Volume 10:1-12

UJEMI+

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

Mapping Potential Transcription Start Sites of the *brka* Gene on pDO6935 in *Escherichia coli* Using the ARF-TSS Method

Gurpreet Sidhu, Sana Alayoubi, Alissa Gama, Jamila Huseynova

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

SUMMARY Autotransporters are membrane proteins expressed on the outer membrane in gram-negative bacteria. BrkA is an autotransporter and virulence factor of the whooping cough-causing bacterium *Bordetella pertussis*. Previous studies aimed at characterizing the autotransporter BrkA have used pDO6935-expressing *Escherichia coli* cells; however, the promoter driving *brkA* gene expression in pDO6935-expressing *E. coli* cells remains unknown. Our study aims to determine the transcription start site (TSS) of the *brkA* gene in pDO6935 in *E. co*li (DH5α). In order to map the TSS, we implemented the adaptor- and radioactivity-free identification of the transcription start site method (ARF-TSS) which involves reverse transcribing *brkA* mRNA into cDNA using a 5'-phosphorylated primer, ligating the cDNA, and PCR-amplifying it before sending it for sequencing to determine the TSS. We were able to more precisely map the region of the plasmid that the transcription start site driving the expression of the *brkA* gene in pDO6935 in *E. coli* which is located in a region more than 270 bp upstream of the *brkA* translation start site. Finding the transcription start site aids in locating the promoter and contributes to our understanding of gene expression and regulation. This knowledge can help improve future studies aimed at characterizing BrkA autotransporter using pDO6935-expressing *E. coli* cells.

INTRODUCTION

Bordetella pertussis, the bacterial pathogen that causes whooping cough, expresses a Type Va autotransporter protein known as BrkA. This protein serves as a virulence factor, contributes to resistance against serum, adhesion, and is secreted across both the inner and outer membranes of *B. pertussis* (1). Similar to other autotransporters, BrkA has an N-terminal signal sequence to guide the translocation across the inner bacterial membrane, a conserved C-terminal sequence for outer membrane translocation, and a passenger domain which is the secreted protein that is exported to the cell surface (1). The plasmid pRF1066 encodes both *brkA* and *brkB* genes. The plasmid pDO6935 encoding *brkA* was generated by cutting the 476-bp AatII fragment from the pRF1066 plasmid leading to the deletion of the 5' region of the *brkB* gene (1). The plasmid pDO6935 shows constitutive low expression in *Escherichia coli*, but the promoter driving the expression of *brkA* in *E. coli* is still unknown (1).

RNA polymerase holoenzyme initiates DNA transcription into RNA in bacteria by recognizing sequence elements in the DNA promoter region (2). Promoters are usually located directly upstream of the Transcription start sites (TSSs); therefore, TSSs are typically utilized to identify promoter elements in the genomes of bacteria (3, 4). Locating the TSS is important for characterizing the promoter of pDO6935 aiding in understanding gene expression and the regulation of BrkA. (5). Locating the TSS of prokaryotic genes is usually done through primer extension analysis or S1 nuclease protection mapping assays (5). These

Published Online: September 2024

Citation: Sidhu, Alayoubi, Gama, Huseynova. 2024. Mapping potential transcription start sites of the *brkA* gene on pDO6935 in *Escherichia coli* using the ARF-TSS method. UJEMI+ 10:1-12

Editor: Shruti Sandilya, University of British Columbia

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

methods involve the use of adaptors and radioactively labeling, and they are technically demanding and harmful to the environment (5). A novel method for determining the TSS, called adaptor-and radioactivity-free identification of TSSs (ARF-TSS), will be used (5). ARF-TSS involves generating cDNA from total RNA using a 5' phosphorylated primer, ligating, and then amplifying the cDNA; the PCR amplicons are cloned into a plasmid vector and sequenced (5). The TSS is the sequence upstream of the 5' end of the phosphorylated primer used to generate the cDNA (Figure 1) (5). Given that Wang et al. were able to use the ARF-TSS method to determine the promoter region of the *lasI* gene, we aim to identify the TSS of the *brkA* gene in pDO6935 and characterize the promoter region using the ARF-TSS method.

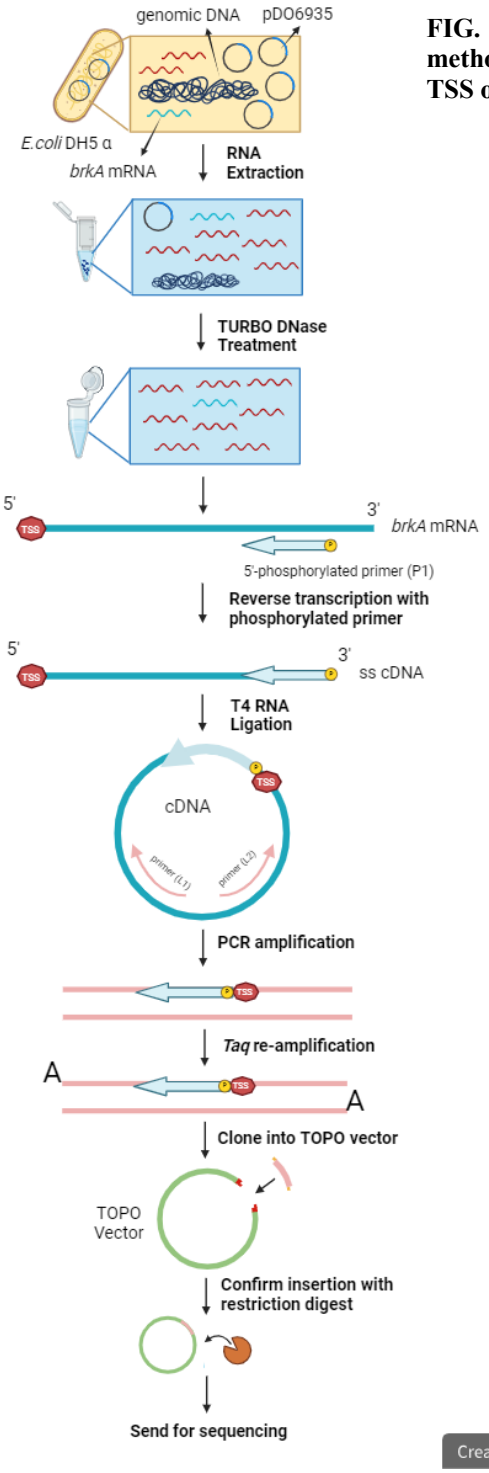

FIG. 1 Schematic diagram of the methodology used to identify the TSS of the *brkA* gene in pDO6935.

This study aims to determine the TSSs that play a crucial role in regulating the BrkA expression in pDO6935-containing *E. coli* DH5-α. This knowledge can help improve future studies using the pDO6935 for further characterization of the BrkA autotransporter. Understanding the TSS of BrkA is pivotal for characterizing its regulatory network and functional relevance in different environmental conditions. This knowledge will contribute to advancing our understanding of bacterial adaptation and aiding the development of targeted therapeutic vaccines against BrkA, the virulence factor of the causative agent of whooping cough.

METHODS AND MATERIALS

Bacterial Culture. 5 mL of 50 mg/mL ampicillin stock solution was prepared from Fischer™ Ampicillin, sodium salt (Cat. # BP1760-25) and distilled H2O and stored at 4°C. 500 mL of Luria-Bertani (LB) broth was made by adding 5g of bacto-tryptone, 2.5g of Bacto-yeast extract, and 5g of NaCl to 500 mL of distilled water. For LB plates containing ampicillin, an additional 7.5 grams of agar was added to the 500 mL LB broths, which were subsequently sent for autoclaving and cooled to around 48°C (1.5% agar plates). 200 µL of 50 mg/mL ampicillin was then added. Bacterial broths containing *E. coli* DH5-α cells expressing pDO6935 were grown overnight in 5mL of LB broth supplemented with 100 µg/mL of ampicillin under constant shaking at 200 rpm at 37°C.

RNA Extraction and Purification. Overnight cultures of *E. coli* DH5-α cells expressing pDO6935 were diluted 1:10 in LB-ampicillin broth for growth until the optical density (OD) of cultures reached an OD600 of 0.5. Optical density measurements of the liquid cultures were read on a Pharmacia Biotech Ultrospec 3000 UV-Vis Spectrophotometer. Total RNA was extracted from *E. coli* DH5-α expressing pDO6935 using the QIAGEN™ RNeasy Kit and the NEB™ Monarch Total RNA Miniprep kit according to the manufacturer protocols (8, 9, 10). To ensure efficient lysis of the bacterial cells and to obtain a high RNA yield, proteinase K and lysozyme solution (30mM Tris-HCl pH 8.0, 1mM EDTA, and 15mg/ml lysozyme) were added to the RNeasy samples; moreover, lysozyme solution was added to the Monarch samples. These additional solutions were added before the addition of the lysis buffer. RNA concentrations and purity were measured using a NanoDrop™ 2000 spectrophotometer. The RNA samples were then treated with Invitrogen's TURBO DNA-*free*™ Kit to remove any plasmid or genomic DNA contamination, following the manufacturer's protocol for routine DNase treatment (11).

Plasmid DNA Contamination Check. The presence of pDO6935 plasmid DNA will contaminate the RNA samples and interfere with downstream processes; therefore, it is crucial to ensure that the RNA samples are free of plasmid DNA contamination. Custom forward and reverse primers, C1 and C2 (Table 1), were designed using the SnapGene software (v. 7.1) by Dotmatics to bind at nucleotide positions 2197-2218 and 4486-4508 of pDO6935 respectively. This would amplify a 2312 bp region that would include our cDNA

TABLE. 1 List of primers used for experiments and their melting temperatures.

Primer Name	Sequence	T_m(°C)
C1	5'-ATTTCGCTGCCAGGCATATACG-3'	60
C2	5'-CGAACTGAGATACCTACAGCGTG-3'	59
P1	5'- phosphorylated- GAGATGTGGATGTCCTTG -3'	51
L1	5' -CAACACGCTCATTGCAGTC-3'	56
L2	5'-CTGCAAGGAAGACGGACATTG-3'	59

sequence, which consists of the region between the P1 (Table 1) sequence in the *brkA* gene to the TSS of *brkA*. The specific binding sites were also chosen due to the GC-rich nature of the plasmid and the limited availability of areas in the plasmid that weren't extremely GC-rich. PCR was performed using these primers and the Q5® High-Fidelity DNA Polymerase

(12). The following thermocycler conditions were used: initial denaturation at 98 ℃ for 30 seconds, 35 cycles of denaturation at 98℃ for 10 seconds and annealing/extension at 72℃ for 90 seconds, and final extension at 72℃ for 5 minutes (13). Following PCR amplification, the samples were held at 4℃, before the amplified products were run on a 1% agarose gel alongside the Thermo Scientific O'GeneRuler DNA Ladder.

cDNA Generation. A custom 5-phosphorylated primer, P1, for reverse transcription (RT) of the cDNA (Table 1) was designed using SnapGene software (v. 7.1). Since the location of the TSS is unknown, only a single primer was designed. To ensure the primer sequence did not conflict with or exclude the TSS, the primer was designed to bind to nucleotide positions 2773 to 2790, just downstream of the *brkA* translation start site to amplify the beginning of the *brkA* gene and the sequence upstream of the gene that should include the TSS. This specific binding site was also chosen due to the GC-rich nature of the plasmid and the limited availability of areas in the plasmid that were not GC-rich. RT was performed on total RNA extracted samples using SuperScript™ II Reverse Transcriptase, in accordance with the SuperScript™ II Reverse Transcriptase manual with minor modifications, to generate a cDNA fragment with the TSS at its 3'-end (14). 500 ng of total RNA was used per RT reaction, and 5% DMSO was added to the total reaction mix to address the GC-rich nature of the *brkA* mRNA. After RT, any remaining RNA was degraded by adding 2µL of 1N NaOH to each 21uL sample, incubating at 70°C, and neutralizing with 3.6µL 0.5M acetic acid (15).

cDNA Purification and Ligation. cDNA fragments were purified using the QIAGEN™ QIAquick PCR Purification Kit, following the manufacturer's protocol (16). Using T4 RNA Ligase 1 from ThermoFisher™, we ligated cDNA in accordance with the manufacturer's instructions (17). The amount of cDNA utilized was informed from the amounts employed by Kreisel et al. (18). The reaction mix consisted of 18 µL of the 1X T4 RNA Ligase 1 reaction buffer, 2.5 µL of BSA at a final concentration of 0.1 mg/ml and 1 µL of T4 RNA Ligase 1. The samples were then incubated at 37°C for 30 minutes. T4 RNA Ligase 1 was inactivated by heating at 70°C for 10 min (17).

PCR Amplification of Ligated cDNA. Custom forward and reverse primers, L1 and L2 (Table 1) were designed using SnapGene software (v. 7.1) to confirm the presence of the ligated cDNA fragment containing the primer P1 and the TSS sequences. PCR was performed using the Q5® High-Fidelity DNA Polymerase according to the manufacturer's guidelines (12). The following thermocycler conditions were used: initial denaturation at 98 ℃ for 30 seconds, 35 cycles of denaturation at 98℃ for 10 seconds and annealing/extension at 68℃ for 90 seconds, and final extension at 72℃ for 5 minutes (13). The amplified products were run on a 1% agarose gel alongside the Thermo Scientific O'GeneRuler DNA Ladder.

PCR Re-amplification with *Taq* Polymerase. Amplification of the ligated PCR products was performed using the L1 and L2 primers (Table 1) and *Taq* Polymerase from BioBasic following the manufacturer's protocol (19). This was done to add 3' polyA overhangs to the ligated PCR products for TOPO® cloning. The PCR used the following thermocycler conditions: initial denaturation at 94 ℃ for 5 minutes, 35 cycles of denaturation at 94℃ for 45 seconds, annealing at 51℃ for 30 seconds, extension at 72℃ for 1 minute, and final extension at 72℃ for 10 minutes (19, 20). The PCR products were run on a 1% agarose gel via gel electrophoresis alongside the Thermo Scientific O'GeneRuler DNA Ladder.

TOPO® TA Cloning and Isolation. The *Taq*-amplified amplicons were then cloned into a TOPO® vector following Invitrogen's TOPO® TA Cloning Kit for Sequencing User Guide (21). 2µL of fresh PCR product and 1µL of TOPO® vector were used for each cloning reaction, and cloned vectors were transformed into competent *E. coli* One Shot® TOP10 cells. Transformants were plated on ampicillin-supplemented (50 mg/mL) LB plates to select for Amp resistance marker on the TOPO® vector and grown overnight at 37°C. Single colonies from these plates were inoculated into LB broth supplemented with ampicillin (50 mg/mL) and cultured overnight at 37°C, shaken at 200 rpm. The cloned TOPO® vector was isolated

using the EZ-10 Spin Column Plasmid DNA Miniprep Kit from Bio Basic, following the manufacturer's protocol. (22)

Restriction Digest. To validate the integration of the PCR insert into the TOPO® vector, we conducted a restriction digest using EcoRI-HF enzyme. The digestion followed the manufacturer's guidelines, incubating 1 µg of DNA, 1x rCutsmart buffer, and EcoRI-HF at 37°C for 15 minutes. Samples displaying successful TOPO® vector cloning displayed two bands on the gel. Samples that underwent restriction enzyme digestion were diluted to a concentration of 30 ng/µl. Whole plasmid nanopore sequencing was performed by Plasmidsaurus (23).

RESULTS

Total RNA was extracted from pDO6935-expressing *E. coli* DH5α. To determine the yield and purity of the RNA samples, the concentration, $A_{260/280}$ values, and $A_{260/230}$ values were measured using the Nanodrop. The yield, $A_{260/280}$, and $A_{260/230}$ values are summarized in Table 2. The Monarch Total RNA Miniprep samples had higher yield overall. Considering

TABLE. 2 Isolated RNA samples yield and purity measured using the Nanodrop.

Kit used for extraction	RNA sample name	Yield (ng/µL)	$A_{260/280}$	$A_{260/230}$
RNeasy Kit from QIAGEN	**R1**	136.6	2.14	1.61
	R2	197.8	2.16	1.05
	R3	174.1	2.15	2.10
	R4	129.3	2.16	1.45
Monarch Total RNA Miniprep kit from NEB	**R5**	85.3	2.22	0.17
	R6	99.6	2.15	2.30
	R7	65.3	2.11	1.97

that mRNA transcripts constitute less than 5% of total bacterial RNA found in the cell and the *brkA* gene exhibits low expression levels on pDO6935, an increased overall yield would probably result in a greater quantity of *brkA* mRNA produced (24, 1). Typical criteria for optimal purity of RNA are indicated by a 260/230 ratio of 2 or slightly above 2. However, contaminants with significant absorption at 230 nm, such as guanidine thiocyanate found in lysis buffer during RNA extraction, have been shown to not affect cDNA synthesis at $A_{260/230}$ values as low as 0.5 (25). Considering that low RNA $A_{260/230}$ values do not affect downstream applications, we proceeded with mRNA reverse transcription.

Extracted RNA samples are free of pDO6935 DNA contamination. The presence of pDO6935 DNA could result in false positive results downstream as our cDNA amplification primers can bind to pDO6935. Therefore, we performed a DNase treatment for the depletion of genomic DNA contamination. Using forward primer C1 and reverse primer C2 that bind at nucleotide positions 2197-2218 and 4486-4508 of pDO6935, we used PCR to amplify a 2312 bp region that would include our cDNA sequence. We performed the DNA check on all seven RNA samples along with a positive control containing pDO6935 and a negative no template control. Genomic DNA was not found in our RNA samples, as evidenced by the absence of bands for all seven samples and the negative control, and as affirmed by a robust band of 2300 bp for the positive control (Figure 2). Therefore, no amplification of the genomic DNA occurred and the RNA samples are pure of pDO6935 DNA.

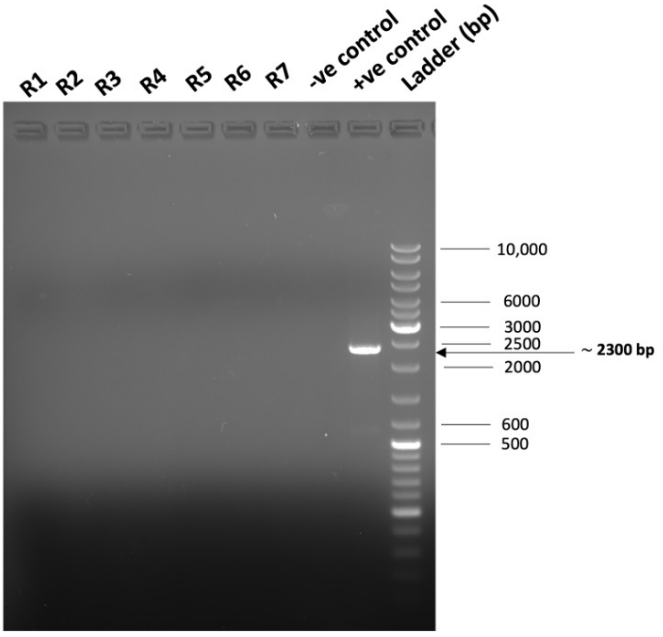

FIG. 2 The RNA samples extracted are free of pDO6935 contamination. PCR amplification of the part of the pDO6935 corresponding to the cDNA fragment was conducted on TURBO DNase-treated RNA samples. Running the PCR products on a 1% agarose gel shows no bands indicating that the samples are free of pDO6935 contamination.

PCR amplification confirms the presence of ligated cDNA. In order to generate cDNA, we reverse transcribed the *brkA* mRNA from our total RNA samples using the 5'-phosphorylated primer, P1, binding pDO6935 at the nucleotide positions 2773 to 2790. We then purified the resulting cDNA products to eliminate potential interfering compounds, before ligating them with T4 RNA ligase 1 to covalently connect the 3′-end (TSS) and the 5′-end (P1) of the cDNA. Un-ligated RT samples were used as a negative control for contamination by pDO6935. Subsequently, we PCR-amplified the ligated purified cDNA fragments, as well as the non-ligated purified -RT controls, with forward primer L1 and reverse primer L2 to generate linear, double-stranded PCR products. All reverse-transcribed RNA samples, except R4, present a band at around 800 bp (Figure 3). Furthermore, the -RT

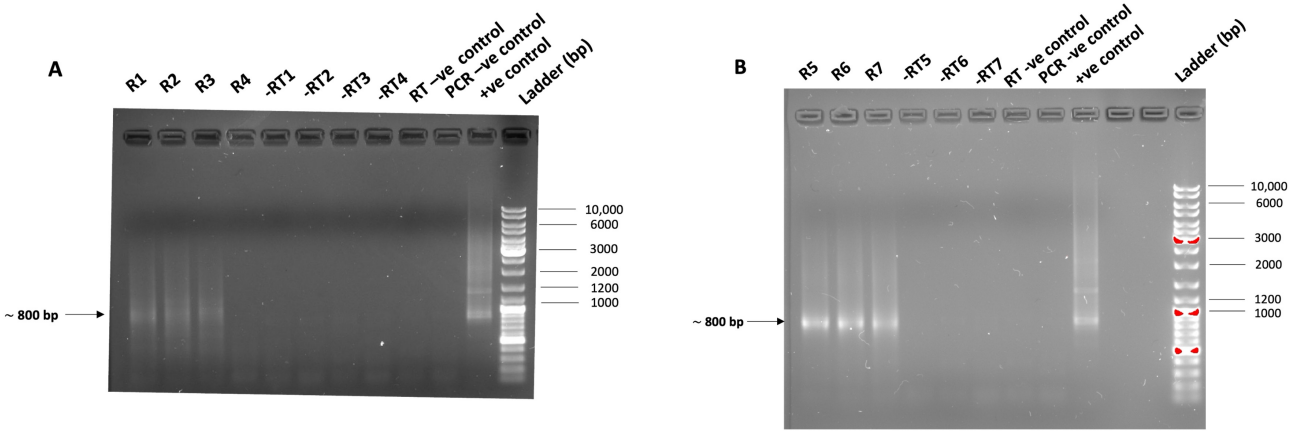

FIG. 3 Confirming the generation of ligated cDNA products through PCR and gel electrophoresis. (A) PCR amplification was conducted on RNA extracts from *E. coli* DH5α cells that were reverse transcribed and ligated. The RNA extracts were obtained using A) the RNeasy Kit from QIAGEN (R1-R4) and **(B)** Monarch Total RNA Miniprep kit from NEB (R5-R7). The PCR products run on a 1% agarose gel were shown to have the expected band sizes of 800-1000 bp. –RT (1-7) indicate negative controls for plasmid contamination. RT and PCR -ve controls, consisting of nuclease-free water only, indicate negative controls for reverse transcription and PCR amplification, respectively. +ve control is a positive control for PCR amplification, consisting of pDO6935. O'GeneRuler 1 kb DNA Ladder (Thermofisher) was used for reference (26)

controls corresponding to each of these samples, all show no bands, confirming there was no pDO6935 contamination during any of the experimental steps up to PCR amplification (Figure 3). The RT negative control, a no-template control, with reverse transcriptase, confirms no contamination during RT, while the PCR negative controls confirm no

contamination during PCR amplification. The PCR positive controls show a smear and bands at around 800 and 1200 bp, contrary to our expectations of a 6766 bp fragment, suggesting non-specific binding of L1 and L2 (Figure 3). However, the bands at around 800 bp are slightly longer than the bands seen in the samples (Figure 3). The positive controls confirm no technical issues with the PCR as we see amplification. Since we do not know the exact position of the TSS, we did not expect any specific-sized fragment. However, given that the amplification of the reverse transcribed RNA samples shows distinct bands that are unlike the smear seen in the positive control with pDO6935, and that all our -RT controls present no bands, our results indicate generation of cDNA from *brkA* mRNA and not simply a false positive due to pDO6935 contamination.

Re-amplification of ligated PCR products with *Taq* for samples R5-R7 showcases successful insertion of 3'-A overhangs. The ligated PCR products were reamplified with Taq polymerase to add a 3' polyA overhang to be inserted into a TOPO® vector. pDO6935 was used as a positive control and a sample without template was included as a negative control. The positive control showed a faint smear indicating non-specific products which were not present in the negative control (Figure S1). The bands at the bottom of the gel of all samples are likely primer dimers. For samples R1-R3, distinct bands were not observed at the expected 800 bp mark, however, smearing was observed indicating the presence of many non-specific products. Despite smearing, the distinct bands at the 500bp mark, corresponding to R1-R3, could be distinguished likely due to the generation of non-specific products (Figure S1). For samples R5-R7, while smearing was again present, distinct bands could be made out at the 800bp mark, indicating the ligated PCR products have successfully been re-amplified by *Taq* polymerase and thus contain the 3' polyA overhangs.

Restriction digest with EcoRI-HF confirms insertion of *Taq*-amplified fragments into TOPO® vector. To determine whether there was an insertion of *Taq*-amplified cDNA fragments into the TOPO® vector, we performed a restriction digest on the vectors with the cDNA insert isolated from *E. coli* One Shot® TOP10 cells. The TOPO® vectors with cDNA inserts were digested with EcoRI-HF restriction enzyme which would cut the EcoRI restriction sites on either end of the insertion site and produce a 3900 bp TOPO® vector, as well as the inserted fragment. R5-R7 samples correspond to the cDNA samples reverse transcribed from RNA, while the subsequent numbers denote the different colonies on the same plate from which the vectors with the inserts were isolated from. The identical band size observed at approximately 4500 bp strongly suggests that it is the TOPO® vector due to its close proximity to the anticipated size of the vector (Figure 4). Various fragments, distinguished by different band sizes, were obtained for the samples. R5.3 and R6.1 samples

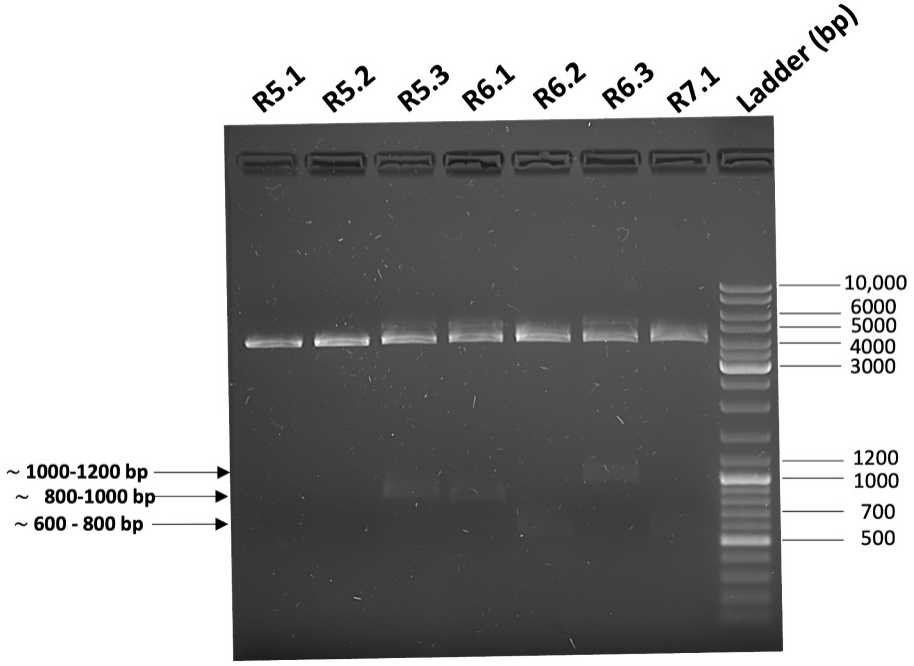

FIG. 4 Restriction digest verifies the success of TOPO® cloning of the desired ligated cDNA PCR product. EcoRI-HF restriction digest was performed and visualized on a 1% agarose gel, confirming the incorporation of the PCR product in the TOPO® vector. R5-R7 samples correspond to the RNA samples extracted using the Monarch Total RNA Miniprep kit. The subsequent numbers in the sample names represent distinct colonies chosen from the same plates. O'GeneRuler 1 kb DNA Ladder (Thermofisher) was used for reference (26).

presented bands at around 800-1000 bp, R6.2 and 7.1 samples manifested bands at around 600-800 bp, and R6.3 sample showed a band at 1000-1200 bp. R5.1 and R5.2 showed no bands. The positive control, which consists of pDO6935 DNA digested with EcoRI-HF, showed bands at around 5000 bp and 1500 bp which is the size of fragments we expected given where the EcoRI cut sites are on pDO6935 (Figure S2). There was also a band at 6000 bp, which was likely un-digested plasmid, and a band at greater than 10 000 bp, which is unknown. Together, these results indicate that samples R5.3, R6.1, R6.2, R6.3 and R7.1 showed insertion of the *Taq*-amplified ligated cDNA fragment; therefore, only these samples were prepared and sent for whole plasmid sequencing.

Nanopore sequencing of TOPO® vectors with *Taq*-amplified inserts shows multiple potential TSSs. In order to determine the TSS, we sent five *Taq* polymerase-amplified fragments inserted into TOPO® vectors for nanopore sequencing. To map the TSS, we located the sequence of the P1 primer and looked at the nucleotide sequence upstream of the 5'-end of the primer to map it back to pDO6935. However, the first part of the sequence, where we expect the TSS to be, did not map to sequences upstream of the *brkA* gene. Instead, the primer sequence mapped within the *brkA* gene sequence downstream of primer P1, before switching to a sequence upstream of the *brkA* gene (Figure 5). The length of the *brkA* gene sequence downstream of primer P1 varied in each fragment as did the length of the sequence upstream of the *brkA* gene, leading to multiple potential TSSs (Figure 5). Fragments R5.3, R6.2, and R6.3 also had overlapping regions of sequence, thus leading to two possible TSSs depending on whether the overlapping region belonged to the *brkA* gene sequence or the region upstream of the *brkA* gene (Figure 5).

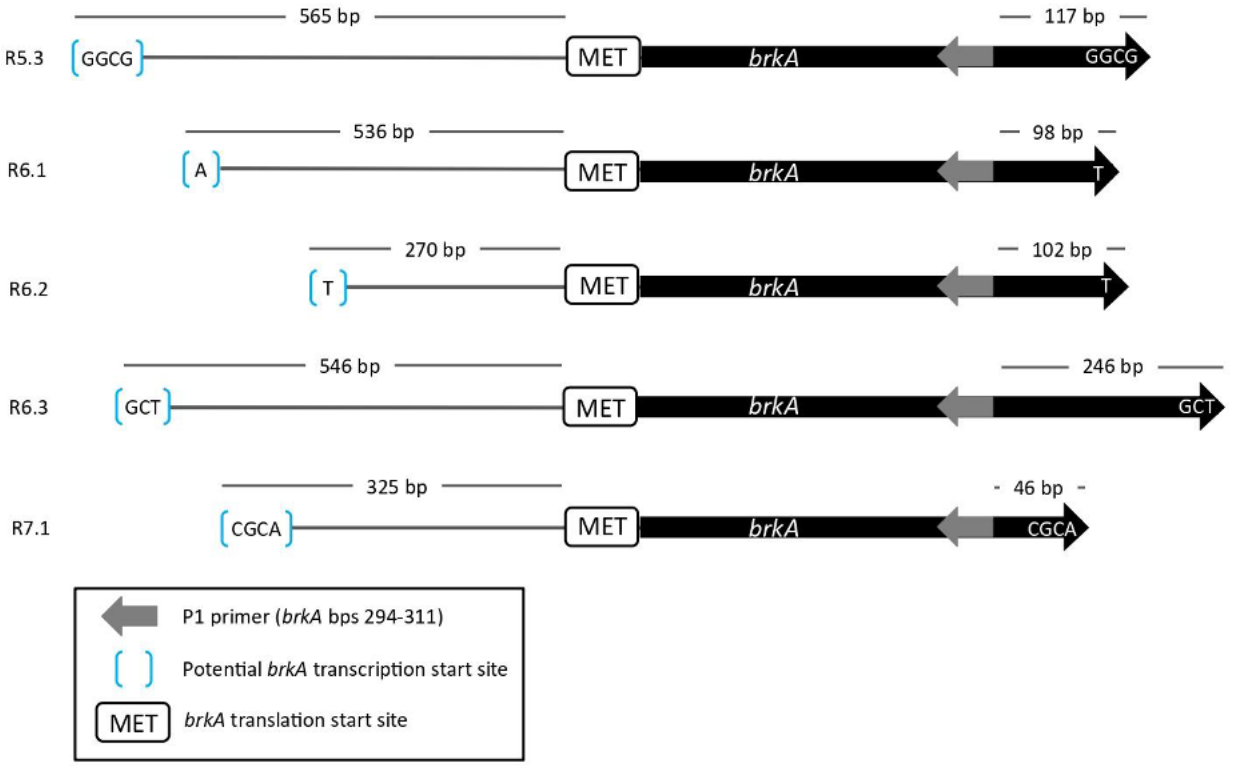

FIG. 5 Nanopore sequencing results of PCR-amplified cDNA fragments. The results are aligned to show the difference in potential TSSs identified, as well as extension of the cDNA downstream into the *brkA* gene. The bp correspond to the template strand of *brkA* and run 3' to 5'. Size of each fragment is not to scale, only used to depict differences.

DISCUSSION

Our study aimed to determine the TSS of the *brkA* gene on pDO6935 to characterize the promoter driving the expression. Using the ARF-TSS method (5), we reverse-transcribed a segment of purified *brkA* mRNA (confirmed through PCR plasmid DNA contamination check) into cDNA with a 5'-phosphorylated primer. We then ligated the cDNA using T4 RNA ligase 1 and amplified the circularized product followed by the addition of 3' polyA overhangs. These fragments were then cloned into a TOPO® vector and confirmed by restriction digest yielding fragments of varying sizes, before sending five clones for nanopore sequencing. The sequencing results revealed that each clone showed a different potential TSS as well as the extension of the cDNA fragment in the 5'-direction of the 5'-phosphorylated primer. Given that the shortest cDNA fragment shows the TSS to be 270 bp upstream of the *brkA* gene, our results confirm that the TSS of *brkA* is not located anywhere within the 270 bp upstream of the *brkA* translation start site. The sequencing results of the PCR-amplified cDNA fragments inserted into TOPO® vectors revealed that each clone showed a different potential TSS (or two possible TSSs due to overlapping sequences) (Figure 5). Interestingly, three of these potential TSSs (fragments R5.3, R6.2, and R6.3) are within 29 bp of each other and further upstream of the *brkA* gene than the other two (fragments R6.1 and R7.1) which are closer to the *brkA* gene and within 55 bp of each other.

Analyzing the 565 bp region upstream of the *brkA* gene which contains all five of the potential TSSs reveals that the region has 62% GC content. Regions of high GC content have been shown to have more secondary structure formation than non-GC rich regions (27). The formation of these secondary structures can cause reverse transcriptases to fall off the RNA transcript and result in early termination of cDNA synthesis (28). Considering that the fragments are not terminating randomly and instead are terminating in two distinct regions upstream of the *brkA* gene, this could suggest the formation of stable secondary structures. Specifically, these secondary structures would be located upstream of 565 bp from the *brkA* translation start site (resulting in the termination of fragments R5.3, R6.2, and R6.3) and upstream of 325 bp from the *brkA* translation start site (resulting in termination of fragments R6.1 and R7.1).

Furthermore, higher temperatures can reduce the formation of secondary structures in RNA (29). Our reverse transcription was run with SuperScript™ II Reverse Transcriptase, which runs at a temperature of 42°C. In contrast, Wang et al. used SuperScript™ III Reverse Transcriptase, which runs at 50°C to determine the TSS of the *lasI* gene, and Round et al. used SuperScript™ IV Reverse Transcriptase which can optimally run at 65°C, however, they ran it at 50°C to determine the TSS of RS36130, a transcriptional regulator (5, 30). Given the high GC content of the region that the TSSs could be located in, we speculate that there may have been secondary structure formation and 42°C was not a high enough temperature to melt these secondary structures. These secondary structures resulted in the early termination of our cDNA amplification before termination at the 5'-end of the *brkA* mRNA transcript.

Another potential explanation for the multiple TSSs is degradation of the *brkA* mRNA from the 5'-end. Compared to eukaryotic mRNA, bacterial mRNA is much more unstable and can have a half-life of a few minutes to a few hours (31). Furthermore, mRNA degradation often occurs in the 5' to 3' direction. Given the 3'-stem loops stabilizes the 3'-end of mRNA and bacterial mRNAs do not have a 5'-methylated guanosine cap, bacterial mRNA is less protected from degradation than eukaryotic mRNA (31, 32). Although mRNA degradation would not likely result in similar-sized cDNA fragments, given there could be varying degrees of mRNA degradation, another mechanism could explain these similar-sized fragments. Post-transcriptional processing such as alkylation has been shown to take place during reverse transcription and could be the potential mechanism behind the generation of similarly-sized cDNA fragments (33). Post-transcriptional modifications can alter nucleotide sequence or charge, influencing interactions with RNA-binding molecules. For instance, reverse transcriptase polymerases may introduce errors or prematurely terminate transcription when encountering modifications that disrupt base-pairing, thus producing shortened cDNA fragments of similar sizes.

In addition, sequencing analysis of our results showed cDNA synthesis in the 5'-direction. The extension of cDNA synthesis was observed in the 5'-direction, downstream into the *brkA* gene (Figure 5). This is an unusual result given a single, 5'-phosphorylated primer during

reverse transcription and reverse transcription was used and should only have amplified in the 5' to 3' direction, towards the 5'-untranslated (UTR) region of the *brkA* mRNA. One possibility for these results could be non-specific binding of the 5'-phosphorylated primer to the *brkA* mRNA. The annealing temperature of the 5'-phosphorylated primer is 51°C according to the SnapGene software it was designed with. Therefore, it is possible that non-specific binding occurred when the reaction was run at 42°C, leading to the extension of the cDNA fragments in the *brkA* gene to variable degrees. However, if this was the case, we should have seen the primer sequence on the 5'-end of the sequenced fragments, which we did not.

Limitations Bacterial mRNA can be very challenging to work with especially due to its instability. Thus, it is very possible that despite our best efforts to work quickly and efficiently to mitigate exposure to RNases, RNA degradation still occurred. Additionally, Nanodrop readings can be unreliable, particularly with low RNA concentrations, if the concentration of the RNA samples is close to the detection limit of the machine (34). It also is not a reliable assessment of RNA integrity; thus, it is entirely possible that the reason for the variability in the resulting cDNA fragments was due to degradation at the 5'end.

Another limitation was the GC-rich nature of the plasmid which made it difficult to design primers and amplify regions of the plasmid necessary for the experiment. Primer sequences were restricted to specific locations in the *brkA* gene where the GC-rich content remained around 50%. This design limited our ability to design highly specific primers to the BrkA region which are crucial to the successful implementation of the ARF-TSS method. Furthermore, GC-rich nature of BrkA also result in secondary structure such as folding in the RNA which may have interfered in the reverse transcription reaction, resulting in the early termination.

Conclusions As demonstrated earlier and corroborated by our results, the ARF-TSS method can be employed as a preferable alternative to conventional approaches for identifying transcription start sites (TSSs) that were both tedious and less cost-efficient (5, 36, 37). Our investigation showcased effective outcomes in various procedures, encompassing, among others, the extraction of RNA from *E. coli* DH5-α cells expressing pDO6935, the reverse transcription of mRNA sequences of interest, and the application of a phosphorylated primer and ligation reaction to identify a potential transcription start site (TSS) for our target gene, *brkA*. Although our results cannot conclusively confirm the TSS of *brkA* on pDO6935, we can confirm that the TSS is not located within 565 bp upstream of the *brkA* translation start site, since our longest cDNA fragment, R5.3, shows that the TSS is at least 565 bp upstream of *brkA.*

Future Directions Since RNA integrity was not controlled for in our experiment and could contribute to inconclusive results, integrity could be measured after RNA extraction. For example, assays such as a denaturing RNA gel can determine RNA integrity and should be added to the experiment (35). Determining the RNA integrity could help elucidate if RNA degradation is a crucial component affecting our results and help us further optimize the protocols for extracting the RNA. High-quality RNA is critical to the ARF-TSS method as it is needed to accurately capture the TSS at its 5'end.

As outlined in the discussion, future investigations could investigate acquiring a reverse transcriptase such as SuperScript™ IV Reverse Transcriptase that can run the reverse transcription reaction at higher temperatures (50°C) in order to reduce the formation of secondary structures in the RNA. Additionally, secondary structures could be playing a role in the observed results due to the GC-rich nature of pDO6935. Thus, secondary structures in the RNA should be investigated prior to reverse transcription to identify potential problematic regions that may be interfering with the reverse transcription reaction. Secondary structure can be predicted using software such as RNAfold that can analyze and calculate secondary structure formation given the RNA sequence. If secondary structures are present, it may be necessary to adjust other parameters for amplification including using reverse transcription protocols refined for high GC content. This reaction is crucial to generating cDNA which

accurately captures the TSS, and the promoter region, thus further optimization is crucial for the proper implementation of the modified ARF-TSS method.

Finally, since there is variability in the TSS site, a large screen of other TOPO® vector transformants could help identify other colonies and greater consensus of the TSS. It is possible that one of these colonies contains a PCR product that maps exactly to the TSS, thus the TOPO vector should be isolated from numerous colonies and sent for nanopore sequencing to screen for the TSS.

ACKNOWLEDGEMENTS

We would like to thank Dr. David Oliver for his guidance, support, and continuous optimism for the duration of our project. We would also like to thank Jade Muileboom for her endless patience and help in the MICB 471 lab, and Charlotte Clayton for her encouragement and helpful suggestions for troubleshooting our project. Additionally, we would like to acknowledge the Department of Microbiology and Immunology at the University of British Columbia for providing the resources and funding necessary to carry out our research project. We would also like to thank the anonymous reviewer for constructive feedback on this manuscript.

CONTRIBUTIONS

All authors contributed to experimental work done in the MICB 471 laboratory. SA proposed the project idea, contributed to writing the abstract, the introduction, creating the figures, writing the methods, writing the results section, and creating the references. AG contributed to writing the methods, the results, the limitations, future directions, and creating the references. JH contributed to creating the figures, writing the introduction, methods, results, discussion and the conclusion. GS also created the figures and wrote the results, discussion, conclusion, and generated the references. AG, GS, JH, and SA all contributed to the final editing of the manuscript.

REFERENCES

1. **Oliver DC, Huang G, and Fernandez RC**. 2003. Identification of secretion determinants of the *Bordetella pertussis* Brka Autotransporter. *J Bacteriol* **185**:489–495.
2. **Cervantes-Rivera R, Puhar A**. 2020**.** Whole-genome Identification of Transcriptional Start Sites by Differential RNA-seq in Bacteria. *Bio Protoc* **10**:e3757.
3. **Matteau D, Rodrigue S**. 2015. Precise Identification of Genome-Wide Transcription Start Sites in Bacteria by 5'-Rapid Amplification of cDNA Ends (5'-RACE). *Methods Mol Biol* **1334**:143-159.
4. **Chevez-Guardado R, Peña-Castillo L**. 2021. Promotech: a general tool for bacterial promoter recognition. *Genome Biology* **22**:318-318.
5. **Wang C, Lee J, Deng Y, Tao F, Zhang L-H.** 2012. ARF-TSS: An alternative method for identification of transcription start site in bacteria. *BioTechniques* **52**:1–3.
6. **Jain S, Van Ulsen P, Benz I, Schmidt, MA, Fernandez, R, Tommassen J, Goldberg, MB.** 2006. Polar localization of the autotransporter family of large bacterial virulence proteins. *J. Bacteriol* **188**:4841-4850.
7. **Zhao L, Nguyen NT, Fernandez RC, Murphy ME**. 2009. Crystallographic characterization of the passenger domain of the Bordetella autotransporter BrkA. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **65**:608-611.
8. **QIAGEN**. 2021. RNeasy mini kit, part 1. https://www.qiagen.com/us/resources/
9. resourcedetail?id=0e32fbb1-c307-4603-ac81-a5e98490ed23&lang=en
10. **QIAGEN**. 2016. RNeasy mini kit, part 2 (EN). https://www.qiagen.com/us/resources/
11. resourcedetail?id=f9b2e5ef-9456-431a-85ed-2a2b9fbd503d&lang=en
12. **New England Biolabs.** 2017. Quick protocol for total RNA extraction (NEB #T2010). https://www.neb.com/en/protocols/2017/11/28/quick-protocol-for-monarch-total-rna-miniprep-kit-neb-t2010.
13. **Ambion.** 2012. TURBO DNA-free™ Kit TURBO™ DNase Treatment and Removal Reagents User Guide. https://tools.thermofisher.com/content/sfs/manuals/cms_055740.
14. pdf
15. **New England Biolabs**. 2013. PCR using Q5® high-fidelity DNA polymerase (M0491). https://www.neb.com/en/protocols/2013/12/13/pcr-using-q5-high-fidelity-dna-polymerase-m0491
16. **Assal N, Lin M.** 2021. PCR procedures to amplify GC-rich DNA sequences of mycobacterium bovis. *J. Microbiol. Methods* **181**:106121.
17. **ThermoFisher Scientific.** 2010. SuperScript™ II Reverse Transcriptase. https://tools.thermofisher.com/content/sfs/manuals/superscriptII_pps.pdf

18. **Avican K, Aldahdooh J, Togninalli M, Mahmud AK, Tang J, Borgwardt KM, et al.** 2021. RNA atlas of human bacterial pathogens uncovers stress dynamics linked to infection. *Nat. Commun.* **12**:3282-3282.
19. **QIAGEN**. 2018. QIAquick PCR Purification Kit and QIAquick PCR & Gel Cleanup Kit Quick-Start Protocol - (EN). https://www.qiagen.com/us/products/discover y-and-translational-research/dna-rna-purification/dna-purification/dna-clean-up/qiaquick-pcr-purification-kit
20. **ThermoFisher Scientific.** 2012. T4 Ligase Product Information. https://www.thermo
21. fisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2FMAN0011993_T4_RNA_Ligase_UG.pdf
22. **Kreisel K, Engqvist MKM, Clausen AR**. 2017. Simultaneous mapping and quantitation of Ribonucleotides in human mitochondrial DNA. *JoVE* **129**:e56551
23. **Bio Basic**. Taq DNA Polymerase Product Information.https://www.biobasic.com/us
24. /amfilerating/file/download/file_id/25311/
25. **Bodykevich G, Alyssa M, Li S, Tom K**. 2022. Cloning *chiC* from insecticidal *Pseudomonas aeruginosa* PAO1. *UJEMI* **27**:1-11.
26. **Invitrogen.** 2014. TOPO® TA Cloning® Kit for Sequencing. https://tools.thermofischer .com/content/sfs/manuals/topotaseq_man.pdf?fbclid=IwAR1T3wxqMJWfloAEFJdmRUCQSh-_uElRe-F3A5OhxDy2CSUZbJS1laq6_ho
27. **Bio Basic.** 2019. EZ-10 SPIN COLUMN HANDBOOK. https://www.biobasic.com/amfi lerating/file/download/file_id/25472/
28. **Plasmidsaurus**. 2023. Whole Plasmid Sequencing. https://www.plasmidsaurus.com
29. /plasmid_learn_more/
30. **Wangsanuwat C, Heom KA, Liu E, O'Malley MA, Dey SS**. 2020. Efficient and cost-effective bacterial mRNA sequencing from low input samples through ribosomal RNA depletion. *BMC genomics* **21**:717-717
31. **QIAGEN**. 2010. What are the effects of low A260/A230 ratios in RNA preparations on downstream applications? https://www.qiagen.com/us/resources/faq?id=c59936fb-4fle-
32. -4191-9c16-ff083cb24574&lang=en
33. **ThermoFisher Scientific.** 2016**.** O'GeneRuler 1 kb DNA Ladder, ready-to-use. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0013033_OGeneRuler_1kb_DNALadder_RTU_UG.pdf
34. **Price A**, **Garhyan J**, **Gibas C**. 2017. The impact of RNA secondary structure on read start locations on the Illumina sequencing platform. *PLOS ONE* **12**.
35. **Zhang Y-J**, **Pan H-Y**, **Gao S-J**. 2001. Reverse transcription slippage over the mRNA secondary structure of the LIP1 gene. *BioTechniques* **31**:1286–1294.
36. **Thomas SE**, **Balcerowicz M**, **Chung BY-W**. 2022. RNA structure mediated thermoregulation: What can we learn from plants? *Front. Plant Sci* **13**.
37. **Round JW**, **Roccor R**, **Eltis LD**. 2019. A biocatalyst for sustainable wax ester production: Re-wiring lipid accumulation in *rhodococcus* to yield high-value oleochemicals. *Green Chemistry* **21**:6468–6482.
38. **Rauhut R**, **Klug G**. 1999. mRNA degradation in bacteria. *FEMS Microbiology Reviews* **23**:353–370.
39. **Galloway A, Cowling VH**. 2019. mRNA cap regulation in mammalian cell function and fate. *Biochimica et biophysica acta. Gene regulatory mechanisms* **1862**:270-279.
40. **Dahlmann, Heidi A**. 2017. Post-Transcriptional Modification Tracked at Single-Base Resolution. *American Chemical Society*. 2919–2921
41. **ThermoFisher Scientific**. 2012. Sample Reproducibility. https://assets.thermofisher.com/TFS-Assets/CAD/Warranties/T044-NanoDrop%20Spectrophotometers-Sample-Reproducibility-EN.pdf
42. **Chacko, Sapna.** 2005. Validating RNA Quantity and Quality: Analysis of RNA Yield, Integrity, and Purity. *BioPharm International*. **18**:10
43. **Kim, S.K., Lee, J.B. and Yoon, J.W**. (2022). Characterization of transcriptional activities at a divergent promoter of the type VI secretion system in enterohemorrhagic *Escherichia coli* O157:H7. *Journal of Microbiology*, **60**: 928–934.
44. **Round, J.W., Roccor, R. and Eltis, L.D.** (2019). A biocatalyst for sustainable wax ester production: Re-wiring lipid accumulation in *Rhodococcus*to yield high-value oleochemicals. *Green Chemistry*. **21**:6468–6482.