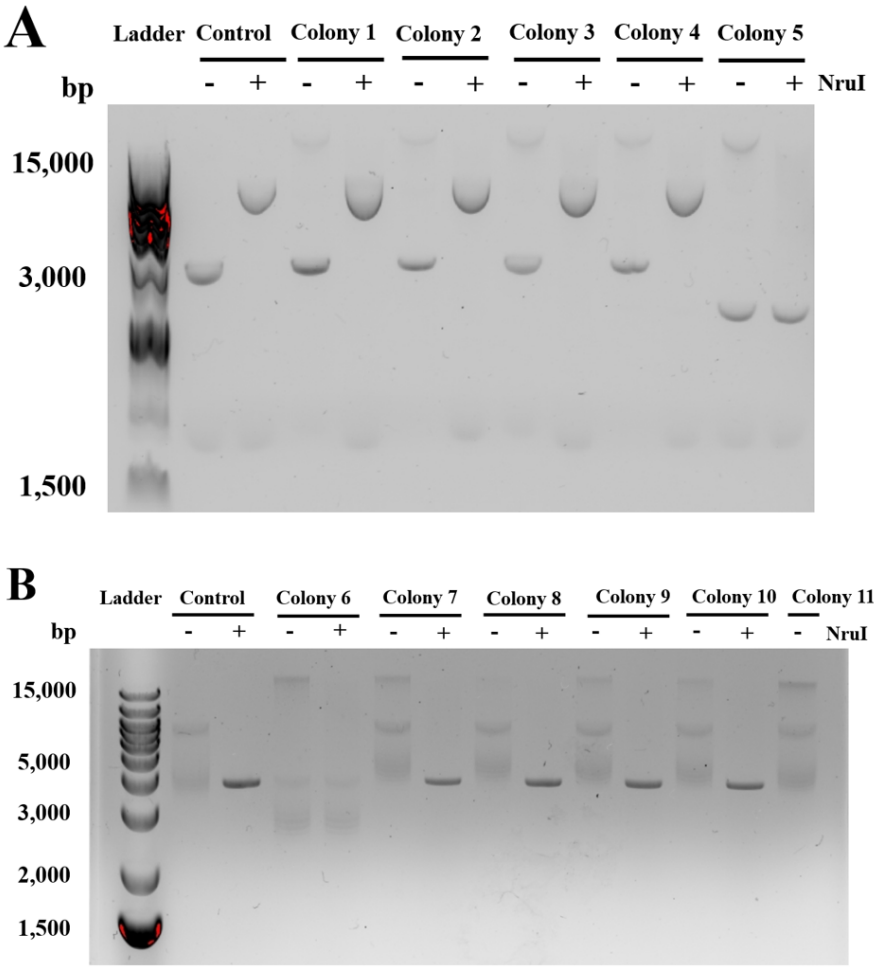

FIG. 1 Clones 5 and 6 were not linearized by NruI digestion as shown in gel electrophoresis, suggesting successful insertion. 1% Agarose Gel Electrophoresis results of clones 1-5 (A) and clones 6-11 (B) in the presence or absence of NruI indicate that clones 5 and 6 did not linearize due to NruI digestion. The DNA ladder in gel A did not migrate as expected, but the profiles of the bands can still be observed. Controls were undigested and digested pSPPH21 plasmid as shown in columns beside the DNA ladder. Invitrogen™ 1Kb Plus DNA Ladder was used while Electrophoresis ran at 100V for 60 minutes with DNA gels visualized on the BioRad ChemiDoc MP.

5 and 6 appear not to differ between digestion conditions, suggesting that the isolated plasmid DNA may have an insert thereby destroying the NruI site.

Clones 5 and 6 show no significant difference in GFP and RFP fluorescence compared to a negative control or when induced by IPTG. To further examine whether colonies 5 and 6 had an inducible *lacp* insert capable of driving GFP or RFP expression, we transformed plasmid DNA from clones 5 and 6 into *E. coli* strain DH5a cells and inoculated overnight cultures containing LB supplemented with chloramphenicol and IPTG for selection and induction, respectively. Samples were then transferred to a 96-well plate for an RFP and GFP fluorescence assay, measured in triplicate and averaged. OD_{600} readings were taken to ensure comparable cell densities among samples (Table 1). Empty pSPPH21 vector containing IPTG was used as a negative control because it is expected to have minimal fluorescence due to no inducible promoter to drive GFP or RFP expression. GFP measures of clone 5 produced 1739 ± 12 relative fluorescence units (RFU) under IPTG induction and 1721 ± 56 RFU without IPTG (Table 1). Therefore, GFP fluorescence does not differ significantly between IPTG conditions. Additionally, GFP fluorescence of clone 5 under both IPTG conditions is not significantly greater compared to the negative control, which produced 1682 ± 48 RFU (Table 1). GFP measures of clone 6 produced 1615 ± 37 RFU with IPTG and 1604 ± 71 RFU without IPTG (Table 1). These readings also do not differ significantly between IPTG conditions nor are they significantly greater than the negative control.

RFP measures for clone 5 produced 21 ± 4 RFU with IPTG and 23 ± 16 RFU without IPTG (Table 1). Clone 6 produced a reading of 15 ± 5 RFU with IPTG and 18 ± 10 RFU without IPTG (Table 1). Both clones 5 and 6 show no significant difference in RFP fluorescence between IPTG conditions or to the negative control, which produced a reading of 16 ± 11 RFU (Table 1). Therefore, IPTG induction does not increase GFP or RFP fluorescence readings in clones 5 or 6 and relative GFP and RFP fluorescence for the examined clones does not differ significantly from our negative control. This may suggest the

TABLE 1. Transformants do not display increased GFP/RFP fluorescence compared to negative control. GFP was measured at 485nm excitation, 525nm emission. RFP was measured at 580nm excitation, 620nm emission. Fluorescence and optical density readings were taken in technical triplicates, averaged, and expressed as a mean ± standard deviation (SD). Optical density readings were taken to ensure that an approximately equal number of bacterial cells were compared per treatment and per replicate and that no well exceeded an OD_{600} of 1.0. pSPPH21 plasmid grown in 1mM IPTG was included as a negative control.

Treatment	OD_{600} ± SD	GFP RFU ± SD	RFP RFU ± SD
pSPPH21 + IPTG	0.624 ± 0.012	1682 ± 48	16 ± 11
Clone 5 + IPTG	0.743 ± 0.002	1739 ± 12	21 ± 4
Clone 5 – IPTG	0.735 ± 0.013	1721 ± 56	23 ± 16
Clone 6 + IPTG	0.681 ± 0.010	1615 ± 37	15 ± 5
Clone 6 – IPTG	0.683 ± 0.007	1604 ± 71	18 ± 10

absence of an inducible promoter to drive fluorescence expression in pSPPH21 or other issues regarding GFP/RFP expression.

Sanger sequencing results of clones 5 and 6 show a truncation of pSPPH21. To visualize the plasmids of clone 5 and 6 and verify whether they have *lacp* insertion, we submitted the clones for Sanger sequencing. Samples were sequenced using RFP forward primer (5'- GGC GTA TCA CGA GGC AGA ATT TC - 3') and RFP reverse primer (5'- GGA AGC CTG CAT AAC GCG AAG -3'). Sequencing results were visualized using SnapGene viewer and consensus sequences between colonies 5 & 6 were determined using NCBI BLAST alignment tools. Clones 5 and 6 appear to have an intact RFP sequence but contain a truncated GFP sequence compared to the pSPPH21 intact construct (Figure 2). Figure 2 also shows that clones 5 and 6 lack an NruI digestion site as well as the *lacp* insert. Therefore, it appears that partial plasmid loss occurred during experimentation. The absence of an NruI site renders these truncated constructs resistant to NruI linearization, which explains the results of the restriction digest gel screen.

DISCUSSION

The substrate-induced gene expression (SIGEX) vector was designed to act as a promoter trap plasmid for the discovery of novel metagenomic substrate-induced promoters. Abrishamkar *et al.* improved upon the design of the original SIGEX vector by adding red fluorescent protein (RFP) in the opposite orientation of the original green fluorescent protein (GFP) gene (11). This modification enables the detection of promoters inserted in either direction. Since this improved vector had never been tested before, we decided to conduct a proof-of-concept experiment to validate the activity of the SIGEX vector by ligating a known substrate-inducible promoter into the vector. With the addition of the RFP gene to the SIGEX vector, we also wanted to assess the ratio of green fluorescent to red fluorescent colonies produced by this experiment to determine if the SIGEX vector has an insertion bias. To answer these questions, our approach was to ligate the *lac* operon regulatory region into the SIGEX vector, screen transformants for successful insertion using a restriction digest gel screen, observe the red and green fluorescence from transformants with insert, and submit plasmid DNA from these transformants for Sanger sequencing. Our goal was to validate the activity of the SIGEX vector and quantify its insertion bias using sequencing data and the ratio of red: green fluorescence observed in these transformants.

After multiple attempts to obtain clones containing plasmids encoding our *lacp* insert in the pSPPH21 plasmid we obtained two clones – referred to as clones 5 and 6 – resistant to NruI digestion. Since we expected our insert to destroy the NruI site in pSPPH21, we reasoned that clones resistant to NruI digestion should contain insert. We then used Sanger sequencing to characterize plasmid DNA from clones 5 and 6 (Figure 2). Analysis of our sequencing results showed that a portion of the GFP gene along with the NruI site had been deleted from these plasmids and that neither plasmid encoded the intended lacp insert. These results were consistent with fluorescence measurements of IPTG induced clones which showed no GFP or RFP signal (Table 1).

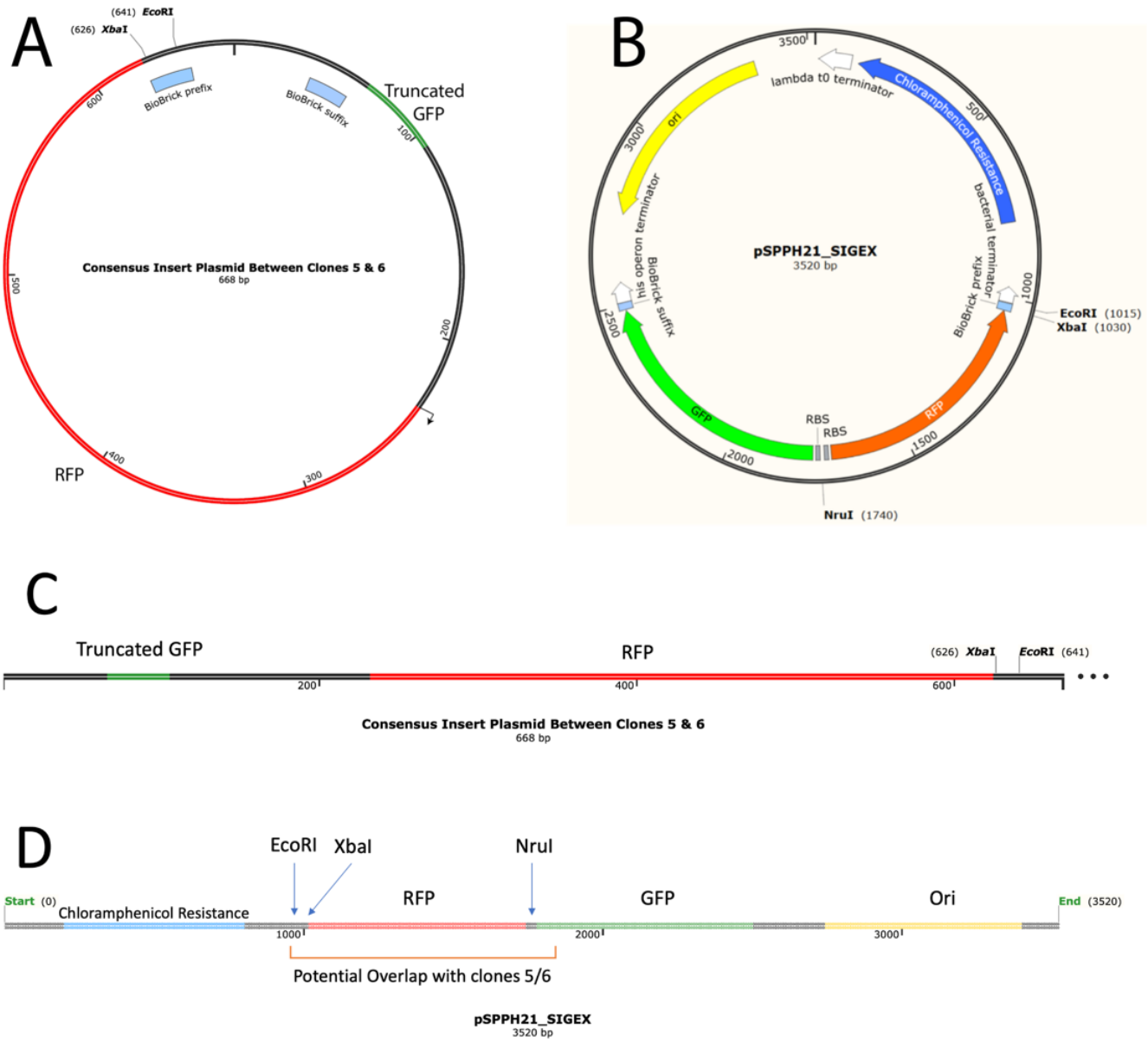

FIG. 2 Sequencing results show an identical consensus sequence containing a truncated GFP gene fragment suggestive of partial plasmid loss common across clones 5 & 6. Samples were submitted according to the specifications created by GENEWIZ as described previously. Consensus sequences between colonies 5 & 6 were determined using NCBI BLAST alignment on their corresponding sequence files. Sequencing results were then visualized using the SnapGene viewer and further processed in Adobe Illustrator with the consensus sequence shown as a circular plasmid on the left (A) and the original circular pSPPH21 SIGEX plasmid on the right (B) for comparison of their size differences indicative of plasmid loss. Linear forms of the consensus sequence (C) and pSPPH21 SIGEX plasmid (D) reveal a potential overlap indicative of the conserved region of the pSPPH21 plasmid that was carried forward in clones 5 and 6.

We were ultimately unable to validate the activity of the SIGEX vector and determine its insertion bias. Instead, we will focus our analysis on the limitations of the SIGEX vector for screening metagenomic DNA and how we believe these limitations can be overcome through optimizations of the SIGEX cloning protocol. Previous studies have noted that promoter trap plasmids like the SIGEX vector have exhibited lackluster performance in the characterization of novel substrate-induced promoters (14). Researchers have identified that limiting factors such as low-quality metagenomic DNA and incompatibility between host and metagenomic DNA contribute to SIGEX's poor performance (14). However, the low efficiency of blunt-end ligations has not yet been reviewed as a weakness of SIGEX technology. In this paper,

we will summarize our own attempts to optimize the efficiency of the SIGEX protocol and offer suggestions to future researchers on how to improve the effectiveness of SIGEX-based screening of metagenomic DNA fragments.

Increasing transformation efficiency can compensate for low efficiency blunt-end ligation. Our first goal was to clone the *lac* operon regulatory region into the SIGEX vector using blunt-end ligation. Blunt-end cloning is used in the SIGEX method for its ability to ligate without complementary sequences between insert and vector. However, the major limitation of blunt-end cloning is the ligation step, where the ligation efficiency is between 10-100 times less efficient than cohesive-end ligations (15). This is due to the lack of hydrogen bonding between complementary nucleotides present in cohesive-end ligation, thus decreasing the affinity for insert and vector and the likelihood of the blunt-end ligation reaction (15). We encountered this limitation in our own research, when our first attempt to ligate the *lac* operon regulatory region into the SIGEX vector failed after transforming into *E. coli* DH5α. We then hypothesized that the inefficient blunt-end ligation reaction could be compensated for by using a competent *E. coli* stock with a higher transformation efficiency.

To test this hypothesis, we performed another transformation using competent cells with a higher transformation efficiency and the same ligation product. This optimization succeeded in generating more ligated transformants (Figure 1D). For this optimization, we used bacteria with a transformation efficiency of 1 x 10^9 CFU/ug DNA. Other researchers employing the SIGEX method have achieved similar success by using the electroporation method of transformation (14). Electroporation is generally considered the highest efficiency transformation method – capable of yielding up to 1 x 10^{10} CFU/ug DNA (16). Since we have shown that using high-efficiency competent cells can compensate for a low-efficiency ligation reaction, electroporation is one technique that may be applied to the SIGEX protocol to yield more ligated transformants. Furthermore, it is important to maximize both the frequency of successful ligation events, as well as the transformation efficiency of competent cells, in order to maximize the likelihood of capturing metagenomic substrate-induced promoters.

Restriction digest gel screening and Sanger sequencing reveal problems with plasmid rearrangement. To verify that transformants possessed plasmids containing ligated insert rather than empty vector, we performed a restriction digest gel screen. Using this method, we found two transformants which we believed carried ligated plasmid (Figure 1). However, upon further analysis we observed that these clones did not fluoresce in the green or red fluorescence channel (Table 1). We also submitted plasmid DNA from the two transformants for Sanger sequencing and found that these transformants had acquired a truncated version of the SIGEX plasmid that was missing the NruI site and part of the GFP gene (Figure 2). This explains why these transformants were not linearized during the restriction digest gel screen. Plasmid rearrangement is a known phenomenon associated with transformation using linearized DNA (17). Competent cells can be transformed with linear DNA; however this does not confer a fitness advantage since only circular DNA can be replicated and passed on to daughter cells. Due to the low efficiency of blunt-end ligation, it is possible that some competent cells were transformed with linearized DNA which then re-circularized *in vivo* – leading to plasmid rearrangement. Circularization by homologous recombination may explain why transformants acquire truncated plasmids and why multiple clones acquire the same truncation (Figure 2). Rearrangement would confer a fitness advantage to the cell since the transformant would acquire Chloramphenicol resistance without the anabolic burden of GFP/RFP production. Both plasmid rearrangements we observed in this study were near the NruI site, further strengthening this hypothesis. Since plasmid rearrangement is independent of ligation, transformants could acquire rearranged plasmids even when an insert is not ligated into the vector. Using colony PCR instead of restriction digest analysis or omitting phosphatase treatment in future proof-of-concept experiments may help eliminate transformants with rearranged plasmids from further characterization. For metagenomic screens, future researchers should try to optimize the blunt-end ligation protocol to increase the frequency of plasmid recircularization with insert and therefore reduce the number of linearized fragments.

Optimizing the SIGEX protocol to screen metagenomic DNA for substrate-inducible promoters. To effectively screen a metagenomic library of DNA fragments where novel substrate-inducible promoters make up a miniscule fraction of all fragments, it is essential that blunt-end ligation be optimized as much as possible. We attempted numerous modifications to the ligation protocol with the goal of increasing the efficiency of ligation (Table 2). Although these modifications were not successful in this study, previous studies have highlighted their effectiveness. Therefore, we believe they are relevant for future researchers looking to ligate metagenomic DNA fragments into the SIGEX vector. One such modification is temperature-cycle ligation, which increases the efficiency of blunt-end ligation by optimizing the activity of T4 DNA Ligase and the annealing of DNA (18). Since T4 DNA Ligase is most active at 37°C and DNA anneals best at low temperature, ligation reactions must achieve a delicate balance between ligase activity and DNA annealing for the reaction to proceed (Table 2). Temperature-cycle ligation purports to maximize this balance by cycling between optimal enzyme temperature and optimal DNA annealing temperature in 30-second windows (18). With this method, the authors were able to obtain up to six times more ligated colonies compared to incubation at room temperature (18). This makes temperature-cycle ligation a valuable tool for increasing the efficiency of SIGEX blunt-end ligations.

Another factor greatly affecting the efficiency of ligations is the molar ratio of insert DNA to vector DNA. Optimal insert: vector ratios can vary substantially depending on the length of the DNA fragments being ligated – therefore we recommend testing a range of molar ratios when optimizing ligation. Past research has found that the optimal insert: vector molar ratio is dependent on the properties of the DNA being ligated – such as length, restriction sites, and phosphatase treatment (19). For short inserts – such as metagenomic DNA fragments or the *lac* operon regulatory region used in this study – Revie et al. recommend using a 5nM insert concentration, or 100fmol in a 20uL ligation reaction (19). We tested this molar ratio as part of our optimization experiments, and we believe that future SIGEX research involving metagenomic DNA fragments should also take the insert: vector molar ratio into account when optimizing ligation (Table 2).

Finally, specific products have been developed to optimize blunt-end ligation reactions and may be tested for use in the SIGEX protocol. These kits usually take advantage of the ability of polyethylene glycol (PEG) to induce macromolecular crowding in a small reaction

TABLE 2. Troubleshooting ligation through modification of protocol. List of various modifications to optimize the ligation protocol along with corresponding rationale and results.

Modification	Rationale
Increasing transformation efficiency by using OneShot™ TOP10 Chemically Competent *E. coli.*	- Transformations using *E. coli* DH5α resulted in few colonies on the transformation positive control plate. - If ligation is inefficient, increasing transformation efficiency will increase the chance that a competent cell takes up ligated plasmid. - TOP10 *E. coli* have a transformation efficiency that is a thousand times higher than *E. coli* DH5α. - Transformants won't require induction since TOP10 *E. coli* do not carry the *lac* repressor.
Temperature cycling the ligation mixture overnight	- Ligation is a delicate balance between enzyme activity and random DNA interactions. - The optimal temperature for T4 DNA Ligase is 37°C. - The optimal temperature for DNA annealing is 10°C. - Cycling between these two temperatures overnight will maximize DNA interactions and enzyme activity, leading to a higher ligation efficiency.
Adding more enzyme to the ligation mix	- Blunt-end ligations are inefficient due to a lack of stabilizing interactions between DNA molecules. - To compensate for this, higher amounts of enzyme are used to increase the chances of collision between vector, insert, and enzyme.
Varying the insert: vector molar ratio from 1:1, 3:1, 5:1, 10:1	- Increasing the number of insert molecules in the ligation reaction will increase the chance of collision between vector and insert. - The insert is 125bp and small inserts tend to ligate better at higher concentrations.
Using high concentration enzyme	- High concentration T4 DNA Ligase is advertised to achieve higher efficiency blunt-end ligation in just five minutes.

volume, thus further increasing the collisions between insert and vector DNA molecules. Commercial kits also tend to use a higher concentration of T4 DNA Ligase to increase blunt-end ligation efficiency by increasing the number of enzyme molecules available to catalyze ligation (20). The Blunt/TA Ligase Master Mix from New England BioLabs® Inc. is one of the few kits on the market specifically designed to increase the efficiency of blunt-end ligation reactions and purports to achieve up to 70% product yield (21). Since blunt-end ligation is orders of magnitude less efficient than cohesive-end ligation, it is important that protocols and products designed specifically for ligating blunt DNA ends are used.

Limitations It would be good to verify that our gBlocks™ gene fragment (insert) and vector have the desired structure and are functional. Verifying the structure can be accomplished by sequencing both our insert and initial SIGEX vector. As for examining functionality, this is outlined in future directions.

When we managed to obtain ligated colonies, the efficiency of transformation appeared quite poor as only two colonies were believed to have inserts. Such poor efficiency could be improved by instead carrying out sticky-end cloning instead of blunt-end cloning – which is less efficient. This would entail remodeling the SIGEX vector to include sticky ends. To accomplish this, one could perform a t-tailing reaction on the vector and an a-tailing reaction on the insert or metagenomic DNA library. This would allow metagenomic DNA fragments to be cloned into the vector using cohesive-end ligation.

It is possible that overexpression of fluorescent protein results in cell death and little to no transformed cells on plates due to toxicity. Our protocol does not tightly regulate expression of fluorescent protein and thereby allows expression of the protein as the cells grow and divide. Perhaps tighter control of expression by having both the lac repressor and promoter encoded into the insert instead of relying on the repressor that is part of the host genome could result in fluorescence and provide control over potential toxicity.

Conclusions The process of validating the SIGEX vector for use as a promoter trap has been focused on optimizing ligation and transformation efficiency in proof-of-concept experiments. The difficulties outlined in this study involving inefficient blunt-end ligation and plasmid rearrangement represent opportunities for future researchers to optimize the SIGEX protocol for screening metagenomic DNA. Given the low frequency of substrate-inducible promoters in metagenomic samples combined with the difficulties involved with obtaining and purifying a sample of metagenomic DNA, ligation and transformation must first be optimized in order for the SIGEX protocol to be effective as originally intended. We hope that lessons learned from our research and our suggestions for optimizing the SIGEX protocol will allow future researchers to discover novel substrate-inducible promoters using SIGEX technology.

Future Directions It would be a good idea if future researchers examined the *lac* operon regulatory region insert if it is to be used in future research. Confirming that the issue may be with our gBlocks™ gene fragment (insert), and not the pSPPH21 vector, can be done by designing an alternative insert that eventually results in fluorescent bearing colonies. Using this alternative insert and obtaining fluorescence will confirm that our vector is functional. Next, to confirm that our insert is functional, one could add overhangs to the insert and ligate it into a vector such as TOPO and then submit the plasmid for sequencing to see if insertion occurred. This offers a different means of ligation to verify insertion of our vector and if successful, pinpoint the issue having to do with our insert and ligation. At this point, it could be examined if this ligation problem is site specific by ensuring that we are able to successfully blunt-end ligate an alternate DNA fragment into the SIGEX vector by carrying out blunt-end ligation at an alternate site. Following verification of the insert and vector as well as understanding the specificity of the ligation problem, it is possible to now explore different ways of improving ligation using the SIGEX vector and our insert. In Table 2, strategies we employed to improve ligation were largely unsuccessful but there are others which we were unable to explore. Several other strategies include having expression of both the *lac* repressor and promoter on the insert, using a PCR amplified DNA insert instead of a gBlocks™ gene fragment, or using a commercial kit specifically tailored for blunt-end

ligation reactions. Finally, we hypothesized that ligated transformants would express GFP/RFP constitutively in *E. coli* TOP10 since it harbors a complete *lac* operon deletion. However, we were unable to confirm this since no transformants contained insert. Future researchers may wish to transform *E. coli* TOP10 with a ligated pSPPH21 plasmid and observe if transformants fluoresce without IPTG induction. If successful, this could be an effective SIGEX positive control and a useful way to generate fluorescent clones for FACS gating.

ACKNOWLEDGEMENTS

We would like to thank Dr. David Oliver for his creative and experimental guidance and Jade Muileboom for her technical support throughout the term. We also want to acknowledge our colleagues Gawol et al., Agnew et al., and Frese et al. for sharing their experiences, scientific knowledge, and materials. Finally, thank-you to the Department of Microbiology and Immunology at the University of British Columbia for providing the funding and lab space required to carry out this research.

CONTRIBUTIONS

KM wrote the methods, discussion, and conclusion. JM wrote the limitations and future directions, NH wrote the abstract, introduction, and figure captions. AM wrote the results section.

REFERENCES

1. **Whitman WB, Coleman DC, Wiebe WJ.** 1998. Prokaryotes; the unseen majority. PNAS **95**:6578–6583.
2. **DeLong EF, Pace NR, Kane M.** 2001. Environmental Diversity of Bacteria and Archaea. Sys Bio **50**:470–478.
3. **Sasidharan S, Divakar K, Saudagar P**. 2020. Culture-Independent Omics-Techniques for Microbiome-Based Molecular Therapeutics Against Infectious Diseases, p. 95–114. *In* Metagenomic Systems Biology. Springer Singapore, Singapore.
4. **Uchiyama, T., and Miyazaki, K.** 2010. Substrate-Induced Gene Expression Screening: A Method for High-Throughput Screening of Metagenome Libraries, p. 153–168. *In* Metagenomics, (Totowa, NJ: Humana Press).
5. **Baric RS, Crosson S, Damania B, Miller SI, Rubin EJ.** 2016. Next-Generation High-Throughput Functional Annotation of Microbial Genomes. mBio **7.**
6. **Wakamatsu T, Mizobuchi S, Mori F, Futagami T, Terada T, Morono Y.** 2021. Construction of Aerobic/Anaerobic-Substrate-Induced Gene Expression Procedure for Exploration of Metagenomes from Subseafloor Sediments. Front Microbiol **12**:726024–726024.
7. **Ngara TR, Zhang H.** 2018. Recent Advances in Function-based Metagenomic Screening. Genomics Proteomics Bioinformatics **16**:405–415.
8. **Watanabe K, Uchiyama T, Abe T, Ikemura T.** 2005. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. Nat Biotechnol **23**:88–93.
9. **Yun J, Ryu S.** 2005. Screening for novel enzymes from metagenome and SIGEX, as a way to improve it. Microb Cell Fact **4**:8–8.
10. **Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G.** 2013. Enhancers: five essential questions. Nat Rev Genet **14**:288–295.
11. **Parsa Tabassi, Sourena Oveisi, Helia Mansouri Dana, Parsa Abrishamkar.** 2022. Construction of SIGEX duo-directional reporter plasmid pSPPH21. UJEMI.
12. **Sambrook J, Russell DW**. 2001. Molecular cloning: a laboratory manual 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
13. **Uchiyama T, Watanabe K.** 2008. Substrate-induced gene expression (SIGEX) screening of metagenome libraries. Nat Protoc **3**:1202–1212.
14. **Meier MJ, Paterson ES, Lambert IB.** 2016. Use of Substrate-Induced Gene Expression in Metagenomic Analysis of an Aromatic Hydrocarbon-Contaminated Soil. Appl Environ Microbiol **82**:897–909.
15. **BiteSize Bio.** 2019. A Beginner's Guide to How Blunt-End Cloning Works. https://bitesizebio.com/45131/a-beginners-guide-to-how-blunt-end-cloning-works/. Retrieved 17 April 2022.
16. **Dower WJ, Miller JF, Ragsdale CW.** 1988. High efficiency transformation of E. coli by high voltage electroporation. Nucleic Acids Res **16**:6127–6145.
17. **Conley EC, Saunders VA, Saunders JR.** 1986. Deletion and rearrangement of plasmid DNA during transformation of Escherichia coli with linear plasmid molecules. Nucleic Acids Res **14**:8905–8917.
18. **Lund AH, Duch M, Skou Pedersen F.** 1996. Increased Cloning Efficiency by Temperature-Cycle Ligation. Nucleic Acids Res **24**:800–801.

19. **Revie D, Smith DW, Yee TW.** 1988. Kinetic analysis for optimization of DNA ligation reactions. Nucleic Acids Res **16**:10301–10321.
20. **Packer H.** 2022. Tips for blunt-end DNA cloning and ligation. Integrated DNA Technologies. https://www.idtdna.com/pages/education/decoded/article/cloning-strategies-part-3-blunt-end-cloning. Retrieved 18 April 2022.
21. **New England BioLabs.** Blunt/TA Ligase Master Mix. https://international.neb.com/products/m0367-blunt-ta-ligase-master-mix#Product%20Information. Retrieved 18 April 2022.