JEMI methods

Vol. 1:7-15

# Antisense RNA Insert Design for Plasmid Construction to Knockdown Target Gene Expression

**Ji, Tom, Lu, Aneka, Wu, Kaylee**
*Department of Microbiology and Immunology, University of British Columbia*

**Regulatory RNA molecules are common tools used in bacterial gene regulation. This paper focuses on the steps in designing an antisense RNA component for insertion into a plasmid in order to silence a gene in the plasmid host cell. By hybridizing to the ribosomal binding site of the target gene mRNA transcript, the antisense RNA transcript from the plasmid is able to inhibit gene translation and lead to eventual mRNA degradation. By regulating or silencing the gene of interest, further experiments could be used to confirm or reject proposed roles for undefined genes. In this paper, *wecD* silencing using pHN678 plasmid in *Escherichia Coli* will be used as an example. The design of the antisense RNA component involves three main steps: identification of the gene of interest, selection of the antisense target sequence of the gene, and modification of the target sequence to generate the antisense sequence. In addition, a sense RNA insert could be used as an orientation control in the experiment. The vector plasmid without any sequence insertion should also be included as a negative control. As a result, a successfully designed antisense RNA insert should allow knockdown of the target gene through hybridization to the mRNA transcript and therefore inhibiting gene translation.**

## INTRODUCTION

Bacteria possess many diverse means of gene regulation using RNA molecules. More specifically, RNA molecules such as antisense RNA (asRNA) can hybridize to distinct regions of the target mRNA for gene regulation. For example, asRNA targeting near the 5'UTR of the mRNA can modulate the gene translation and mRNA stability (1). By sharing extensive sequence complementarity with this region, asRNA can inhibit ribosome binding and gene translation, and it often leads to mRNA degradation (1).

The most well-studied asRNAs reside on plasmids which makes it a practical tool for studying the function of a gene of interest (1). By regulating or silencing the gene of interest, further experiments could be used to confirm or reject hypothesized roles of the gene in the bacteria. For example, growth curves could be carried out to determine if the gene has an essential role in the survival of the bacteria.

As an example for this method paper, asRNA residing in pHN678 plasmid and targeting *wecD*, in *E. coli*, is used. The asRNA insert is specifically designed to target regions containing the ribosomal binding site (RBS) and the start site of the *wecD* mRNA. The pHN678 plasmid was chosen as it contains a *lac* operator sequence that allows for isopropyl β-D-1-thiogalactopyranoside (IPTG) induction control of the asRNA sequence (2). Additionally, the pHN678 plasmid establishes a paired-termini design that flanks the asRNA to stabilize it and assist it to resist degradation by double-stranded RNA-specific RNAses (summarized in the supplementary section) (2). This increased stability allows for the accumulation of the asRNA in the cell which helps amplify the effect of it (2).

**Key words:** regulatory RNA molecule, gene regulation, gene knockdown, ribosomal binding site, antisense RNA, plasmid, *wecD*, pHN678

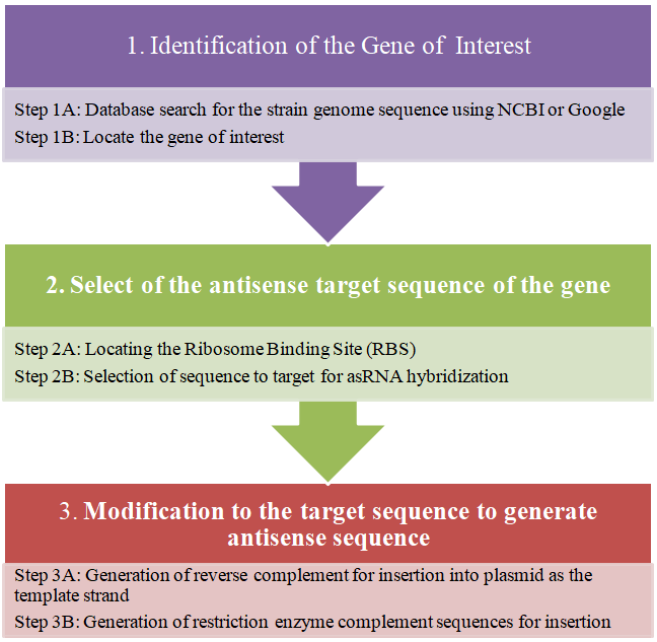
**Submitted:** May 1, 2017 **Accepted:** May 1, 2017
**Published:** May 1, 2017

## MATERIALS AND EQUIPMENT

**Web Resource Links**

- https://www.ncbi.nlm.nih.gov/
- http://parts.igem.org/Help:Ribosome_Binding_Sites/Mechanism
- http://parts.igem.org/Ribosome_Binding_Sites/Design
- http://www.idtdna.com/calc/analyzer

## PROTOCOLS AND METHODS




### 1. Identification of the gene of interest

a) **Database search for the strain genome sequence:** The initial step in designing the antisense insert for the plasmid is to locate the gene of interest. The complete genome sequence of many common strains of *E. coli*, such as MG1655 and BW25113, are available on the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/) under the search parameter "Nucleotide".

NCBI requires very specific keywords to be entered in order for it to return the desired information. For example, the query "*Escherichia coli* MG1655" returns whole genome shotgun sequences (Figure 1), while the query "*Escherichia coli* str. k-12 substr. MG1655" returns the complete genome sequence (Figure 2). As such, it is recommended to use Google search instead and use keywords including "*Escherichia coli*", "Strain designation", and "Genome".

b) **Locating the gene of interest:** To locate the gene of interest within the complete genome of MG1655, click "Graphics" and enter the gene name in the "Find" box (Figure 3). However, due to the design of the website and characterization of the gene by the authors, the graphical search option may not be able to locate genes even when they are present. Alternatively, the gene or protein product can be searched directly on the NCBI publication page after the entire page has been loaded (Figure 4).

### 2. Selection of the antisense target sequence of the genes

a) **Locating the ribosome binding site (RBS):** The RBS of well characterized genes, such as *ftsZ*, can often be found in previous literature while the RBS of poorly characterized genes must be located by examining the gene sequence directly. Based on previous knowledge, a few key characteristics are known to help locate the RBS (Figure 5).

i. The RBS is located upstream from the gene and may contain the entire or part of the Shine-Dalgarno (SD) sequence.
ii. The SD sequence is located within 6-7 base pair (bp) of the start codon of the gene (http://parts.igem.org/Help:Ribosome_Binding_Sites/Mechanism).
iii. The ribosomal protein S1 binds to sequences enriched with adenine bases found further upstream of the SD sequence (http://parts.igem.org/Ribosome_Binding_Sites/Design).

It is important to note that there exists alternate start codons other than ATG, and only a portion of the SD sequence may be present instead of the entire AGGAGG sequence.

b) **Selection of sequence to target for asRNA hybridization:** Based on characteristics defined in the previous section, a sequence can be selected upstream of the gene of interest. Integrated DNA Technologies, Inc. is used by

the lab to construct primers and gene sequences. Due to cost and accuracy, the maximum length of the insert should be limited to 60 base pairs.

3. **Modification to the target sequence to generate antisense sequence.**

a) **Generation of reverse complement for insertion into plasmid as the template strand:** During transcription, the template strand is transcribed and the resulting mRNA strand will have an identical sequence to the coding strand of DNA, but with uracil instead of thymine. Therefore, the coding strand of the target gene must be inserted as the template strand of the plasmid to generate a plasmid transcript which is complementary to the mRNA.

Direct insert of the target sequence into the plasmid will result in a 3′ to 5′ transcript that is identical to the 5′ to 3′ mRNA. However, the transcript will not hybridize to the mRNA and the gene will still be expressed. Therefore, a reverse complement sequence must be generated and inserted into the plasmid to ensure that there is 5′ to 3′ complementarity and asRNA hybridization to the target mRNA (Figure 6).

Many tools can be found online using the following search phrase: complement DNA tool. We recommend choosing a website with options that clearly indicate how the input sequence will be manipulated. For example the output sequence may be the reverse sequence, complement sequence, or reverse and complement sequence. The sequence inserted to generate proper transcript for asRNA complement can be generated by **reversing** the **coding strand sequence** or **reverse complement** the **template strand sequence** (Figure 7).

b) **Generation of restriction enzyme complement sequences for insertion:** Since the insert strands are designed, the nucleotides on 5′ and 3′ ends of the sequences must be modified to complement the product of restriction enzyme digest sites to allow ligation into the plasmid (Figure 8, 9). It is recommended to take a look at the chosen plasmid map and pick ideal restriction enzyme sites (summarized in the supplementary section).

## ANTICIPATED RESULTS AND CONTROLS

The antisense insert, as part of the constructed plasmid, is together transformed into bacterial cells. Upon induction of the plasmid, the antisense RNA transcript should be complementary to the RBS sequence of the target gene mRNA transcript, which is *wecD* in this case. This complementary binding prevents ribosome attachment and therefore silences the target gene expression by inhibiting translation. The pHN678 plasmid with no sequence insertions should be used as the negative control.

A sense insert can also be designed and used as the orientation control. The sense RNA transcript from the plasmid is designed to be the same sequence as the coding strand of the target gene mRNA instead of the complementary of it. As a result, complementary binding will not occur and the target gene will not be silenced. The orientation control sequencing results should be the same as your target gene sequence.

## ACKNOWLEDGEMENTS

We would like to thank Dr. David Oliver and Gyles Ifill for their continued guidance and endless support throughout the process of designing our antisense RNA sequence.

## REFERENCES

1. **Waters, LS, Storz, G**. 2009. Regulatory RNAs in Bacteria. Cell. 136:615-628.
2. **Nakashima, N, Tamura, T, Good, L.** 2006. Paired termini stabilize antisense RNAs and enhance conditional gene silencing in *Escherichia coli*. Nucleic Acids Res. 34**:**e138.

## FIGURES AND TABLES

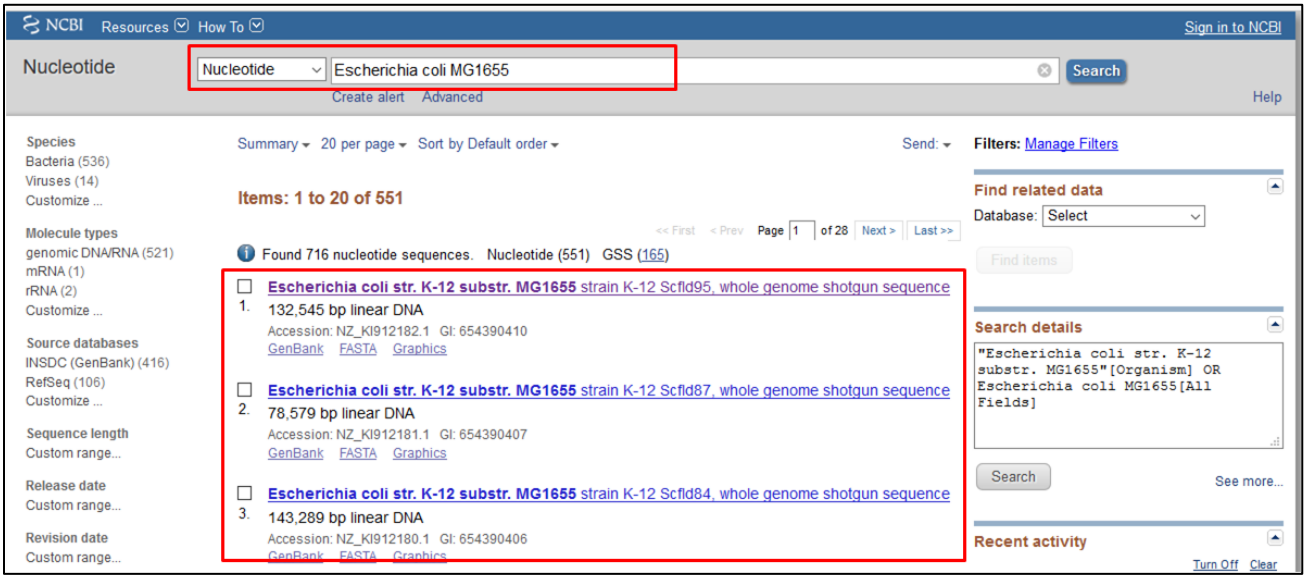



**FIG 1 Result of searching NCBI: Nucleotide with *E. coli* MG1655.** The search results generated are of shotgun fragment sequences which makes it very difficult to locate the gene of interest.

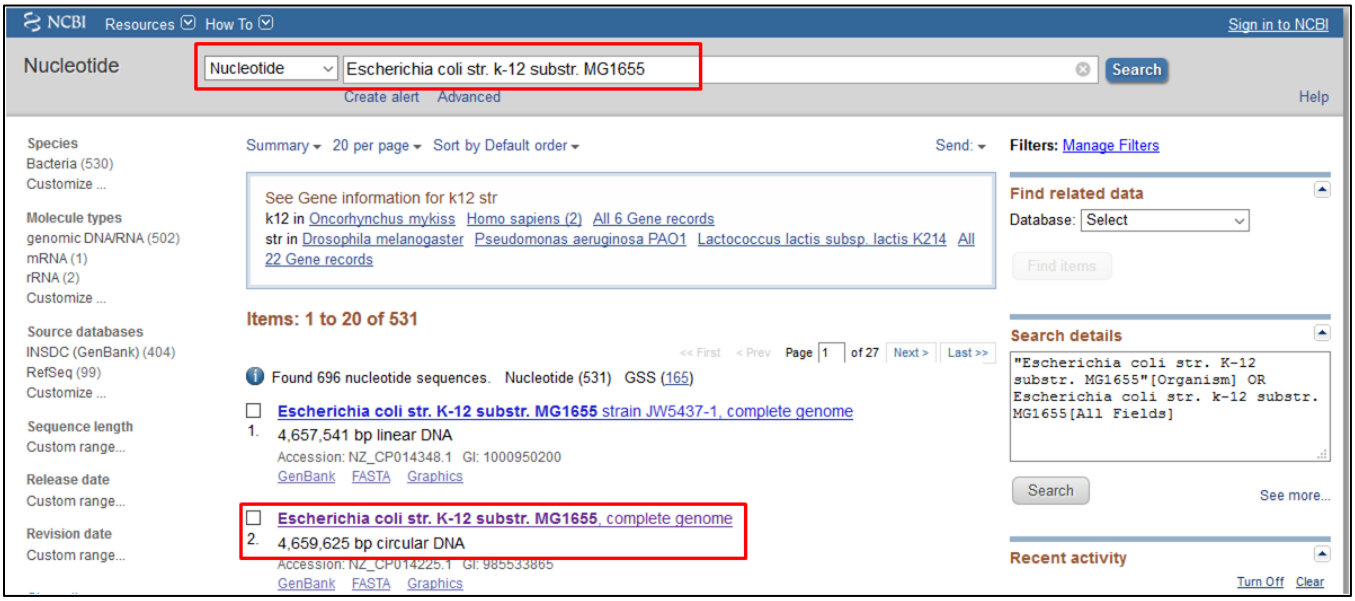



**FIG 2 Result of searching NCBI: Nucleotide with *E. coli* str. k-12 substr. MG1655.** The search result found the complete genome sequence of *E. coli* MG1655 instead of fragments.

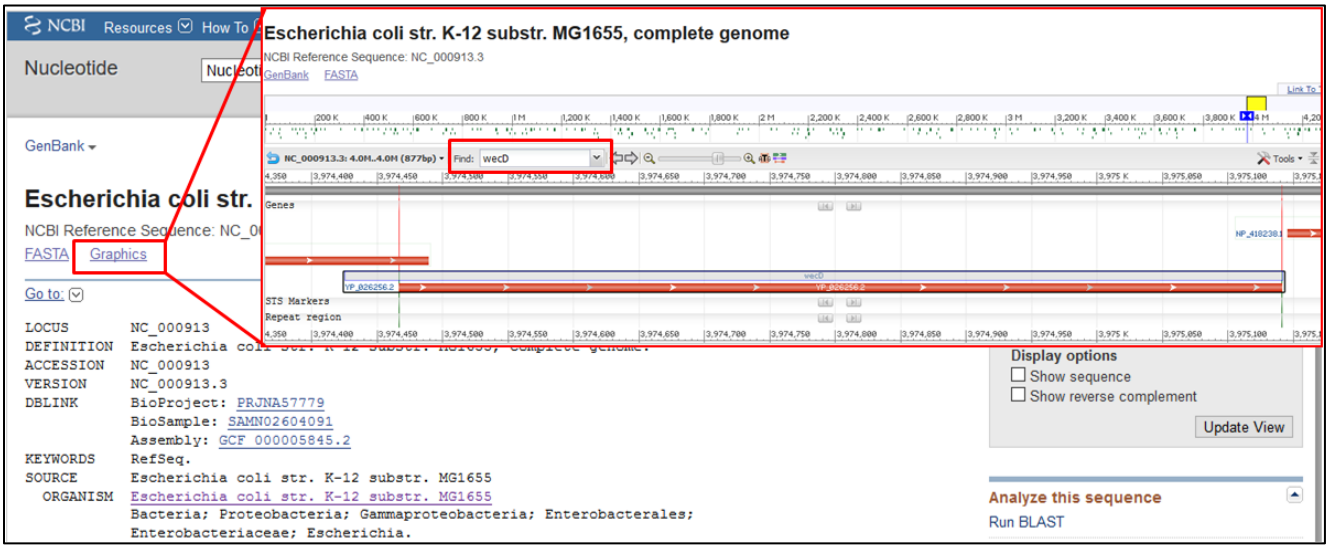



**FIG 3 Locating the gene of interest in the genome with the "Find" box on the "Graphics" webpage.**

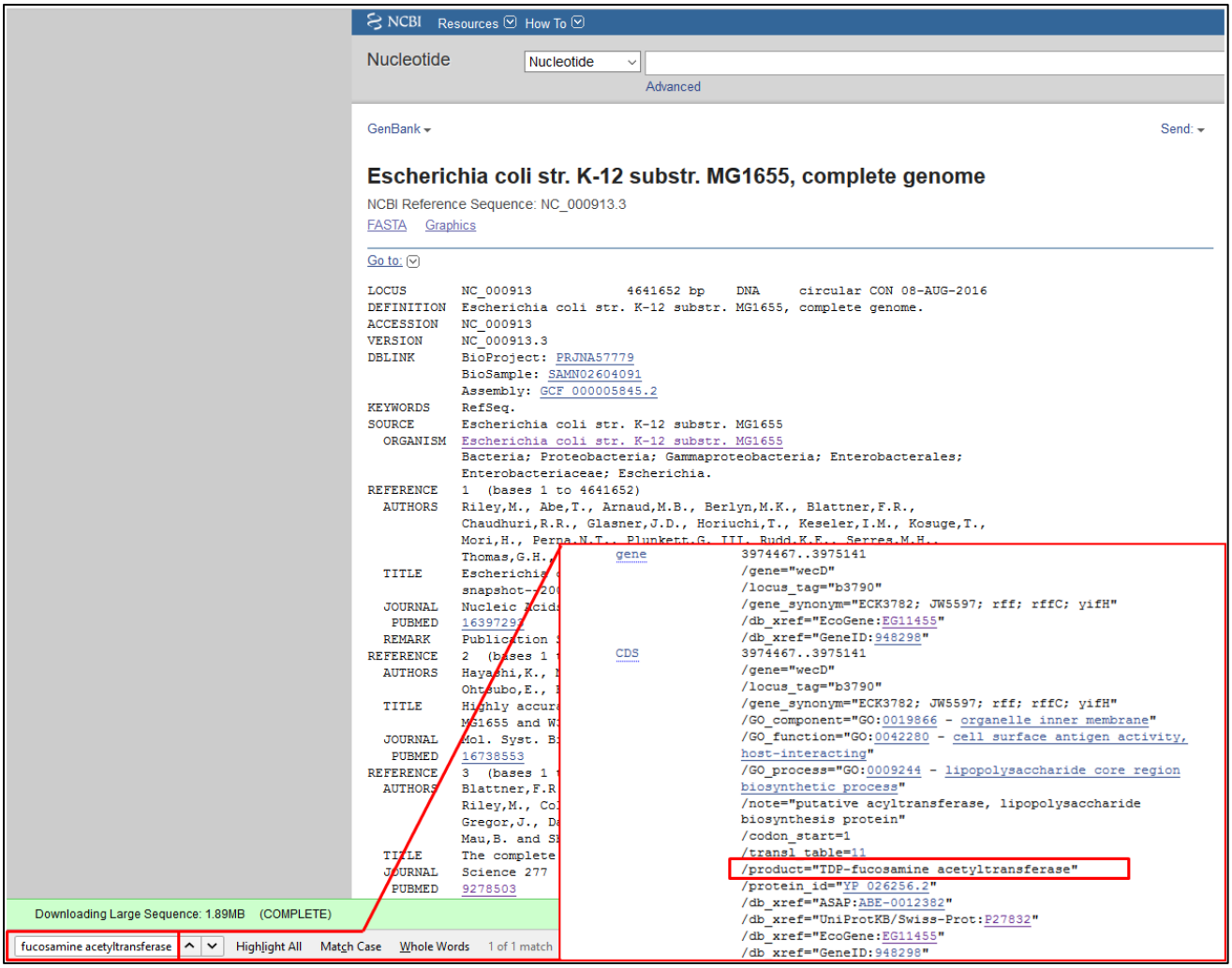



**FIG 4 Locating the gene of interest by searching for the protein product on the NCBI publication page after the entire page has loaded.**

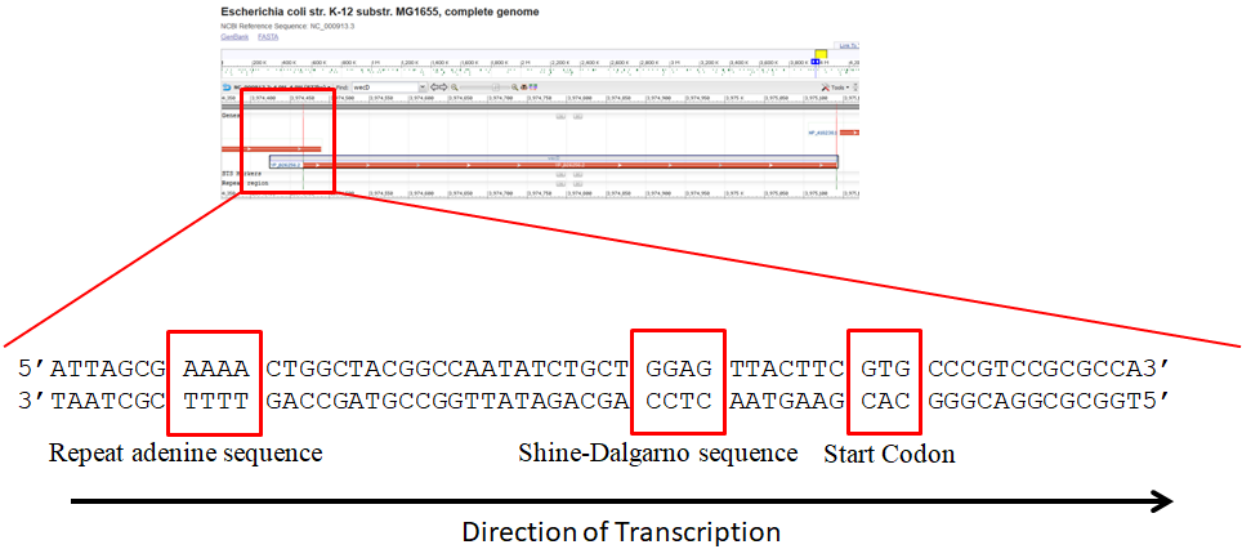



**FIG 5 Sequence chosen upstream of the *wecD* gene (also known as *rffC* and *yifH*) which encodes for TDP-fucosamine acetyltransferase.**

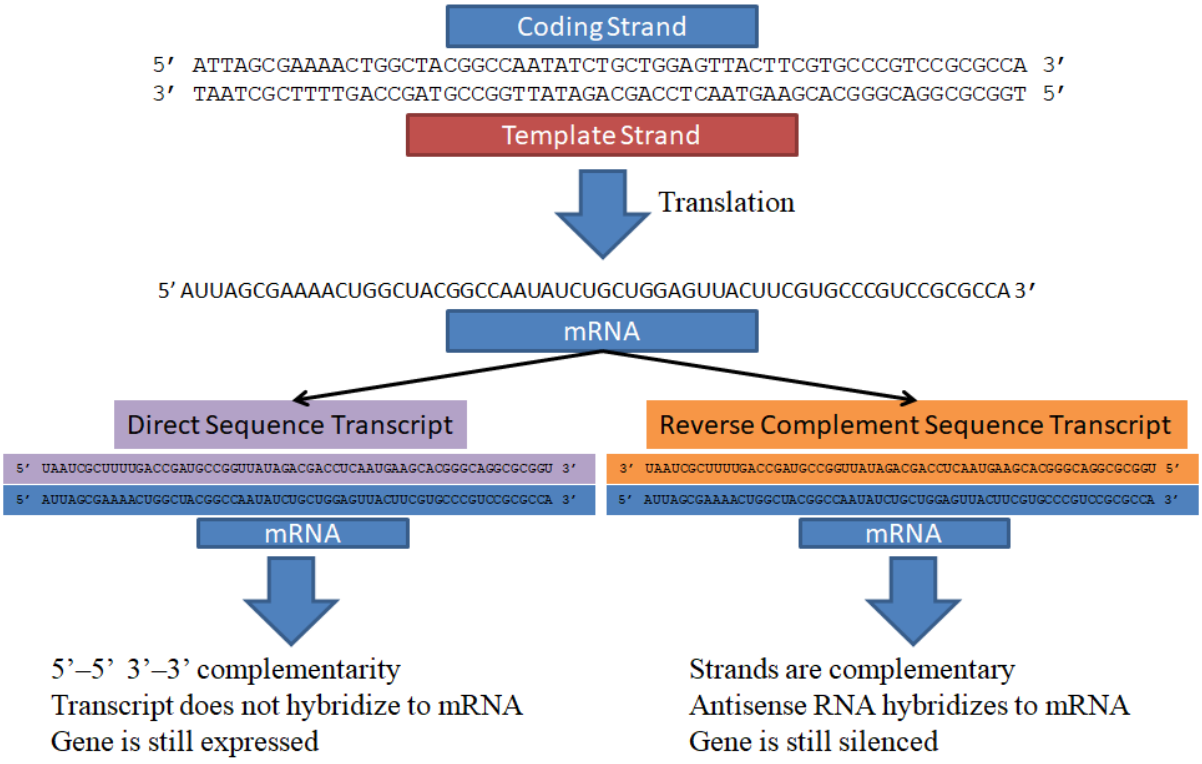



**FIG 6 Comparison between plasmid transcription products of direct sequence transcript insertion and reverse complement sequence transcript insertion.**

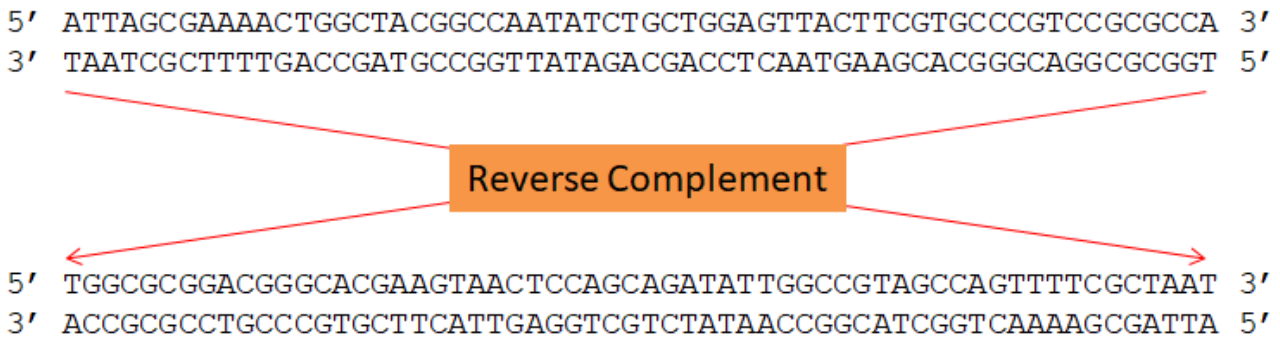



**FIG 7 Insert sequence generated by reverse complement to generate complementary plasmid asRNA transcript which hybridizes to mRNA of the target gene to achieve gene knockdown at translation.**

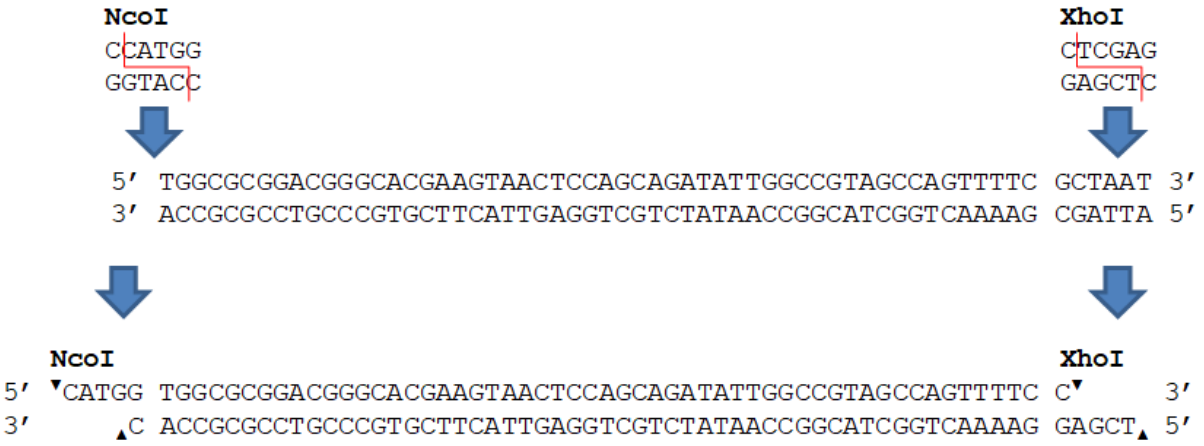



**FIG 8 Substitution of base pairs at 5′ and 3′ ends of the insert sequence to generate sticky ends and facilitate ligation into plasmid.** Using pHN678 plasmid as an example, the restriction enzyme pair between the paired termini are NcoI and XhoI.

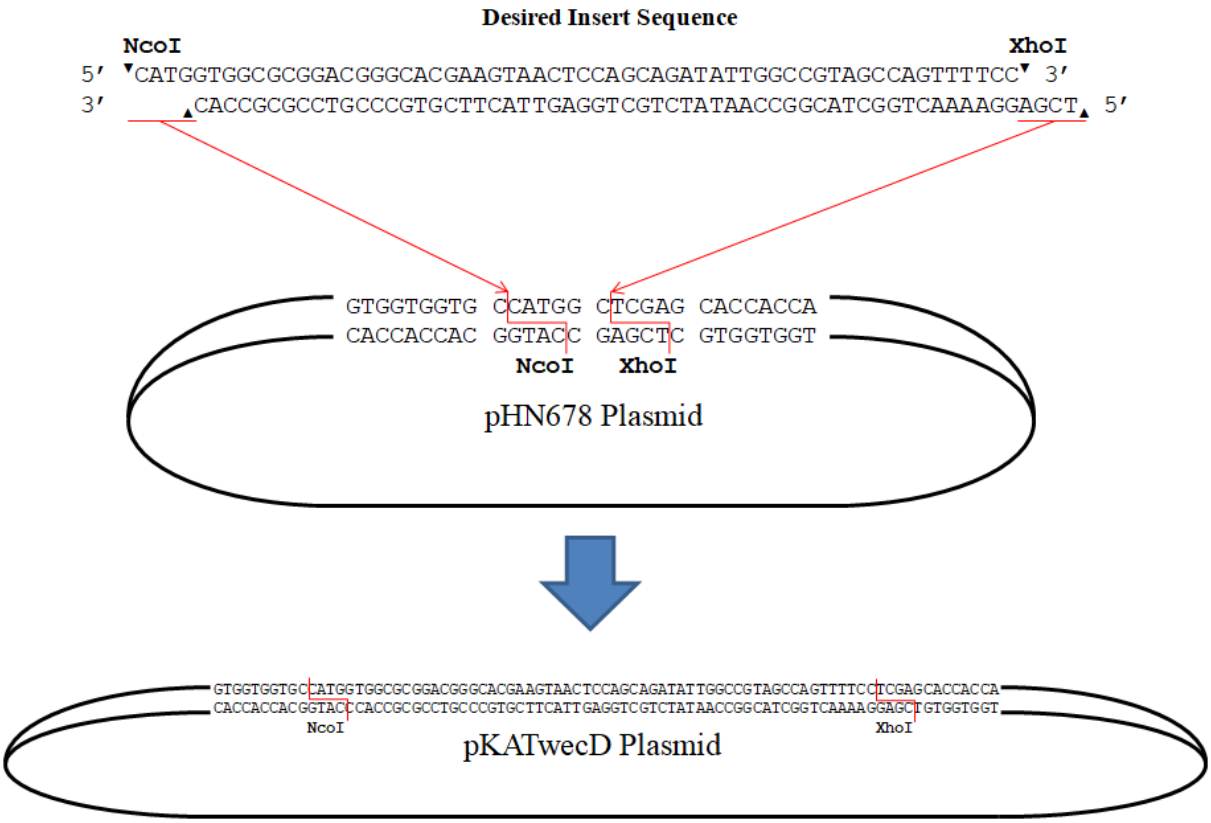



**FIG 9 Insertion of antisense *wecD* sequence into pHN678 plasmid to inhibit *wecD* expression at translation.**

## TROUBLESHOOTING TABLE

| Issues | Possible Explanations | Solution |
|---|---|---|
| Gene of interest is not found within the complete genome of the bacteria | An assortment of names may be used across the scientific literature and public biological databases for a certain gene. For example, *wecD* is also known as *rff*C or *yifH*. | ○ Verify if your gene of interest has multiple name listings or not<br>○ UniProt is a good resource for protein sequence and functional information (http://www.uniprot.org/) |
| RBS of the gene of interest is not found in sequence | Genes not well-defined often do not have the RBS sequence cited in primary literature. | ○ Look for the key sequence characteristics known to help find the approximate RBS region<br>○ Locate the start codon, the Shine-Dalgarno sequence, and the adenine-enriched region |
| Secondary structures formed by the antisense RNA | Although the asRNA is stabilized by the paired termini, it could form secondary structures which would interfere with downstream applications. The probability differs based on the nucleotide sequence and antisense transcript length. The longer the sequence, the higher the probability for secondary structure formation. | ○ Choose antisense regions where there are lower chances for secondary structure formation<br>○ IDT Oligo Analyzer can be used to predict the likelihood of secondary structure formation from a sequence. (http://www.idtdna.com/calc/analyzer) |
| Antisense RNA Target region is longer than 60 base pairs | The lab uses Integrated DNA Technologies, Inc. to construct primers and gene sequences. Due to increased error and cost, the maximum length of the sequence is recommended be limited to 60 nucleotides. | ○ Two 60 bp sequences can be designed with complementary sequences at 3′ end of the coding strand and 5′ end of the template strand and the ends are extended by PCR.<br>○ PCR extension of the sequences require digestion with restriction enzymes to generate sticky ends.<br>○ Like the 60 bp sequence, restriction enzyme digestion sites and additional nucleotides must be added to the ends of the sequence to allow proper restriction enzyme function.<br>○ The number of nucleotides required for restriction enzyme function can be found on websites of the enzyme supplier (For example, NEB: https://www.neb.com/tools-and-resources/usage-guidelines/cleavage-close-to-the-end-of-dna-fragments) |