Volume 8:1-11

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

Construction of SIGEX duo-directional reporter plasmid pSPPH21

Parsa Abrishamkar, Helia Mansouri Dana, Sourena Oveisi, Parsa Tabassi

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

SUMMARY Substrate-induced gene expression (SIGEX) methods use gene reporter systems to screen metagenomic libraries and identify novel catabolic genes. Commonly, these reporter systems incorporate promoter-less fluorescent protein genes downstream of an overhang-producing restriction site which allows metagenomic DNA fragments to be inserted and analyzed. SIGEX methods, however, lack the ability to detect genes which are inserted in the reverse orientation based on the unidirectional design of the SIGEX vector as determined by the overhangs produced during the cloning process. The insertion of a second oppositely oriented fluorescent reporter gene and a blunt-end producing restriction site would circumvent this limitation. In this study, we adapted a vector design from UBC iGEM to construct a duo-directional SIGEX plasmid reporter system (pSPPH21) containing green fluorescent protein (GFP) and red fluorescent protein (RFP) reporter genes. To build plasmid pSPPH21, a 766 base pair DNA fragment containing an NruI restriction site and an RFP gene was synthesized, digested, and cloned into the pSB1C3 vector containing a GFP and chloramphenicol resistance gene to construct the duo-directional reporter vector. Successful assembly of the vector was verified using gel electrophoresis and Sanger sequencing. The location of the NruI restriction site between the oppositely oriented GFP and RFP genes allows for high efficacy cloning of inducible promoters from metagenomic libraries and for the identification of novel catabolic genes.

INTRODUCTION

More than 99% of microorganisms in the natural environment cannot be cultivated using standard laboratory techniques, and due to difficulties in isolating microorganisms in pure culture, large portions of microbial diversity remain unknown (1, 2). This has led to increased usage of metagenomics, where genes of interest are isolated from a mixed microbial genome from the environment without cultivation or isolation of individual microorganisms (2, 3). For isolating catabolic genes, the two current approaches used in metagenomics are enzyme activity-based screening and nucleotide sequence-based screening (3). In a sequence-based analysis, the metagenomic DNA will be analyzed for the presence of a diagnostic nucleotide sequence, while a functional-based screening will rely on successful heterologous expression of gene products in a host organism (4). The limitation of these two approaches is that catabolic gene expression is usually determined by the presence of a relevant compound and is usually controlled by a regulatory element (3). This reduces the number of novel catabolic operons that can be isolated using these two methods.

Substrate-induced gene expression (SIGEX) is a new high throughput method which involves using a gene reporter system to screen a metagenomic library and identify metagenomic clones of interest based on the expression of a downstream fluorescent reporter gene (5). A SIGEX procedure includes a cloning vector (an operon-trap vector) containing a multiple cloning site upstream of a promoter-less fluorescence marker gene, commonly GFP (2,3). A metagenomic library is first obtained by fragmenting DNA from environmental

Published Online: September 2022

Citation: Parsa Abrishamkar, Helia Mansouri Dana, Sourena Oveisi, Parsa Tabassi. 2022. Construction of SIGEX duo-directional reporter plasmid pSPPH2. UJEMI+ 8:1-11

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: Parsa Abrishamkar, parsa.abrishamkar@gmail.com

samples, then cloning these fragments into the cloning vector (3). The activation of cloned metagenomic promoters in the presence of inducing substrates is then determined by observing the expression of the fluorescent marker, and clones positive for fluorescence are sorted using fluorescence activated cell sorting (FACS) (2). Cells of interest are then isolated and characterized via sequencing (2). Ultimately, this allows for the identification of catabolic genes whose expression is modulated by the presence of environmental stimuli (3). This approach can be used to identify novel catabolic genes for which enzymatic activities are difficult to detect, and for isolating novel genes with genetic sequences which significantly differ from those of known genes (3). A limitation of the SIGEX vector design is that it does not allow for the identification of catabolic genes that are distantly located relative to the relevant transcriptional regulator, since both the gene and its regulatory domain must be present within the construct vector to allow for gene expression to occur (5). Another limitation to SIGEX plasmids is their directional structure, which only allows detection of inducible promoters oriented in the same direction as the reporter gene (5).

The directional design of SIGEX plasmids means that genes inserted in the reverse orientation relative to the reporter gene will not be detected (5). A possible solution for this would involve the use of a vector encoding reporter genes in both orientations so that promoter activity can be detected from all insertions. Previously, the UBC iGEM team explored the use of a duo-directional reporter system with a vector containing a forward-oriented GFP reporter gene and a reverse-oriented RFP reporter gene for the discovery of novel genes which act as transcription factor-based biosensors of harmful environmental biotoxins (https://2019.igem.org/Team:British_Columbia). In particular, their work targeted the discovery of regulatory units that were induced in the presence of saxitoxin, a harmful marine neurotoxin, and they were able to successfully identify multiple potential biosensor candidates based on increased fluorescence in samples with saxitoxin present (https://2019.igem.org/Team:British_Columbia/Results). Due to time constraints, they were unable to sequence their biosensor candidates or optimize their plasmid vector further, and unfortunately, their plasmid could not be recovered from freezer stocks while documentation around its construction was limited. Here, we revisit this project and describe the construction of pSPPH21, a duo-directional SIGEX reporter plasmid. This plasmid is identical in its design with the screening plasmid previously created by the UBC iGEM team (https://2019.igem.org/Team:British_Columbia/Experiments) and we hope that its re-creation would allow for further investigation into its structure, its optimization and its future applications.

METHODS AND MATERIALS

Bacterial strains and plasmids. The original pSB1C3 vector was isolated from *Escherichia coli* DH5α (https://parts.igem.org/Part:pSB1C3#:~:text=pSB1C3%20is%20a%20high%20copy,reading%20out%20into%20the%20vector), and was obtained as a streaked plate from the UBC iGEM team. Commercially available, chemically competent *E. coli* DH5α cells (ThermoFisher) were purchased and were used for transformation and propagation of vectors in this study.

Luria Bertani (LB) and chloramphenicol media preparation. LB broth was prepared using tryptone, yeast extract, NaCl, and distilled water, with the recipe adapted from Hancock Lab Methods(http://cmdr.ubc.ca/bobh/method/media-recipes/). For agar plates, 20 g agar per liter was added to the LB medium prior to autoclaving. LB agar plates with 20 ng/ul chloramphenicol were prepared using a 34 mg/ml chloramphenicol stock dissolved in ethanol.

PCR amplification of RFP construct. The RFP construct was ordered from GeneWiz and amplified using high-fidelity Platinum ™ SuperFi ™ DNA polymerase according to the manufacturer's instructions (ThermoFisher), using RFP forward primer (5'-CTTCGAATTCGCGGCCGCTTCTAG-3') and RFP reverse primer (5'-CTTCCTGCAGCGGCCGCTACTAGT - 3') (GeneWiz). pUC19 amplification was used as positive control, with pUC19 forward (5’- CCCAGTCACGACGTTGTAAAACG -3’) and pUC19 reverse (5’- AGCGGATAACAATTTCACACAGG- 3’) primers (Invitrogen), and a

no template reaction mix was used as negative control. PCR was performed with an initial denaturation step at 98°C for 30 seconds, followed by 35 rounds of PCR according to the Platinum™ SuperFi™ DNA Polymerase protocol (ThermoFisher). Denaturation was set at 98°C for 10 seconds, annealing was set at 60°C for 10 seconds, and extension was set at 72°C for 40 seconds. The PCR products were analyzed following resolution a 1% agarose gel (120 V for 65 minutes) and stained with RedSafe nucleic acid staining solution (FroggaBio).

Digestion and ligation of RFP construct into pSB1C3. Restriction sites were identified on the RFP construct and pSB1C3 vector and selected according to the iGEM list of BioBrick restriction sites, recommended for DNA digestion and cloning experiments (http://parts.igem.org/Help:Standards/Assembly/RFC10). The pSB1C3 vector was double digested with EcoRI-HF and XbaI (New England BioLabs). The PCR fragment encoding RFP was double digested with EcoRI-HF and SpeI (New England BioLabs). The digested XbaI site on the vector and the digested SpeI site on the construct are compatible for ligation (https://international.neb.com/applications/cloning-and-synthetic-biology/dna-assembly-and-cloning/biobrick-assembly). Digestion was completed at 37°C for 15 minutes. Ligation was performed using a 3:1 molar ratio of insert (60 ng) to vector (20 ng) and using 1X T4 DNA ligase at room temperature for 10 minutes (New England BioLabs). Purification of DNA was completed after each digestion/ligation reaction using a PCR cleanup kit (ThermoFisher).

Transformation of pSPPH21 into *E. coli* DH5α. Transformation of *E. coli* DH5α with pSPPH21 (and control plasmids) was performed using heat-shock method (6). Undigested, circular pSB1C3 was used as positive control for transformation. Transformation with pSB1C3 linearized once by XbaI, and once by EcoRI-HF, were used to control for the functionality of the restriction enzymes. Transformed cells were spread-plated on LB plates with 20 ng/ul chloramphenicol for selection, and grown overnight in a 37°C incubator. Colonies were selected from plates and propagated in overnight LB and 20 ng/ul chloramphenicol broth for plasmid isolation.

Plasmid isolation and quantification. Isolation of vectors was done using the alkaline lysis method via EZ-10 Spin Column Plasmid DNA Miniprep (Kit BS413; BioBasic). The final pSPPH21 vector was propagated in and isolated from *E. coli* DH5α, grown on LB and 20 ng/uL chloramphenicol plates. Quantification of DNA and assessment of DNA concentration and purity were carried out using a NanoDrop2000® Spectrophotometer (ThermoFisher) at 260 nm.

Agarose DNA gel electrophoresis. DNA concentration of vectors was calculated based on NanoDrop2000 spectrophotometer readings at 260 nm. Undigested, circular pSB1C3 was used as positive control. Six colonies were selected for plasmid isolation. Undigested, circular vectors from each colony were included as controls for restriction enzyme digestion. Isolated plasmids from each colony were linearized with NruI single digest, XbaI and NruI double digest, and XbaI and NruI sequential digest for gel analysis. Sequential digestion was performed by XbaI digestion with a no-salt buffer (rCutSmart) (New England BioLabs). Salt concentration was then adjusted to 100 mM for digestion with NruI. DNA samples were mixed with 6X DNA green loading dye (ThermoFisher) to allow loading of 200 ng of DNA per well with a final 1X loading dye concentration. 100 mL 1.0% agarose gels were prepared using 1.0 g UltraPure® Agarose powder (ThermoFisher) and 1X TBE buffer, and Invitrogen DNA 1 kb Plus Ladder was loaded at 200 ng alongside the samples in each gel. For the PCR gel, a 100 bp ladder was also included (Invitrogen). Gels were run at 105 V for 65 minutes, and imaging was done using an ultraviolet (UV) gel imaging system (BioRad).

Sanger sequencing. Samples for Sanger sequencing were prepared according to instructions by GeneWiz. Two 500 ng aliquots of pSPPH21 from colony 5 were prepared. One was submitted with the pSPPH21_RFP forward primer (5'- GGC GTA TCA CGA GGC AGA ATT TC - 3'), and one was submitted with the pSPPH21_RFP reverse primer (5'- GGA AGC CTG CAT AAC GCG AAG -3'). Sequencing results were compared to the pSPPH21

reference sequence (A. Noonan, unpublished document) for determining sequence identity of the RFP insert- including the NruI site- using the NCBI alignment tool.

RESULTS

We first obtained a text file containing the nucleotide sequence for an RFP insert (Figure 1A), containing the RFP gene, a ribosome binding site, and an NruI cut site (Avery Noonan, personal communication). We also obtained the pSB1C3 vector from the UBC iGEM team (Figure 1B), containing the GFP gene and the chloramphenicol resistance gene. The aim was to ligate the RFP construct into the pSB1C3 vector by utilizing XbaI, SpeI, and EcoRI restriction sites within the BioBrick prefix and suffix sequences located on the construct and the vector, hence creating the pSPPH21 duo-directional reporter vector (Figure 1C). To confirm the existence of the pSPPH21 vector in the transformants, we performed DNA gel electrophoresis and Sanger sequencing of the isolated plasmid DNA.

PCR amplification yields higher concentration of the RFP construct. Several attempts were made to ligate the RFP construct into the digested pSB1C3 vector but no colonies were obtained that contained the insert. It was surmised that the initial concentration of the RFP

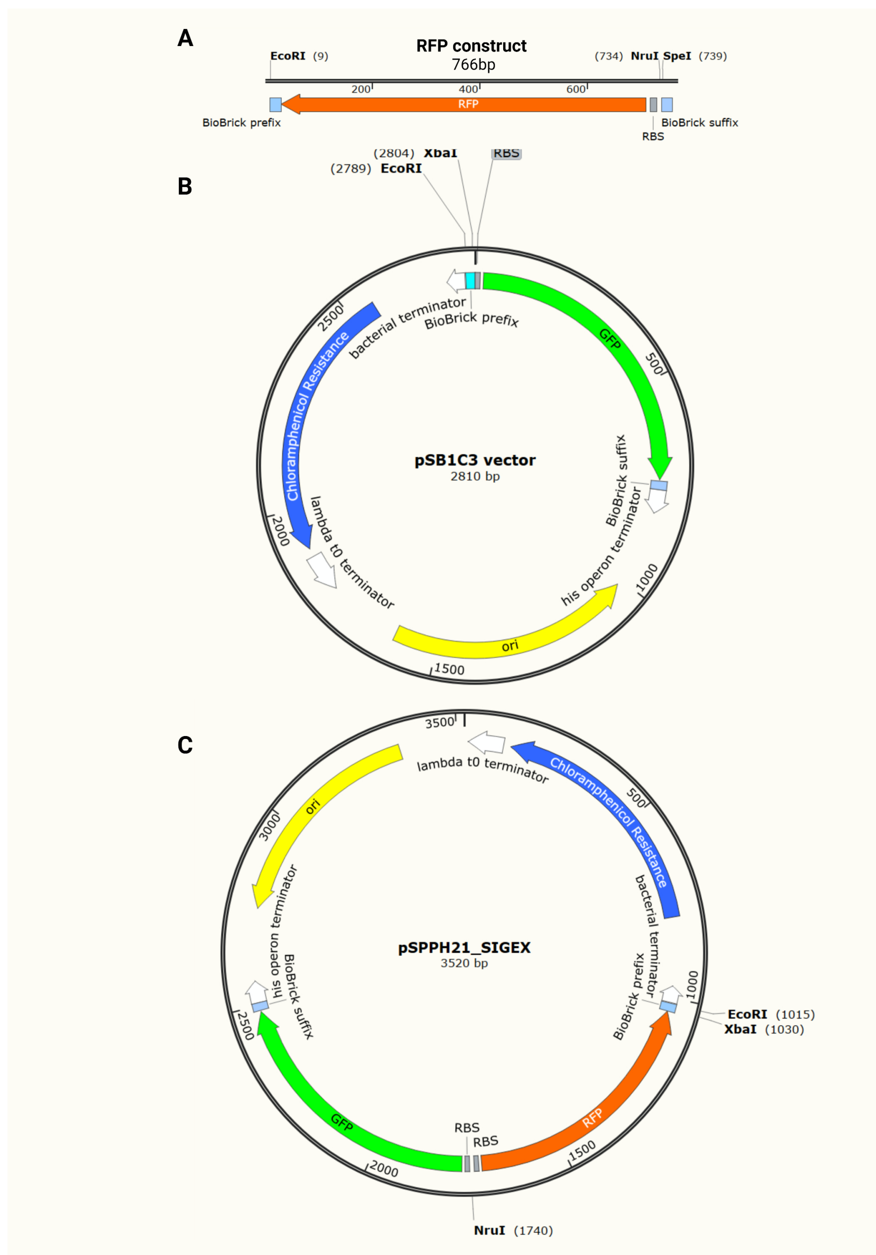

FIG. 1 Graphical Depiction of DNA plasmids and construct. Graphical depiction of the RFP construct (A), the pSB1C3 vector (B), and the pSPPH21 (C).

construct (28.6 ng/ul) was too low to allow for its successful insertion into the digested pSB1C3 vector. Therefore, to raise the RFP construct concentration, we designed primers for amplification of the RFP construct and performed PCR, raising its concentration to 35.1 ng/ul. PCR products were resolved on 1% agarose gel and produced a band at approximately 700-800 bp, indicating the presence of the 766 bp long RFP construct (Figure 2). While the increase in the concentration of the RFP construct was minor, its amplification may have increased the DNA quality, hence increasing the chance for successful ligation into the pSB1C3 vector. pUC19 plasmid was amplified as well, intended to serve as a positive control; however, it did not yield a band on the gel (Figure 2). This may have been due to the far lower concentration of the pUC19 DNA template (0.1 ng/ul) when compared to that of the RFP construct (28.6 ng/ul), causing pUC19 not to get amplified during the PCR experiment. A water sample with no DNA was used as negative control and it did not yield any bands as expected, indicating the absence of DNA contamination.

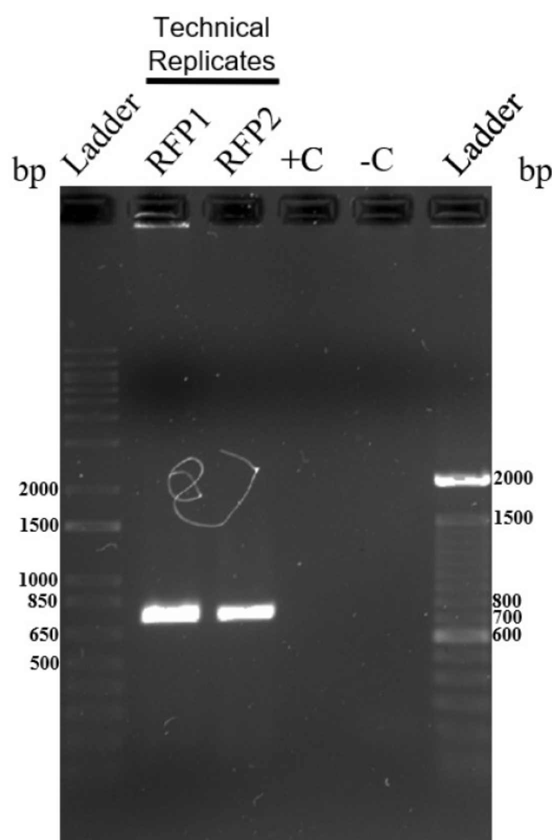

FIG. 2 1% agarose DNA gel electrophoresis of the RFP construct PCR products. RFP1 and RFP2 are technical replicates of the PCR-amplified RFP constructs. pUC19 plasmid was amplified as a positive control (+C), however, a band did not appear on the gel. PCR amplification with no template DNA was done as a negative control (-C). A 1kb and a 100 bp ladder were resolved on the left-most and right-most lanes, respectively.

pSB1C3 vector digested with XbaI and EcoRI-HF yields no transformants, verifying functionality of the restriction enzymes. To control for the functionality of both XbaI and EcoRI-HF restriction enzymes cutting the pSB1C3 plasmid, we linearized the pSB1C3 vector with XbaI and EcoRI-HF single digests, and then transformed *E. coli* DH5α cells with each linearized pSB1C3 vector (Figures S1B and S1C). It was expected that no colonies would be observed on plates containing chloramphenicol, as linearization of the pSB1C3 vector by each restriction enzyme would prevent replication of the vector; hence, no daughter cells would contain the vector in order to express chloramphenicol resistance. As expected, no colonies were observed on chloramphenicol plates, indicating that both restriction enzymes cutting the pSB1C3 plasmid were functional.

NruI and XbaI digestion of the pSPPH21 vector yields two DNA fragments. To construct the pSPPH21 vector, we cloned the PCR-amplified RFP construct into the pSB1C3 vector and obtained colonies on plates containing chloramphenicol (Figure S1D). The presence of colonies validates the re-circularization of the plasmid following digestion and ligation of the

RFP construct, as these colonies are able to express chloramphenicol resistance. However, it was possible that some of these colonies contained the uncut pSB1C3 vector that survived the digestion procedure as both the pSPPH21 vector and the pSB1C3 vector confer chloramphenicol resistance. Therefore, it was necessary to distinguish the colonies that contained the pSB1C3 vector from the ones that contained the pSPPH21. The size difference between the two vectors and the existence of the NruI restriction site on the pSPPH21 were used to differentiate the two vectors when resolved on a gel. It was expected that digestion of the pSPPH21 vector would generate a linear 3520 bp long fragment. It was also expected that digestion of the pSPPH21 vector with both NruI and XbaI would generate a shorter 710 bp fragment and a longer 2810 bp fragment.

DNA gel electrophoresis results suggest the absence of the pSPPH21 vector in colonies 4 and 6, but its presence in colony 5. To ensure the detection of the 710 bp and 2810 bp DNA fragments, the vector sample from each colony was digested, once in a double digestion reaction and once in a sequential digestion reaction. This was done because it was speculated that the reaction buffer used in the double digest reaction had higher than optimal salt concentration for XbaI activity and could lead to inefficient XbaI digestion; therefore, a sequential digest reaction would allow for efficient XbaI digestion followed by increasing the salt concentration before the addition of NruI. Gel results for colonies 4 and 6 suggest that these colonies do not contain the pSPPH21 vector (Figure 3). This is because the NruI single digest for these colonies yields a smeared pattern close to 2800 bp, similar to their uncut plasmid smear pattern, suggesting that both plasmid samples from these colonies did not contain the NruI cut site and had a smaller size than the 3520 bp pSPPH21 vector. Furthermore, the sequential digest of plasmids from both of these colonies generated only a single band at around 2800 bp, suggesting that their plasmid was cut only by one of the two restriction enzymes. In addition, the double digestion of colonies 4 and 6 show multiple bands, with the brightest being around 2800 bp, suggesting partial linearization of the corresponding plasmids. This partial linearization was likely done by XbaI as the single digest by NruI alone suggested the absence of an NruI cut site on the plasmids from colonies 4 and 6. As a result, evidence from the gel suggests that the pSPPH21 vector was not present in the plasmids from colonies 4 and 6. It can be observed that the double digestion of colony 5 plasmid shows a bright band at around 3500 bp, corresponding to linearized vectors, and a pale band at around 2800 bp likely corresponding to only a small number of those linearized vectors that got successfully cut by XbaI, hence suggesting suboptimal XbaI activity in the double digest setting (Figure 3). However, the sequential digestion of colony 5 plasmid generates a shorter band at around 700 bp and a larger band at around 2800 bp, similar to the fragment sizes that were expected to be generated from digestion of pSPPH21 with XbaI and NruI. Furthermore, digestion of colony 5 plasmid with NruI alone, generates a single bright band at around 3500 bp which is similar to the size of the pSPPH21 vector. As a result, evidence from the gel suggests that colony 5 plasmid could be the pSPPH21 vector.

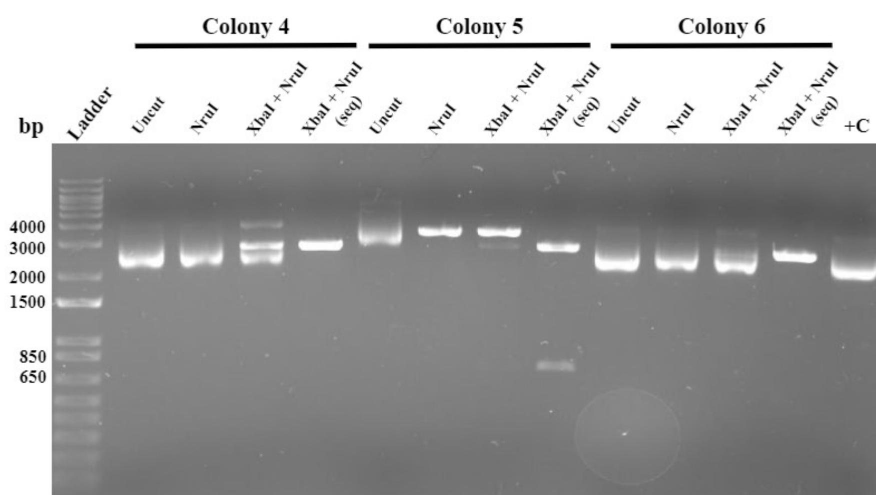

FIG. 3 1% agarose DNA gel electrophoresis of plasmids extracted from candidate DH5α colonies transformed with the constructed pSPPH21 vector. Plasmid samples from each colony were run undigested (uncut), single digested with NruI, double digested with XbaI and NruI (XbaI + NruI), and sequentially digested with XbaI and NruI (XbaI + NruI (seq)). Undigested pSB1C3 vector acted as a positive control (+C). A 1kb DNA ladder was run as well.

Sequencing results suggest the presence of RFP construct in the pSPPH21 vector. Following the gel electrophoresis results, Sanger sequencing was performed to confirm the presence of the NruI and RFP construct in the resultant vector from colony 5. The NruI restriction site and RFP gene are unique to the RFP construct that was initially inserted into the vector, hence the identification of these sequences using our custom forward and reverse primers would validate the successful assembly of the pSPPH21 vector (Figure 4). Analysis of the sequencing results from the pSPPH21_RFP forward primer revealed that the construct was inserted with the NruI site located at base pairs 764 - 769 and upstream of the RFP gene (Figure 4, Figure 5). The pSPPH21_RFP reverse primer reveals the presence of the GFP gene found in the original pSB1C3 backbone, though it shows mutations in the ribosome binding site (RBS) upstream the GFP gene. Nonetheless, the sequence derived from the pSPPH21_RFP reverse primer overlaps with the sequence obtained from the pSPPH21_RFP forward primer at the NruI restriction site (Figure 5), confirming the creation of pSPPH21.

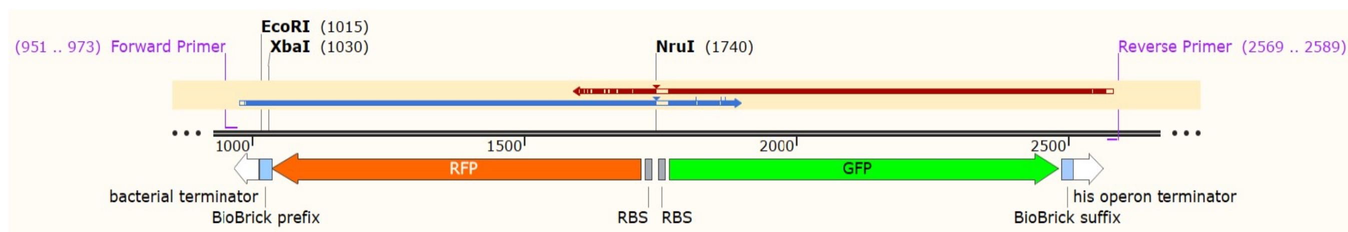

FIG. 4 Graphical depiction of the alignment of the sequenced results within the pSPPH21 vector. Solid blue arrow above the plasmid map represents the sequenced result from the pSPPH21_RFP forward primer. Solid red arrow above the plasmid map represents the sequenced result from the pSPPH21_RFP reverse primer. Gaps in the solid arrow indicate regions with uncertain identity of nucleotides. The RFP gene (orange arrow) and NruI restriction site are found within the aligned sequence of the pSPPH21_RFP forward primer. The GFP gene (green arrow) and NruI restriction site are aligned with the sequence of the pSPPH21_RFP reverse primer.

DISCUSSION

SIGEX technology uses reporter genes in vectors, which can be activated by upstream cloning of DNA fragments from metagenomic libraries (3). However, SIGEX vectors typically contain a single reporter, which limits its application due to a need for specific directionality of the inserted fragment (3). As such, we aimed to construct a novel duo-directional vector containing a blunt-ended NruI restriction site, flanked by GFP and RFP genes oriented in opposing directions. To confirm the insertion of the RFP insert into the pSB1C3 vector, DNA gel electrophoresis (Figure 3) and Sanger sequencing (Figure 5) were done. Confirmed double digestion and ligation of the RFP insert with pSB1C3 thus resulted in the creation of the pSPPH21 duo-directional reporter plasmid, outlined in Figure 1.

Features of the pSPPH21 construct. The pSPPH21 vector uses the pSB1C3 backbone (Figure 1) and consists of two fluorescence reporters, GFP and RFP, in opposite orientations, with a ribosome binding site (RBS) upstream and a terminator sequence downstream of each gene. The RBS sequence upstream of the GFP gene however contains mutations. The reference sequence of pSB1C3 consists of 12 guanine and adenine bases (5'- ATT AAA GAG GAG AAA -3'), followed by a 5 nucleotide spacer upstream the GFP start codon. On the other hand, the obtained sequence of pSPPH21 contains a shorter sequence of guanine and adenine repeats (5'- AGG AAA G-3') followed by a 4 nucleotide spacer, though it still theoretically contains the *E. coli* Shine-Dalgarno consensus sequence (5'- AGG AGG -3'). Additionally, the RBS spacing remains within the optimal range of 4-9 nucleotides required for GFP expression (7). Despite this, we cannot infer functional efficiency of GFP expression, as this is outside the scope of our experiment. A possible reason for this is the presence of mutations in the original pSB1C3 vector. Between and upstream the two fluorescence markers, an NruI restriction site is present to facilitate the blunt-ended insertion of metagenomic DNA fragments for SIGEX applications. The plasmid further consists of a chloramphenicol resistance marker, as well as a high copy-number (100-300 copies per cell)

origin of replication derived from the ColE1/pMB1 vector family (https://parts.igem.org/Part:pSB1C3), suitable for replication in *E. coli* (8). A BioBrick prefix is present downstream of the RFP gene, and a BioBrick suffix is present downstream of the GFP gene.

```
E-value: 0.0        Identity: 901/919 (98%)

13  TAGCTTTCGCTAAGGATGATTTCTGGAATTCGCGGCCGCTTCTAGAGTTAAGCACCGGTG  72   <(Query)
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
40  TAGCTTTCGCTAAGGATGATTTCTGGAATTCGCGGCCGCTTCTAGAGTTAAGCACCGGTG  99   <(Subject)

73  GAGTGACGACCTTCAGCACGTTCGTACTGTTCAACGATGGTGTAGTCTTCGTTGTGGGAG  132
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
100 GAGTGACGACCTTCAGCACGTTCGTACTGTTCAACGATGGTGTAGTCTTCGTTGTGGGAG  159

133 GTGATGTCCAGTTTGATGTCGGTTTTGTAAGCACCCGGCAGCTGAACCGGTTTTTTAGCC  192
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
160 GTGATGTCCAGTTTGATGTCGGTTTTGTAAGCACCCGGCAGCTGAACCGGTTTTTTAGCC  219

193 ATGTAGGTGGTTTTAACTTCAGCGTCGTAGTGACCACCGTCTTTCAGTTTCAGACGCATT  252
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
220 ATGTAGGTGGTTTTAACTTCAGCGTCGTAGTGACCACCGTCTTTCAGTTTCAGACGCATT  279

253 TTGATTTCACCTTTCAGAGCACCGTCTTCCGGGTACATACGTTCGGTGGAAGCTTCCCAA  312
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
280 TTGATTTCACCTTTCAGAGCACCGTCTTCCGGGTACATACGTTCGGTGGAAGCTTCCCAA  339

313 CCCATGGTTTTTTTCTGCATAACCGGACCGTCGGACGGGAAGTTGGTACCACGCAGTTTA  372
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
340 CCCATGGTTTTTTTCTGCATAACCGGACCGTCGGACGGGAAGTTGGTACCACGCAGTTTA  399

373 ACTTTGTAGATGAACTCACCGTCTTGCAGGGAGGAGTCCTGGGTAACGGTAACAACACCA  432
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
400 ACTTTGTAGATGAACTCACCGTCTTGCAGGGAGGAGTCCTGGGTAACGGTAACAACACCA  459

433 CCGTCTTCGAAGTTCATAACACGTTCCCATTTGAAACCTTCCGGGAAGGACAGTTTCAGG  492
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
460 CCGTCTTCGAAGTTCATAACACGTTCCCATTTGAAACCTTCCGGGAAGGACAGTTTCAGG  519

493 TAGTCCGGGATGTCAGCCGGGTGTTTAACGTAAGCTTTGGAACCGTACTGGAACTGCGGG  552
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
520 TAGTCCGGGATGTCAGCCGGGTGTTTAACGTAAGCTTTGGAACCGTACTGGAACTGCGGG  579

553 GACAGGATGTCCCAAGCGAACGGCAGCGGACCACCTTTGGTAACTTTCAGTTTAGCGGTC  612
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
580 GACAGGATGTCCCAAGCGAACGGCAGCGGACCACCTTTGGTAACTTTCAGTTTAGCGGTC  639

613 TGGGTACCTTCGTACGGACGACCTTCACCTTCACCTTCGATTTCGAACTCGTGACCGTTA  672
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
640 TGGGTACCTTCGTACGGACGACCTTCACCTTCACCTTCGATTTCGAACTCGTGACCGTTA  699

673 ACGGAACCTTCCATACGAACTTTGAAACGCATGAACTCTTTGATAACGTCTTCGGAGGAA  732
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
700 ACGGAACCTTCCATACGAACTTTGAAACGCATGAACTCTTTGATAACGTCTTCGGAGGAA  759

                                       *
733 GCCATTACTAGAGTTTCTCCTCTTTAATTCGCGATACTAGAGTCACACAGGAAAGTACT-  791
    |||||||||||||||||||||||||||||||||| | || ||  |   || ||| ||||
760 GCCATTACTAGAGTTTCTCCTCTTTAATTCGCGA-ATTAAAG--AGG-AG-AAA-TACTA  813

792 -AGATGCGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGNANTTAGA  850
     ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
814 GAGATGCGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAA-TTAGA  872

851 TGGT-ATGTTAATGGGCACAAATTTTCTGTCAGTGGANAAGGGTGAAAGGTGATGCAACA  909
    |||| |||||||||||||||||||||||||||||||| | ||||| ||||||||||||||
873 TGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGA-GGGTG-AAGGTGATGCAACA  930

910 TACGGAAAACTTACCCTTA  928
    |||||||||||||||||||
931 TACGGAAAACTTACCCTTA  949
```

FIG. 5 Sanger sequencing alignment with pSPPH21. Sequencing results of the vector isolated from colony sample 5 aligned with the pSPPH21 vector reference sequence analyzed using the NCBI nucleotide alignment tool. Top line (query) shows the sequence of the submitted sample. Bottom line (subject) shows the pSPPH21 reference sequence. Identity percentage shows the number of bases matched between the query and subject sequences. E-value displays probability of results occurring by chance. Vertical lines in between the two sequences indicate a match between bases. Horizontal lines indicate gaps in the sequence. The letter 'N' shows regions where bases could not be determined with accuracy during sequencing. Bolded letters in the sequence indicate the RFP gene sequence (5' → 3'). The NruI restriction site is bolded and denoted by (*).

Comparison to conventional SIGEX vectors. Current SIGEX vectors, namely p18GFP (Figure 6A), have been adapted from a single reporter architecture, with the use of a GFP gene downstream of an overhang producing restriction site (BamHI) for insertion of environmental DNA fragments (2, 3, 5, 9). This vector design is promising in the identification of substrate-induced promoters, though it is limited by the unidirectionality of the fluorescence reporter; to be detected, DNA fragments containing a promoter of interest must be inserted in an orientation that allows the expression of the downstream fluorescence genes by the cell transcriptional machinery (2, 10). The pSPPH21 plasmid, however, uses two oppositely-oriented fluorescence genes (GFP and RFP), which allow expression of fluorescence proteins regardless of insert directionality (Figure 6B). Furthermore, the SIGEX vector p18GFP uses "sticky-ended" restriction endonucleases, which require specific endonuclease recognition sites that may not flank naturally occurring gene fragments (3). To resolve this limitation, the pSPPH21 plasmid contains a blunt-ended NruI recognition site (Figure 6B), which allows for the testing of a larger selection of DNA fragments. In spite of these improvements, a caveat of the use of blunt-ended ligations may be the need for greater concentrations of inserts and poorer efficiency of ligation reactions when compared to sticky-ended cloning (11).

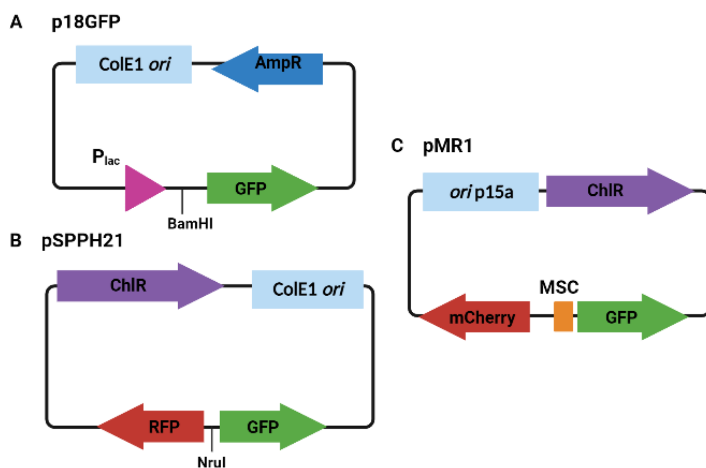

FIG. 6 Diagram of existing SIGEX vectors compared to pSPPH21. (A) The p18GFP SIGEX vector contains an ampicillin resistance marker, a high-copy number ori site, and only one fluorescent marker (GFP) controlled by a lac promoter and downstream of a BamHI sequence. (B) The duo-directional pMR1 vector consists of a chloramphenicol resistance gene, a low-copy number ori site, and mCherry and GFP genes in opposite orientations and flanking an MSC. (C) The duo-directional pSPPH21 vector contains a chloramphenicol resistance marker, a high-copy number ori site (same as p18GFP), and GFP and RFP genes in opposite orientations, flanking an NruI restriction site.

An additional difference between the pSPPH21 and p18GFP SIGEX vectors is the presence of an active *lac* promoter in p18GFP, which controls the expression of GFP when induced with isopropyl-β-D-thiogalactopyranoside (IPGT) (9). The *lac* promoter allows for the detection of self-ligating plasmids that do not contain DNA inserts, and hence reduces false positives in FACS analysis (3, 9). The pSPPH21 vector does not include this feature. Instead, the fluorescence reporters, GFP and RFP, are promoterless, and the lack of DNA inserts would not lead to the expression of any fluorescence proteins; self-ligating vectors can thus be excluded from analysis as no fluorescence would be detected.

Other duo-directional SIGEX vectors. Aligning with the objectives of our paper, a small proportion of available literature have discussed a novel vector with duo-directional reporters, which has been tested to address the challenges of SIGEX technology (10). Guazzaroni and Silva-Rocha synthesized a duo-directional reporter system named pMR1, with oppositely-oriented GFP and mCherry genes and a chloramphenicol resistance marker, which closely resembles the pSPPH21 plasmid (Figures 6B and 6C) (12). An important difference between the pMR1 and the pSPPH21 vectors is that the former contains a low copy number origin of replication; this may be due to the fact that pMR1 was not originally designed for SIGEX cloning experiments (12). The low-copy number of the vector may limit its application in SIGEX experiments, as this technology relies on fluorescence protein expression for detection of inducible promoters, and lower copy numbers will yield lower fluorescence (13). Lower copy numbers may also increase the chance of plasmid loss in colonies (14).

Additionally, the two reporter genes within pMR1 flank a multiple cloning site (MSC) for insertion of DNA fragments as opposed to a blunt-ended restriction site (Figure 6C) (10, 13). While successful creation of a genomic library was achieved using the pMR1 plasmid, a highlighted limitation of this SIGEX vector would be the insertion of a gene fragment in closer proximity to one of the fluorescence reporters, which may further introduce bias in the detection of substrate-induced promoters (10). The design of the pSPPH21 vector would theoretically reduce the chance of this error, as NruI is located in approximately equal distance from the two fluorescent markers (Figure 6B); nevertheless, the functionality of pSPPH21 requires testing in future studies to better understand the potential occurrence of these biases.

Conclusions A duo-directional reporter system, pSPPH21, was created as designed by iGEM and Avery Noonan, and its construction was confirmed using gel electrophoresis and Sanger sequencing. Further studies are needed to test the functionality of this novel vector in SIGEX experiments.

Future Directions Prior to using the pSPPH21 vector for SIGEX applications, a proof-of-concept experiment may be beneficial to test the vector and any significant biases. Using the NruI cloning site in pSPPH21, blunt-ended ligation of inducible promoters and the vector can be done. The functionality of pSPPH21 can then be confirmed by transforming the vector into a suitable host and examining the expression of the GFP and RFP genes after treatment with an inducing substrate. As the RBS sequences of the two reporter genes are different due to the reported mutation in the GFP Shine-Dalgarno sequence, we hypothesize that the inserted promoter would be less likely to form secondary structures through specific promoter-RBS combinations. Theoretically, the blunt-ended production of fragments would cause insertion into the vector without bias towards a particular orientation. To validate the construct, we further hypothesize that the blunt-ended insertion of the promoter sequence into the pSPPH21 vector will result in a pool of colonies with 50% GFP and 50% RFP expression, implying unbiased insertion orientation.

The novel pSPPH21 vector allows for screening of metagenomic libraries of catabolic genes, where gene expression can be induced bidirectionally due to presence of oppositely oriented RFP and GFP genes in the vector. A SIGEX duo-directional reporter system can also be used for making a screening platform for finding novel promoters, some of which have been reported to be responsive to biomarkers useful in cancer research (https://2021.igem.org/Team:British_Columbia/Design#SIGEX). This can be applied in screening inducible promoters in bacteria such as Salmonella, which are reportedly sensitive to specific biomarkers found in the tumor microenvironment (15). Other applications of this vector can be in environmental microbiology research, specifically in testing promoters which can be induced via carcinogenic and recalcitrant aromatic compounds (5). Researchers have previously used a single-reporter SIGEX vector to identify catabolic genes that target aromatic compounds in a soil sample contaminated with aromatic hydrocarbons (5), though there is no information regarding this research using duo-directional reporters. Incorporating pSPPH21 in environmental research can allow a broader analysis of catabolic genes by eliminating the limitation of insertion directionality. This can further aid in finding sustainable methods for reducing environmental contamination using unknown catabolic pathways in complex bacterial communities (5). Furthermore, this system can be useful in the discovery of novel biocatalysts, bioactives and antimicrobials (16). As the majority of foods and pharmaceuticals are produced industrially, pSPPH21 can aid in identifying biocatalytic genes that may be used to process compounds such as starch (16). Furthermore, these identified genes can be cloned and amplified in host bacteria, allowing sustainable production of food products (16).

ACKNOWLEDGEMENTS

We thank Mr. Avery Noonan for designing the RFP construct sequence and for providing us with the pSB1C3 vector and the reference pSPPH21 sequence. We also express gratitude to Dr. David Oliver for their knowledge, supervision, and continuous support throughout our experiments, and Ms. Jade Muileboom for their technical guidance. We thank the UBC

Department of Microbiology and Immunology for funding our project. We would also like to thank two anonymous reviewers for constructive feedback on this manuscript.

CONTRIBUTIONS

All participants participated equally in the project and in writing the manuscript. All four members invested an average of 12 hours per week at the UBC Microbiology and Immunology Laboratories. In completing the manuscript, Parsa Tabassi and Parsa Abrishamkar collaborated on the analysis of results. Helia Mansouri Dana and Sourena Oveisi further cooperated in completing the discussion. Helia Mansouri Dana further described the methods used to complete this project, with the assistance of Sourena Oveisi in some sections. All four members conducted background research, and Parsa Tabassi and Sourena Oveisi completed the Introduction section based on the team's research. Sourena Oveisi and Parsa Abrishamkar further completed the section on Future Directions. Furthermore, Parsa Abrishamkar and Helia Mansouri Dana collaborated on completing the supplemental data. Lastly, Parsa Tabassi completed the Abstract of the project, and the team chose an appropriate title for the manuscript.

REFERENCES

1. **Henne A, Daniel R, Schmitz RA, Gottschalk G.** 1999. Construction of Environmental DNA Libraries in Escherichia coli and Screening for the Presence of Genes Conferring Utilization of 4-Hydroxybutyrate. J Appl Environ Microbiol 65:3901–3907.
2. **Yun J, Ryu S.** 2005. Screening for novel enzymes from metagenome and SIGEX, as a way to improve it. Microb Cell Fact 4:8.
3. **Uchiyama T, Watanabe K.** 2008. Substrate-induced gene expression (SIGEX) screening of metagenome libraries. Nat Protoc 3:1202–1212.
4. **Leis B, Angelov A, Liebl W.** 2013. Chapter One - Screening and Expression of Genes from Metagenomes, p. 1–68. *In* Sariaslani, S, Gadd, GM (eds.), Advances in Applied Microbiology. Academic Press.
5. **Meier MJ, Paterson ES, Lambert IB.** 2015. Use of substrate-induced gene expression in metagenomic analysis of an aromatic hydrocarbon-contaminated soil. J Appl Environ Microbiol 82:897–909.
6. **Chang AY, Chau VWY, Landas JA, Pang Y.** 2017. Preparation of calcium competent *Escherichia coli* and heat-shock transformation. UJEMI 1:22–25.
7. **Chen H, Bjerknes M, Kumar R, Jay E.** 1994. Determination of the optimal aligned spacing between the Shine – Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. Nucleic Acids Res 22:4953–4957.
8. **Solar G del, Giraldo R, Ruiz-Echevarría MJ, Espinosa M, Díaz-Orejas R.** 1998. Replication and control of circular bacterial plasmids. Microbiol Mol Biol Rev 62: 434-464.
9. **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.
10. **Westmann CA, Alves L de F, Silva-Rocha R, Guazzaroni M-E.** 2018. Mining novel constitutive promoter elements in soil metagenomic libraries in *Escherichia coli*. Front Microbiol 9:1344.
11. **Motohashi K.** 2019. A novel series of high-efficiency vectors for TA cloning and blunt-end cloning of PCR products. Sci Rep 9:6417.
12. **Guazzaroni M-E, Silva-Rocha R.** 2014. Expanding the logic of bacterial promoters using engineered overlapping operators for global regulators. ACS Synth Biol 3:666–675.
13. **Friehs K.** 2004. Plasmid copy number and plasmid stability. Adv Biochem Eng Biotechnol 86:47–82.
14. **Chen S, Larsson M, Robinson RC, Chen SL.** 2017. Direct and convenient measurement of plasmid stability in lab and clinical isolates of *E. coli*. Sci Rep 7:4788.
15. **Leschner S, Westphal K, Dietrich N, Viegas N, Jablonska J, Lyszkiewicz M, Lienenklaus S, Falk W, Gekara N, Loessner H, Weiss S.** 2009. Tumor invasion of *Salmonella enterica* serovar Typhimurium is accompanied by strong hemorrhage promoted by TNF-alpha. PLoS One 4:e6692.
16. **Coughlan L, Cotter P, Hill C, Alvarez-Ordóñez A.** 2015. Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. Front Microbiol 6:672.