Volume 8:1-12

UJEMI+

The Undergraduate Journal of Experimental Microbiology & Immunology (+Peer Reviewed)

Duo-directionality of the substrate-induced expression screening vector pSPPH21 confirmed with a lac operon screen

Diana Gawol*, Rachel Floyd*, Kyla Kohara, Yoyo Lee

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

* These authors contributed equally to this work

SUMMARY Within the microbial metabolic diversity lies the potential to utilize catabolic genes for medical, industrial and biotechnological applications. Substrate-induced gene-expression (SIGEX) screening vectors are promoter traps used to screen environmental metagenomic DNA libraries for novel catabolic genes. When a substrate-sensitive promoter is ligated into the plasmid and induced by a substrate, the host cell of the library will express a downstream promoter-less fluorescent reporter. Fluorescence activated cell sorting (FACS) can then be used to isolate positive clones for further study of the ligated promoter and to determine its associated catabolic genes. Despite their importance, traditional SIGEX vectors are limited to detecting promoters ligated in the same 5' to 3' direction as the reporter gene. This led to the development of pSPPH21, a duo-directional SIGEX plasmid, by Abrishamkar *et al*. It was designed to contain two oppositely oriented fluorescent reporters, green fluorescent protein (GFP) and red fluorescent protein (RFP) to allow for the detection of promoters ligated in either direction. We tested the proposed duo-directional functionality of pSPPH21 by using the well-studied inducible promoter of the *lac* operon. Fluorescence imaging and quantification on a plate reader confirmed that pSPPH21 expresses both GFP and RFP upon induction with the allolactose analog isopropyl-β-D-thiogalactopyranoside (IPTG). We further confirmed that GFP and RFP expression occurs at a 1:1 ratio as expected from the directionally unbiased blunt-end ligation. With the proof-of-concept established, further optimization steps can now be undertaken in preparation for an experimental DNA library screen.

INTRODUCTION

The microbial world harbours most of the metabolic diversity on Earth. Microbes are continuously responding to evolutionary pressures of their ever-changing environments by evolving metabolic pathways. Since only a small fraction of microorganisms can be cultured, metagenomic methods are at the forefront of novel gene discovery (1). To offer an alternative approach to phenotype- and genotype-based metagenome library screens, Uchiyama *et al.* developed a method for substrate-induced gene-expression (SIGEX) screening that takes advantage of the fact that catabolic regulatory elements are often directly activated by a substrate or intermediate metabolite (1, 2).

SIGEX screening involves a plasmid-based promoter trap that allows for high-throughput screening of metagenomic libraries by use of a fluorescent reporter (2). In brief, DNA is extracted from environmental samples of interest and a metagenomic library is created by ligating DNA fragments into a SIGEX vector. This library is then transformed into a host and induced with a substrate of interest. If a ligated library fragment contains a promoter sensitive to the substrate, the down-stream fluorescent reporter will be expressed. Grown in liquid

Published Online: September 2022

Citation: Diana Gawol, Rachel Floyd, Kyla Kohara, Yoyo Lee. 2022. Duo-directionality of the substrate-induced expression screening vector pSPPH21 confirmed with a lac operon screen. UJEMI+ 8:1-12

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

Address correspondence to: Rachel Floyd, rachelcfloyd@gmail.com

culture, positive clones can then be selected by fluorescence-activated cell sorting (FACS). Relying on the common proximity of promoters to their respective genes, sequencing can then be used to identify the substrate-induced catabolic gene (2, 3). The advantage of SIGEX screening over genotype-based methods is that inducible catabolic genes can be detected without relying on prior sequence information. In contrast to phenotype-based screens, SIGEX vectors are designed to trap regulatory elements that interact with an inducer, eliminating the need for functional enzyme expression in the screening host cells. The utility of SIGEX vectors was demonstrated by the identification of previously unknown metal-ion induced genes from subseafloor sediment samples in Japan (17).

Commonly SIGEX plasmids are unidirectional, containing only one reporter gene such as green fluorescent protein (GFP) (2). Hence, they do not allow for the identification of promoters ligated in reverse. To solve this problem, Abrishamkar *et al.* designed a duo-directional reporter plasmid pSPPH21 by adding a second, oppositely directed fluorescent reporter, red fluorescent protein (RFP), to the pSB1C3 vector which originally contained only a GFP reporter (4). A blunt-end NruI cut site for library ligation was added equidistant between GFP and RFP (4). Sanger sequencing confirmed the successful integration of RFP and the NruI cut site, but also revealed mutations in the ribosomal binding site (RBS) upstream of the GFP reporter (4). These mutations could affect GFP expression by reducing ribosomal binding. While Abrishamkar *et al.* successfully constructed pSPPH21, the proof-of-concept of the duo-directional function has yet to be demonstrated (4).

Here, we confirmed the unbiased duo-directional reporter expression of pSPPH21 by using the inducible promoter of the lactose operon (*lac* operon). The *lac* operon is a well-understood inducible gene system and regulates the transport and metabolism of lactose in *Escherichia coli* (*E. coli*) (5, 6). In the absence of lactose, the repressor protein LacI is bound to the *lac* operator and prevents the binding of RNA polymerase to the *lac* promoter. Upon transport into the cell, lactose is converted to allolactose by β-galactosidase. Allolactose then allosterically binds the repressor protein and causes its release from the *lac* operator, which then allows for transcription initiation by RNA polymerase at the *lac* promoter. Alternatively, the non-hydrolysable allolactose analog isopropyl-β-D-thiogalactopyranoside (IPTG) functions as an inducer of the *lac* operon. Due to the high degree of understanding of the *lac* operon, it lends itself to this type of experimentation.

The *lac*-insert used in our proof-of-concept test comprised the *lac* promoter, *lac* operator, as well as the *lacI* gene, and the *lacI* promoter. The *lacI* gene encodes for the *lac* repressor protein, LacI. The rationale for including *lacI* in the screening insert is that pSPPH21 is a high-copy plasmid. We suspected that the native LacI expression of the DH5α host cells would not suffice to regulate the increased number of *lac* promoters presented by *lac*-insert containing pSPPH21. By including *lacI* in the insert design, we ensured a 1:1 molar ratio of LacI:*lac* promoter to achieve tight expression control.

We were able to demonstrate expression of both GFP and RFP by fluorescently imaging colonies, as well as fluorescence measurements on a 96-well plate reader. We were furthermore able to confirm that, as expected for directionally unbiased ligation, reporter expression occurs without strong bias for either GFP or RFP (n=23). Lastly, we identified ligation and transformation procedures as areas that require further optimization before conducting a DNA library screen with pSPPH21.

METHODS AND MATERIALS

Bacterial Strains and Plasmids. pSPPH21 stored in *E. coli* DH5α was obtained on a starter plate from the UBC MICB 401 laboratory stocks. *E. coli* MG1655 cells were also obtained from UBC MICB 401 laboratory stocks. Commercial chemically competent DH5α cells (Invitrogen) were used for transformation experiments.

Primers. Custom primers for amplification of the *lac*-insert from MG1655 gDNA were designed using NCBI primer BLAST.
Lac-Insert FWD: 5'- AGA AGG GGT TGA ATC GCA GG -3'
Lac-Insert REV: 5'- GAC GAC AGT ATC GGC CTC AG -3'

Primers from Abrishamkar *et al*. were used for colony PCR (4).

pSPPH21_RFP FWD: 5'- GGC GTA TCA CGA GGC AGA ATT TC -3'
pSPPH21_RFP REV: 5'- GGA AGC CTG CAT AAC GCG AAG -3'

Primers from Agnew *et al.* were used for Sanger sequencing (18).
pT7ForP FWD: 5'- ATT TCG AAC TCG TGA CCG TT -3'
pT7RevP REV: 5'- ACT GAC AGA AAA TTT GTG CC -3'

Plasmid minipreps. Minipreps were prepared according to manufacturer instructions from 1-3 mL of overnight culture using the PureLink Quick Plasmid Miniprep Kit (Invitrogen). DNA concentration and quality were determined by NanoDrop (Thermo Scientific).

Luria-Bertani (LB) broth and agar. LB broth was prepared with 1% tryptone, 1% NaCl, and 0.5% yeast extract with distilled water. For LB agar, 1.5% agar powder was added prior to autoclaving. If required, chloramphenicol (to a final concentration of 25 µg/mL) and IPTG (to a final concentration of 0.5 mM) were added once the LB cooled to ~50°C.

Restriction enzyme digest. Suitable restriction enzymes for structure verification of pSPPH21 were identified using Benchling (7). 1 µg of pSPPH21 was digested with AhdI (NEB), BspHI (NEB), or both according to protocols provided by NEBcloner on the NEB website. The reaction was incubated at 37°C for 15 minutes. 10 µL of each reaction were run on a 1% agarose gel for analysis.

Agarose DNA Gel Electrophoresis. All gel electrophoresis experiments were performed using 1% agarose gels prepared in 1X TAE. RedSafe DNA stain (Bulldog Bio) was added (1:20,000) before casting the gel. Samples were mixed with 6X gel loading dye Purple (NEB). 3 µL of O'GeneRuler DNA ladder Mix (Thermo Scientific) were loaded as reference. Gels were run at 105 V for 65 minutes and visualized using the Bio-Rad ChemiDoc MP imaging system on the "ethidium bromide" setting.

Sanger Sequencing. Purified plasmid samples were prepared with primers from Agnew *et al*. and Abrishamkar *et al.* for Sanger sequencing according to GeneWiz instructions (protocol for "US pre-defined") (4, 18). Alignments with the putative pSPPH21 sequence (provided by Abrishamkar *et al.*), as well as putative *lac*-insert-containing pSPPH21 were performed using SnapGene® (4, 8). Low quality ends of aligned sequence traces were trimmed with high stringency and remaining gaps in the traces were checked manually for legitimacy.

Genomic DNA Extraction. Overnight culture of *E. coli* MG1655 was prepared by inoculating 3 mL of LB broth with a single colony and incubation at 37°C. Approximately 2 x 10^9 cells (~1 mL of overnight culture) were pelleted by centrifugation at 4,000 x g for 10 minutes. Genomic DNA was then extracted using the PureLink™ Genomic DNA Mini Kit (Invitrogen).

PCR Amplification of the *lac*-insert. 50 µL PCR reactions contained 1X SuperFi Buffer (Invitrogen), 1 mM of dNTP mix (Invitrogen), 0.5 µM of Lac-Insert FWD and Lac-Insert REV primers, 130 ng of template MG1655 gDNA, 0.02 U/µL Platinum™ SuperFi™ DNA Polymerase (Invitrogen), and PCR-grade water to 50 µL. The thermocycler was set for a 30 sec initial denaturation at 98°C, followed by 35 cycles of amplification with denaturation for 10 seconds at 98°C, annealing for 10 seconds at 55°C, and extension for 90 seconds at 72°C, and a final extension for 5 minutes at 72°C. 10 µL of the PCR product were run on a 1% agarose gel.

PCR clean-up. PCR products and linearized pSPPH21 were cleaned up using the GeneJET PCR Purification kit (Thermo Scientific). DNA concentration and quality were determined by NanoDrop.

5'-Phosphorylation of *lac*-insert. Purified *lac*-insert was 5'-phosphorylated with T4 polynucleotide kinase (T4 PNK) (Thermo Scientific). Five replicate tubes of the following

reaction were prepared on ice: 300 pmol of *lac*-insert, 10 µL 5X T4 DNA ligase buffer (NEB), 1 µL T4 PNK, and nuclease-free water up to 50 µL were combined in a PCR tube on ice. Reactions were incubated in a thermocycler for 30 minutes at 37°C and heat inactivated for 20 minutes at 65°C. Reactions were pooled, cleaned up and concentrated with the GeneJET PCR Purification kit. Final concentration was verified by Nanodrop. 5'-phosphorylated *lac*-insert (5'-P-*lac*-insert) was stored at -20°C.

Linearization and dephosphorylation of pSPPH21. pSPPH21 was linearized with NruI (NEB) and dephosphorylated with Shrimp Alkaline Phosphatase (rSAP) (NEB) to prepare for subsequent ligation. 1 pmol of pSPPH21, 1 µL NruI, 2 µL 10X NEBuffer r3.1 (NEB), 1 unit rSAP and nuclease-free water to 20 µL were combined in PCR tubes on ice, incubated for 30 minutes at 37°C and heat inactivated for 5 minutes at 65°C. Reactions were pooled, cleaned up and concentrated with the GeneJET PCR Purification kit. Final concentration of linearized, dephosphorylated pSPPH21 was verified by Nanodrop.

Ligation. For all ligations, a 3:1 molar ratio of insert:vector was used. Two different ligation protocols, called "simultaneous" and "sequential", were performed. In the simultaneous method adapted from Green and Sambrook, and Costa *et al.*, 50 ng pSPPH21, 70 ng *lac*-insert, 1 µL of T4 ligase (NEB), 2 µL 10 mM riboATP (NEB), 2 µL 10X NEBuffer3.1, 4 units NruI and nuclease-free water to 20 µL were combined and incubated at 37°C for 2 hours (9, 10). No heat inactivation was performed before proceeding to heat-shock transformation.

In the sequential method (adapted from NEB protocols), 4 µL of 5X T4 DNA ligase buffer, 50 ng of linearized and dephosphorylated pSPPH21, 70 ng of 5'-P-*lac*-insert, 1 µL T4 DNA ligase and nuclease-free water to 20 µL were combined and incubated at room temperature for 3 hours. No heat inactivation was performed before proceeding to heat-shock transformation.

Heat-Shock Transformation of pSPPH21 into DH5α *E. coli.* Heat-shock transformation of commercially competent DH5α cells was performed as described by Chang *et al.* (11). In brief, 5 µL of ligation reaction was added to 50 µL of competent DH5α cells which were then incubated on ice for 30 minutes. Water was used as a negative control. The cells were then heat-shocked in a 42°C water bath for exactly 30 seconds and chilled on ice for 2 minutes. 1 mL of pre-warmed S.O.C. medium (Invitrogen) was then added and the tubes were shake-incubated at 200 rpm at 37°C for 1 hour of outgrowth. The cells were then pelleted at 4000 x g for 10 minutes, and all but 200 µL of supernatant was removed. The cells were gently resuspended and 100 µL was plated on LB + 25 µg/mL chloramphenicol plates and incubated overnight at 37°C.

Colony PCR. First, an LB + chloramphenicol patch plate with transformants was used to maintain a reference of screened colonies. This patch plate was incubated at 37°C overnight and afterwards stored at 4°C. Using a sterile pipette tip, a small amount from each colony was patched and then transferred directly to PCR tubes containing Platinum™ Taq DNA Polymerase mastermix. For each reaction, the mastermix contained 2.5 µL 10X PCR buffer (Invitrogen), 0.75 µL 50 mM $MgCl_2$ (Invitrogen), 0.5 µL 10 mM dNTP mix (Invitrogen), 0.5 µL pSPPH21_RFP FWD primer, 0.5 µL pSPPH21_RFP REV primer, 0.1 µL Platinum™ Taq DNA Polymerase (Invitrogen) and PCR-grade water to 25 µL. The thermocycler protocol used was an initial cell lysis at 94°C for 10 min followed by 35 cycles of 30 sec at 94°C, 30 sec at 55°C and 90 sec at 72°C, and a final extension at 72°C for 5 min. The PCR products were run on a 1% agarose gel and visualized. Colonies containing inserts were then transferred to another patch plate for storage.

Fluorescent Agar Plate Imaging. Transformants were patched onto an LB + 25 µg/mL chloramphenicol plate (uninduced), as well as a LB + 25 µg/mL chloramphenicol plate with 0.5 mM IPTG (induced), and incubated overnight at 37°C. The next day, plates were visualized using the Bio-Rad ChemiDoc MP imaging system. Alexa 546 and Alexa 488 settings excite GFP and RFP fluorescence, respectively. Statistical significance was calculated with a X^2 test.

Induction in broth. 3 mL of LB + chloramphenicol broth was inoculated with a single colony and incubated for ~6 hours. Then, 100 µL of culture was transferred into a fresh 3 mL of LB + chloramphenicol broth and induced with 0.5 mM IPTG. Cells were then grown overnight at 37°C.

Fluorescence quantification with plate reader and data processing. To prepare for data collection with the fluorescent plate reader (BioTek Epoch 2 microplate spectrophotometer), 1 mL of overnight culture was pelleted at 4,000 x g for 5 min. The supernatant was discarded and the cells were washed twice with 1 mL of 1X PBS to remove background fluorescence from the yeast extract in the LB broth. A black-walled, clear-bottom 96-well microplate was loaded by adding 40 µL of PBS-suspended cells to 160 µL of PBS in each well. Each sample was prepared in quadruplicate. Plate reader settings were adapted from Kuiper (12). First, OD600 reads were obtained. For GFP reads, excitation and emission wavelengths were 485 and 528 nm, respectively. For RFP reads, excitation and emission wavelengths were 580 and 620 nm, respectively. To account for cell density, relative fluorescent units were divided by OD600 values of each well. Then, the average RFU of the empty vector measurements were subtracted as background from all wells. Data was plotted with Prism®. Fold-change between induced and uninduced samples was calculated by averaging the quadruplicates and dividing values for induced by values for uninduced samples. Statistical significance of fold-changes was calculated with a X^2-test.

RESULTS

Restriction enzyme digest confirmed the overall integrity of pSPPH21. To verify the suggested construct of pSPPH21, a restriction enzyme (RE) digest was performed. Benchling was used to identify cut sites within the RFP gene and in the plasmid backbone. The selected enzymes AhdI and BspHI cut once within the RFP gene and twice in the plasmid backbone, respectively (Fig. 1A). Virtual double-digest on Benchling predicted the expected band sizes at 608, 1028 and 1884 base pairs (bp).

pSPPH21 was digested with only AhdI, only BspHI, and both enzymes simultaneously and analyzed on a 1% agarose gel. Expected band sizes at ~600, 1000 and 1880 bp were visible (Fig. 1B). In the double-digest, the faint band at the size of the top fragment from the single digest with BspHI is indicative of an incomplete digestion by AhdI (Fig. 1B).

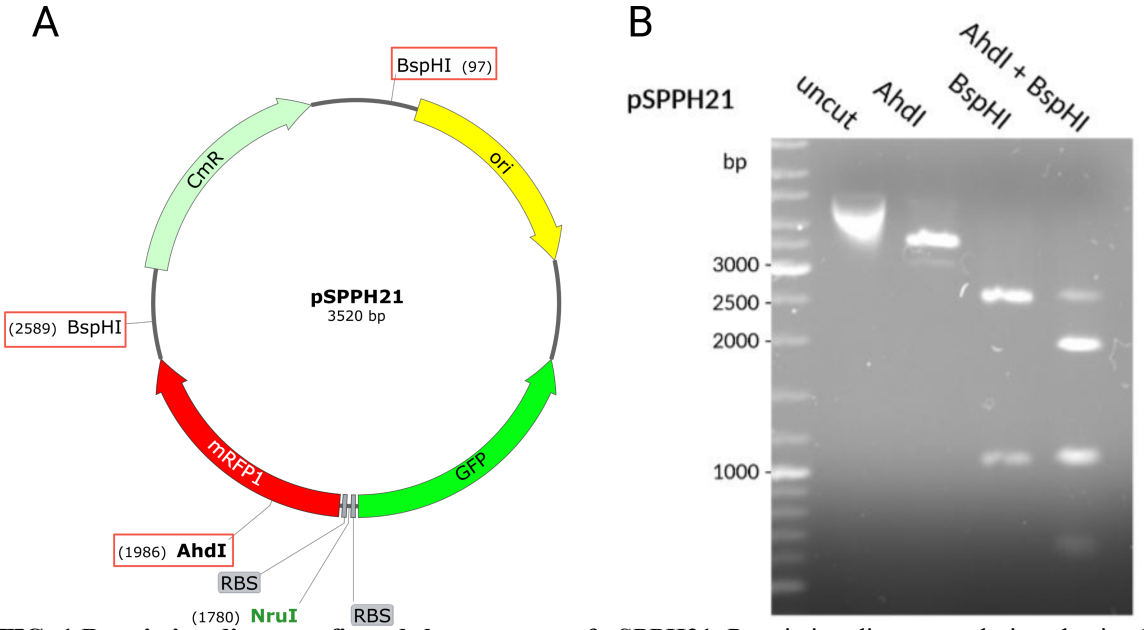

FIG. 1 Restriction digest confirmed the construct of pSPPH21. Restriction digest was designed using Benchling. **(A)** Restriction cut sites of AhdI and BspHI on pSPPH21. **(B)** 1% gel electrophoresis of restriction digests. Lane 1: DNA ladder. Lane 2: undigested pSPPH21. Lane 3: single digestion by AhdI. Lane 4: single digestion by BspHI. Lane 5: double digestion by AhdI and BspHI.

Sanger sequencing revealed changes at the GFP-associated ribosomal binding site. To investigate the mutations at the GFP-associated RBS, pSPPH21 was sent for Sanger sequencing using primers provided by Abrishamkar *et al.* and Agnew *et al.* (4, 18). The primers provided by Abrishamkar *et al.* bind further from the NruI site, downstream of the GFP and RFP genes in their respective directions, while the primers from Agnew *et al.* bind closer to the NruI site and within the fluorescent reporter genes (Fig. S1) (4, 18). Alignments with SnapGene® revealed that all sequences contained the same change in sequence; instead of the predicted GFP-associated RBS (part:Bba_J34801), a 25 bp insertion is present (Fig. 2A, B). Comparison of this 25 bp sequence to the Biobrick RBS library in the Standard Registry of Biological Parts (http://parts.igem.org/Ribosome_Binding_Sites/Catalog) revealed the presence of a medium-strength RBS (Bba_K082001) (Fig. 2C). Hence, GFP and RFP are associated with a medium-strength and strong RBS, respectively.

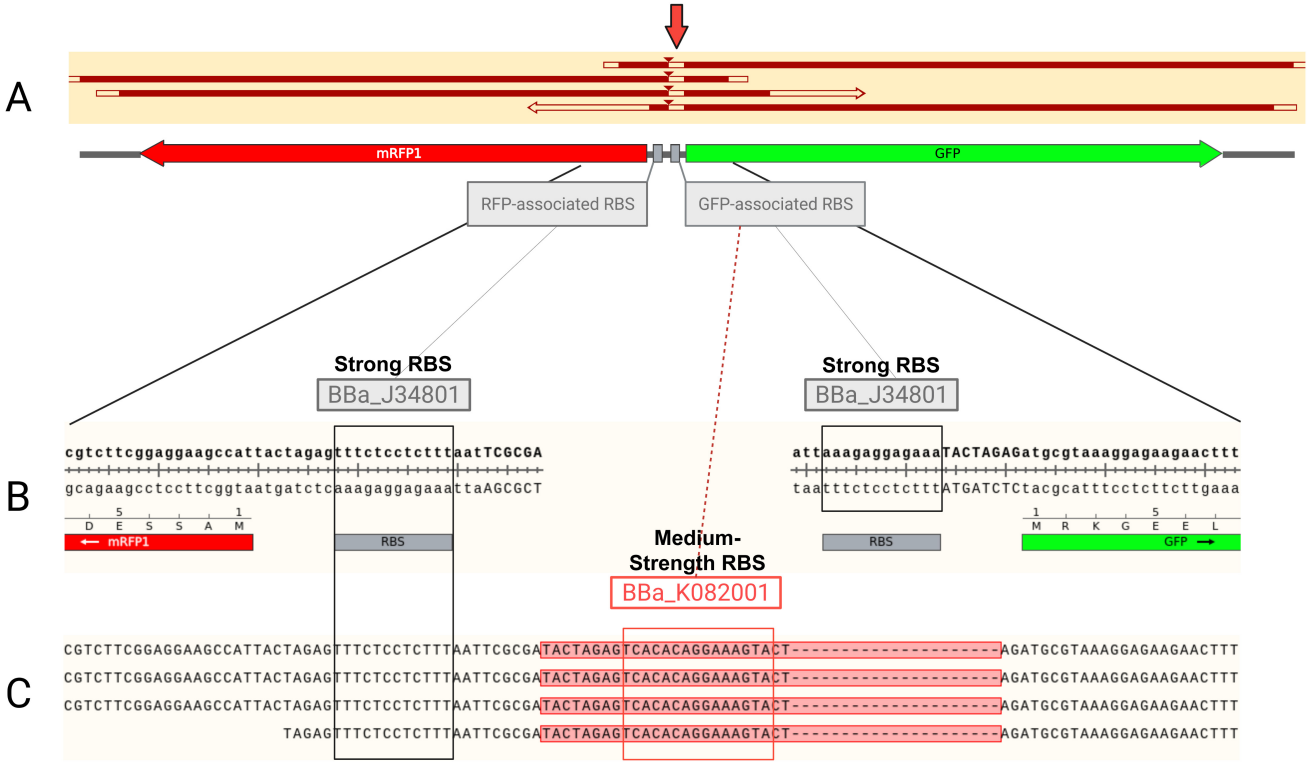

FIG. 2 Alignments of pSPPH21 Sanger sequencing results to putative pSPPH21 sequence reveal mutations in the GFP-associated RBS. pSPPH21 was sequenced using primers from Abrishamkar *et al.* and Agnew *et al.* (4, 18). **(A)** Map view of Sanger sequencing results aligned with the putative pSPPH21 sequence indicate mismatch at the GFP-associated RBS (arrow). **(B)** Sequence view of the putative pSPPH21 sequence with highlighted putative RBS and its BioBrick part name. **(C)** Sequence view of the aligned Sanger sequencing results with highlighted mismatch at the GFP-associated RBS and its BioBrick part name.

PCR amplification of the *lac*-insert from MG1655 gDNA. Genomic DNA (gDNA) extracted from *E. coli* MG1655 served as a template for the *lac*-insert. The proximity of the *lacI* promoter, *lacI*, *lac* promoter and *lac* operator in the MG1655 genome allowed for direct amplification of 1603 bp from the gDNA template using the custom primers Lac-Insert FWD and Lac-Insert REV (Fig. 3A). The PCR product was run on a 1% agarose gel. A single band at ~1600 bp indicated that the *lac*-insert was amplified from the MG1655 *E. coli* genomic DNA (Fig. 3B). No non-specific amplification was visible on the gel.

Ligation and transformation of *lac*-insert containing pSPPH21 into DH5α. pSPPH21 and *lac*-insert DNA were prepared for either simultaneous or sequential ligation as described. Overall, the number of transformants was extremely low in all three transformation trials with only 8 - 23 transformants per reaction. Nonetheless, colony PCR was performed to screen

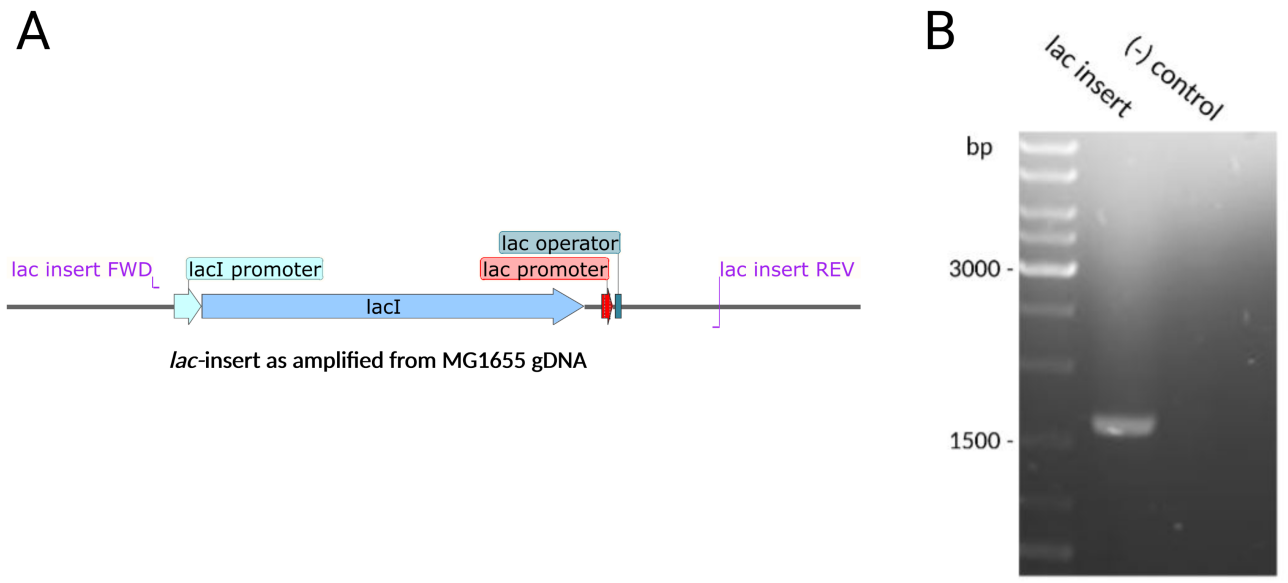

FIG. 3 Successful PCR amplification of the *lac*-insert. 1% gel electrophoresis of PCR product from amplification of the *lac*-insert from *E. coli* MG1655 gDNA using custom primers. Lane 1: DNA ladder. Lane 2: *lac*-insert. Lane 3: negative control.

transformants for successfully ligated *lac*-insert. Using the primers from Abrishamkar *et al*., amplification of empty and *lac*-insert-containing vector is predicted to result in a band at ~1600 and ~3200 bp, respectively (4). Representative colony PCR results showed successful insertion in two of eight transformants in the simultaneous protocol, compared to successful insertions in 15 of 23 transformants in the sequential protocol (Fig. 4). Thus, we obtained more transformants, as well as a greater proportion of successful ligations, with the sequential protocol.

From a total of three ligation and transformation trials, we obtained 23 clones with *lac*-insert containing pSPPH21 for further investigation of GFP and RFP expression.

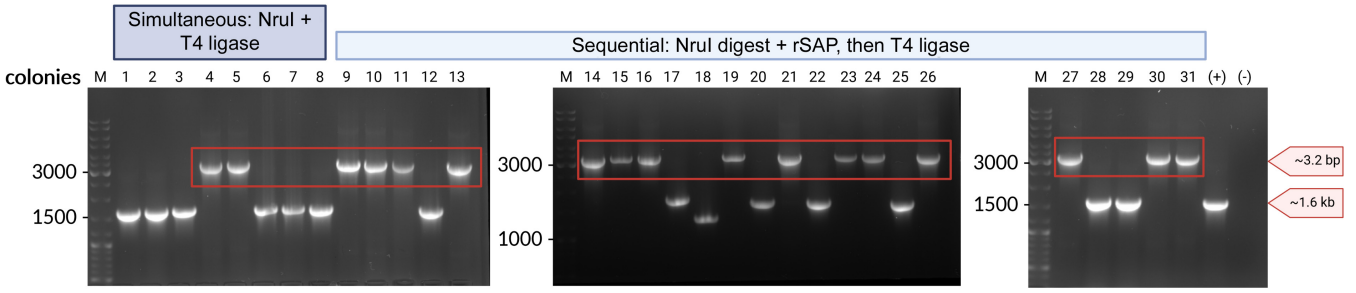

FIG. 4 Successful ligation and transformation of *lac*-insert containing pSPPH21. 1% gel electrophoresis of colony PCR products from 31 representative clones. Lanes numbered 1-8 represent clones obtained from simultaneous digestion and ligation. Lanes numbered 9-31 represent clones obtained from sequential digestion and ligation.

GFP and RFP are expressed at the predicted ~1:1 ratio. Transformants containing the *lac*-insert (n=23) were screened for GFP and RFP expression on the Bio-Rad ChemiDoc MP imaging system. Colonies expressing GFP and RFP can be visualized using settings for Alexa 546 and Alexa 488, respectively. For screening, patch plates were inoculated with the 23 transformants and incubated at 37°C overnight. LB + chloramphenicol agar was supplemented with 0.5 mM IPTG for induction, while LB + chloramphenicol plates served as uninduced control.

Out of the 23 induced clones, 10 showed GFP fluorescence and 13 showed RFP fluorescence, while none of the uninduced clones fluoresced (Fig. 5A, B). The RFP excitation overlaps slightly with GFP, as the GFP-expressing clones showed slight signal under RFP excitation. However, the 13 bright clones under RFP excitation are indicative of RFP

fluorescence (Fig. 5B). The pseudo-colored, merged channels showed a clear distinction between the GFP and RFP clones (Fig. 5C). Hence, the GFP:RFP expression ratio is 10:13. A X^2 test was performed to determine whether this ratio is significantly different from the hypothesized 1:1 ratio. A p-value of 0.91 confirmed that the observed ratio of 13:10 was not significantly different from the hypothesized 1:1 ratio. Hence, it can be concluded that pSPPH21 expresses GFP and RFP without bias.

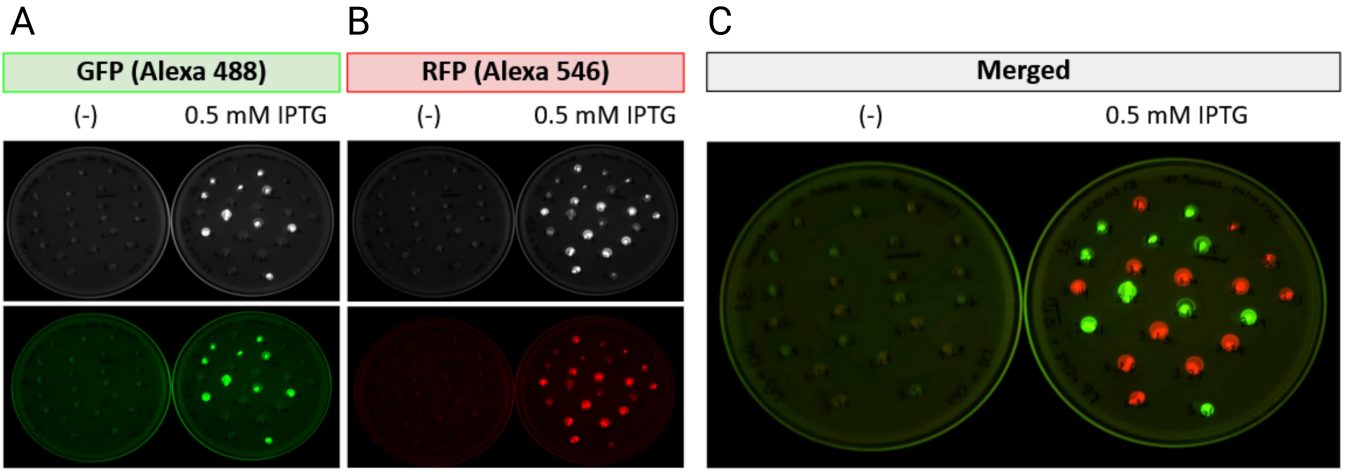

FIG. 5 Approximate 1:1 ratio of GFP to RFP demonstrated by fluorescent agar plate imaging. Patch plates of *lac*-insert containing clone colonies imaged on GelDoc. Settings for Alexa 488 and Alexa 546 were used to visualize GFP and RFP, respectively. Uninduced plates were used as negative controls. **(A)** GFP fluorescence in 10 colonies. Lower image is pseudo-coloured green to aid in visualization. **(B)** RFP fluorescence in 13 colonies. Lower image is pseudo-coloured red to aid in visualization. **(C)** Superimposed, pseudo-coloured image allowing for the visualization of RFP and GFP expressing colonies (n = 23), demonstrating a GFP:RFP ratio of 10:13.

Strong fluorescent expression of GFP and RFP upon induction. To obtain a relative quantification of fluorescence of induced versus uninduced controls, GFP and RFP expressing clones were analyzed with a fluorescent plate reader. Compared to the uninduced controls, induced cells showed a significant fold change of 80x and 253x for GFP and RFP expression, respectively (p = 0) (Fig. 6). No fluorescence was observed in empty vector controls. These results demonstrate the tight regulation of *lac* promoter activity by the *lac* repressor included in the insert design.

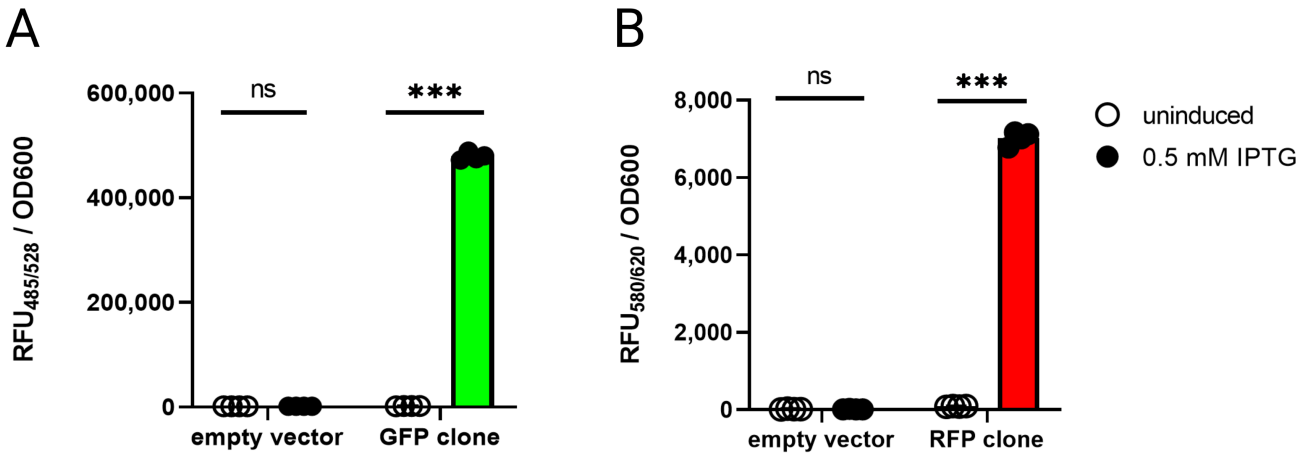

FIG. 6 Strong relative increase in fluorescence upon induction as measured by plate reader. Empty vector pSPPH21 containing cells were used as a negative control. **(A)** Average relative fluorescence units/OD600 of a GFP clone demonstrating an 80 x fold increase in fluorescence compared to the uninduced control (p = 0). **(B)** Average relative fluorescence units/OD600 of a RFP clone demonstrating a 253 x fold increase in fluorescence compared to the uninduced control (p = 0).

Sanger sequencing confirmed predicted *lac*-insert ligation. To verify the predicted blunt-ended ligation at the NruI cut site, plasmids were extracted from two random RFP- and one GFP-expressing clone and Sanger sequenced. Using SnapGene®, these sequences were then aligned with the putative sequences of *lac*-insert containing pSPPH21 in either the GFP- or RFP-expressing orientation (Fig. 7A, B). The alignments matched as expected, except at the above-mentioned GFP-associated RBS disruption (Fig. 7A, B). The sequences confirm (i) the integrity of the *lac*-insert and ligation at the NruI site, and (ii) that the orientation of the *lac*-insert matched the observed fluorescent reporter expression.

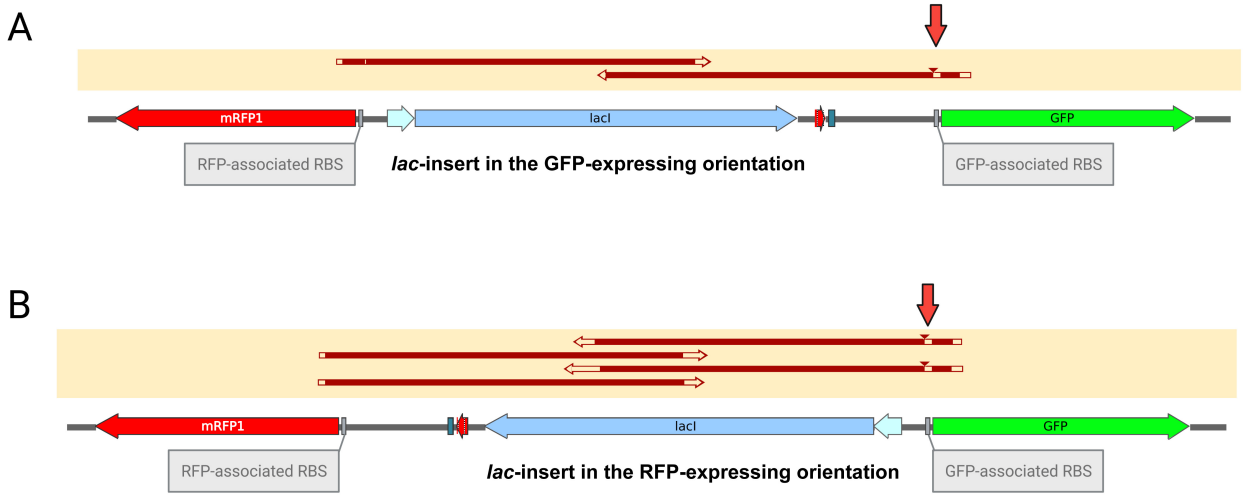

FIG. 7 Sanger sequencing confirmed the presence of *lac*-inserts in selected GFP and RFP expressing clones. *Lac*-insert-containing pSPPH21 was sequenced using primers from Agnew *et al.* (18) **(A)** Sanger sequencing results of pSPPH21 isolated from a GFP-expressing clone aligned with the putative sequence of pSPPH21 with the *lac*-insert ligated in the GFP-expressing orientation. Presence of the GFP-associated RBS mismatch is indicated by the down-pointing arrow. **(B)** Sanger sequencing results of pSPPH21 isolated from two RFP-expressing clones aligned with the putative sequence of pSPPH21 with the *lac*-insert ligated in the RFP-expressing orientation. Presence of the GFP-associated RBS mismatch is indicated by the down-pointing arrow.

DISCUSSION

Proof-of-Concept Testing. Abrishamkar *et al*. constructed the duo-directional SIGEX vector pSPPH21 to allow for detection of inducible promoters in a DNA library ligated in either direction within the plasmid (4). Before proceeding to an experimental DNA library screen for a substrate of interest, thorough proof-of-concept testing was required. Different approaches can be taken to test a SIGEX vector. Testing can be performed amidst a genuine library investigation by screening with a known inducer for which promoters are expected to be present, as performed by the 2019 UBC iGEM team (results seen here: https://2019.igem.org/Team:British_Columbia/Results#pre). This approach comes with the high stakes of committing resources to library preparation, as well as the need for access to FACS equipment. A much more conservative approach has been taken by Meier *et al.* (13). While still utilizing FACS, Meier *et al.* created a gDNA library from Lac^+ *E. coli* and screened the library in Lac^- host *E. coli* upon induction with IPTG (13).

We, however, sought the simplest and least ambiguous method to investigate whether pSPPH21 expressed GFP and RFP at a 1:1 ratio. Hence, we used PCR-amplified elements of the *lac* operon as insert instead of a gDNA library because this approach guaranteed that each insert-containing clone will fluoresce upon induction, whereas in a gDNA library only a minuscule fraction of insert-containing clones would carry a library fragment with the *lac* promoter and operator. This would have made assessment of the GFP:RFP ratio difficult due to too small sample size.

Relevance of the reduced-strength GFP-associated RBS. We confirmed the mutations at the GFP-associated RBS discovered by Abrishamkar *et al.* by Sanger sequencing of empty vector, as well as *lac*-insert containing vector (4). Instead of the expected strong RBS

(Part:Bba_J34801), we identified the presence of a different, medium-strength RBS (Bba_K082001).

The lower fold change of GFP compared to RFP expression observed in measurements with the fluorescent plate reader may provide evidence for the reduced transcription of GFP due to the reduced-strength RBS. However, these differences could also stem from inherent differences in fluorescent intensity of the two fluorescent reporters. Overall, GFP expression is well above background, indicating that GFP expression is sufficient for signal detection in pSPPH21 despite the reduced strength RBS. Therefore, we conclude that the RBS difference does not affect pSPPH21 suitability as a SIGEX vector in DH5α cells. Other screening host cells may be more sensitive to changes in RBS and their performance would need to be assessed in separate optimization experiments.

GFP and RFP controls for future experiments. The *lac*-insert was PCR-amplified from MG1655 *E. coli* gDNA and comprises the *lacI* gene and promoter, *lac* promoter, and *lac* operator. We anticipated that the high copy number of pSPPH21 would stoichiometrically outnumber the *lac* repressor expressed by the host DH5α cells. High copy number plasmids and insufficient repressor are known causes of leaky expression which may be detrimental to experiments requiring tight transcriptional regulation (14). Inclusion of the *lacI* gene and promoter in the *lac*-insert resulted in tight expression control of the fluorescent reporter genes due to the 1:1 stoichiometry of *lac* repressor to *lac* operator. Hence, no leaky expression of fluorescent reporters was observed in uninduced clones. RFP and GFP-expressing clones from this study can be used as positive controls in future experiments, as well as to optimize FACS settings.

FACS screening of false positives. In a real genomic screen, false positives from leaky expression and constitutive promoters can be screened out by using FACS to remove fluorescing cells from an uninduced culture, before proceeding to induce the remaining cells and using FACS to recover the true positives (1). By removing the false positives, the number of cells that need to be sequenced is greatly reduced. This may, however, result in eliminating cells containing leaky or constitutive promoters for the substrate of interest. Depending on the number of cells removed in the preliminary FACS, it may still be desirable to analyze the removed cells.

Ligation and transformation efficiencies. pSPPH21 contains a blunt-ended NruI cut site for library ligation. Blunt-ended ligation results in a directionally unbiased insertion of the ligated fragment (15). pSPPH21 has been intentionally designed to detect promoters ligated in either direction (4). Blunt-ended ligation has the advantage of requiring minimal library preparation, however, it is 10-100x less efficient than sticky-end ligation (15). Our results reflect the lower ligation efficiency of blunt-ended ligation as we obtained very few transformants from all of our ligation and transformation trials with only 2-15 *lac*-insert-containing transformants per reaction.

In this study, we performed ligations using two different protocols: In the simultaneous protocol, NruI and T4 ligase are present in the same reaction mixture (9, 10). In theory, this protocol favours the ligation of the insert over the ligation of the empty vector, since the NruI cut site will be disrupted upon successful insert ligation. Self-ligated vector, however, will continue to be cut by NruI. In the more conventional sequential protocol adapted from NEB protocols, pSPPH21 is linearized in a separate reaction and dephosphorylated to prevent self-ligation before being added to the ligation reaction mixture. Although the simultaneous protocol should improve ligation efficiency according to the literature, we achieved a greater number of total transformants, as well as a greater proportion of insert-containing clones with the sequential protocol (15-17). Several strategies suggested in the literature could be employed to improve blunt-ended ligation efficiency, such as addition of polyethylene glycol (PEG) to increase molecular crowding, longer incubation times, and variations of insert to vector ratios and reaction concentrations (15-17). Other SIGEX vectors utilize sticky-end or TOPO-ligation, which offer better ligation efficiencies (3, 13, 17). However, the directionality of these ligation methods results in missed hits when promoters are ligated in reverse.

Other SIGEX libraries were created by electroporation into host cells (3, 13, 17). Electroporation offers greater transformation efficiency than the heat shock method, which was utilized in this study (15). While we achieved a sufficient number of transformants to reach the conclusions of this study, a DNA library screen undertaken in earnest would require higher ligation and transformation efficiencies.

Limitations A major limitation of this study is the lack of statistical power in the form of increased sample size and number of biological replicates. A greater number of transformants would provide stronger evidence for non-biased expression of GFP and RFP. Similarly, biological replicates for fluorescence measurements on the plate reader would strengthen our estimates of fold change.

Agnew *et al.,* Frese *et al.* and MacPherson *et al.* observed rearrangements in pSPPH21 after ligation attempts, hence Sanger sequencing of our empty vector transformants could have revealed inconsistencies in the pSPPH21 structure as well (18-20). This could have elucidated whether our empty vector transformants contained self-ligated or faulty vectors.

Ultimately, a greater sample size of transformants would have allowed for greater confidence in the performance of pSPPH21 as a SIGEX vector, especially in the face of the difficulties described by Agnew *et al.*, Frese *et al.* and MacPherson *et al.* (18-20).

Conclusions Here, we confirmed the duo-directionality of pSPPH21 in expressing both GFP or RFP depending on the direction of the insert. Examination of fluorescence signals by screening clones using a GelDoc showed un-biased expression of GFP and RFP as expected for directionally un-biased ligation at the blunt-ended NruI cut site. Fluorescence measurements on a 96-well plate reader showed significant fold change of 80x for GFP fluorescent signal and 253x fold change for RFP in induced clones compared to uninduced clones. With the *lac*-insert, we created a tightly controlled inducible system for testing SIGEX vectors. The discovered inconsistencies between the GFP- and RFP-associated RBS did not provide a challenge for GFP expression in DH5α cells. However, ligation and transformation efficiencies have been a major challenge throughout this study.

Future Directions With the functionality of pSPPH21 confirmed, more complex experiments with this vector are reasonable to pursue. Improvement of the ligation and transformation efficiencies is paramount for the application of pSPPH21 in a metagenomic screen. As mentioned above, some approaches could include the addition of PEG, longer incubation times, different insert-to-vector ratios and reaction concentrations in the ligation reaction, and electroporation as a transformation method. The next step could then be to use pSPPH21 in a gDNA library screen using a known inducer. Ideally, a known promoter would then be retrieved from the library.

To optimize FACS settings, a small quantity of the control cells produced in this study could be spiked in with a cell sample.

One caveat of SIGEX screening is that promoters from species taxonomically distant from the screening host cell (commonly *E. coli*) may not be perfectly compatible with the screening host's RNA polymerase. Therefore, it may be of value to confirm that inducible promoters from taxonomically distant species result in fluorescent reporter expression. The groundwork for such a screen has already been laid by Frese *et al.* (19). The incompatibility between promoters and host transcription machinery could potentially be overcome by using different host cells containing the necessary transcription machinery for those promoters (1). Once the use of pSPPH21 has been optimized and sensitivity has been further confirmed, this vector could be used to screen metagenomic libraries with different substrates, and in different screening hosts.

ACKNOWLEDGEMENTS

We acknowledge that this land we work, recharge and feel at home on is the traditional, ancestral, and unceded territory of the xwməθkwəy̓əm (Musqueam) People. This project was funded by the Department of Microbiology and Immunology at the University of British Columbia. We thank Dr. David Oliver, Jade Muileboom, and Gara Dexter for their mentorship, technical advice and constant support throughout this project. Their feedback and

guidance were invaluable throughout the experimental design and troubleshooting process. We further thank Abrishamkar *et al.* and Agnew *et al*. for providing the pSPPH21 plasmid and primers. Lastly, we thank Frese *et al.* and MacPherson *et al.* for the lively discussions. We would also like to thank two anonymous reviewers for constructive feedback on this manuscript.

CONTRIBUTIONS

D.G. generated the figures and wrote the introduction, Sanger sequencing-related results, and discussion. R.F. wrote the abstract, sections of the methods, figure captions, study limitations and the future directions. Y.L. wrote sections of the methods, results, figure captions, and wrote the conclusions. K.K. wrote sections of the methods, results, and references. All authors contributed to the conduction of experiments and editing the manuscript.

REFERENCES

1. **Uchiyama T, Watanabe K.** 2007. The SIGEX Scheme: High Throughput Screening of Environmental Metagenomes for the Isolation of Novel Catabolic Genes. Biotechnol Genet Eng Rev **24**:107–16.
2. **Uchiyama T, Abe T, Ikemura T, Watanabe K.** 2005. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. Nat Biotechnol **23**:88–93.
3. **Uchiyama T, Watanabe K.** 2008. Substrate-induced gene expression (SIGEX) screening of metagenome libraries. Nat Protoc **3**:1202–1212.
4. **Abrishamkar P, Dana HM, Oveisi S, Tabassi P.** 2022. Construction of SIGEX duo-directional reporter plasmid pSPPH21. UJEMI+ (submitted for publication).
5. **Jacob F, Monod J.** 1961. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol **3**:318–356.
6. **Beckwith JR.** 1967. Regulation of the *Lac* Operon: Recent studies on the regulation of lactose metabolism in *Escherichia coli* support the operon model. Science **156**:597–604.
7. Benchling [Biology Software]. 2022. Retrieved from https://benchling.com.
8. SnapGene software (from Insightful Science; available at snapgene.com)
9. **Green MR, Sambrook J.** 2012. Molecular cloning: a laboratory manual 4th ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
10. **Costa GL, Weiner MP.** 1994. Protocols for cloning and analysis of blunt-ended PCR-generated DNA fragments. Genome Res **3**:S95–S106.
11. **Chang A, Chau V, Landas J, Pang Y.** 2017. Preparation of calcium competent *Escherichia coli* and heat-shock transformation. UJEMI **1:**22-25
12. **Kuiper S.** 2019. GFP-RFP plate reader assay. Protocols.io https://dx.doi.org/10.17504/protocols.io.784hryw
13. **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.
14. **Anthony LC, Suzuki H, Filutowicz M.** 2004. Tightly regulated vectors for the cloning and expression of toxic genes. J Microbiol Methods **58**:243–250.
15. **Bercovich JA, Grinstein S, Zorzopulos J.** 1992. Effect of DNA concentration on recombinant plasmid recovery after blunt-end ligation. Biotechniques **12**:190, 192–193.
16. **Damak S, Bullock DW.** 1993. A simple two-step method for efficient blunt-end ligation of DNA fragments. Biotechniques **15**:448–450, 452.
17. **Wakamatsu T, Morono Y, Futagami T, Terada T, Nishikawa S, Morisawa T, Ohshita K, Inagaki F, Ashiuchi M.** 2018. Metal-ion-induced expression of gene fragments from subseafloor micro-organisms in the Kumano forearc basin, Nankai Trough. J Appl Microbiol **125**:1396–1407.
18. **Agnew S, Bangayan A, Fan J, Lin A. 2022.** Testing the functionality of SIGEX duo-directional reporter plasmid pSPPH21 using an inducible promoter. UJEMI (submitted for publication).
19. **Frese K, Martens K, Nichvolodoff.** 2022. Multiple deletions arise from attempted insertion of the salicylate nahR/Psal biosensor into a duo-directional SIGEX vector. UJEMI (submitted for publication).
20. **McPherson K, Makhi J, Haroon M, Musa A.** 2022. Cloning optimization for substrate-induced gene expression technology. UJEMI (submitted for publication).