Transforming *E. Coli* with mOrange Plasmid and Sequencing the Isolated DNA

Zan Chaudhry Partners: Andres Parra & Anjan Singh
Methods in Nucleic Acid Sequencing Lab, Spring 2023 10:30 AM

Introduction

Bacterial transformation is an important technique for academic and industrial pursuits. The ability to incorporate foreign genetic material into bacteria with functioning cellular machinery provides a mechanism for producing proteins of interest, such as life-saving drugs. Thus, transformation remains a topic of interest for scientific investigation, especially considering the recent advances in sequencing technology, which permit a more thorough analysis of the incorporated plasmid DNA than simple phenotype observations. The following report details a transformation experiment involving the incorporation of the mOrange-pBAD plasmid into DH5α *E. Coli* cells and the subsequent isolation and sequencing of plasmid DNA from these cells.

The plasmid, mOrange-pBAD, encodes the mOrange fluorescent protein and provides ampicillin (antibiotic) resistance, allowing for selection of transformed cells via ampicillin-containing media.[1] The DH5α strain consists of competent *E. Coli* cells optimized for cloning/transformation by pretreating with calcium solution and modifying the bacterial genome slightly to improve stability/yield.[2] Successful transformation was observed, both by fluorescence observations and by plasmid sequencing and assembly. A circular plasmid was reconstructed and visualized from sequencing data, with considerable overlap observed between the assembled plasmid and a reference sequence. Spectrophotmetry results corroborate these findings, suggesting the presence of high purity plasmid DNA, with an A_{260}/A_{280} value of 1.86 and A_{260}/A_{230} values as high as 2.03.

Methods

<u>Transformation</u>

Competent DH5α cells were thawed from -80 ℃ on ice. Two agar plates were incubated at 37 ℃ to prepare them for bacterial culture. Two bacterial samples were prepared in 1.5 mL Eppendorf tubes, with 50 µL of cells added to each tube. To the first tube, 2 ng of mOrange plasmid DNA were added, as this plasmid contains the gene of interest and an ampicillin resistance gene. To the second, 2 ng of Puc19 plasmid DNA were added, as this plasmid confers ampicillin resistance and thus these cells could be used as a control. The control group provides a method by which to confirm that the mOrange plasmid confers fluorescence by presenting living comparison cells that lack the mOrange gene. Cells were then cooled on ice in preparation for heat shock. The cooling increases the temperature differential between the cells and the heating block, increasing effectiveness of the heat shock. After 30 minutes, cells were placed on 42 ℃ heat blocks for 30 seconds, followed by 5 minutes on ice. The sudden heat stress increases permeability of the bacterial cell membranes by increasing pore size, allowing them to take up the plasmids in solution. The return to ice quickly reduces pore size for plasmid retention. The

cells were then incubated for 30 minutes at 37 ℃ with shaking to allow them to return to normal growth conditions. Finally, the Puc19+ and mOrange+ cells were plated on the agar+ampicillin plates, which allowed for selection of ampicillin-resistant (and thus plasmid-possessing) bacteria. The cultures were grown overnight.

Colony Selection

Following overnight incubation, plates were inspected for colonies. From the mOrange+ plate, two colonies were selected and extracted using a pipette tip. These colonies were then added to two separate 14 mL tubes, containing 5 mL of LB (growth media) with ampicillin and arabinose. The arabinose activates the production of mOrange fluorescent protein through the araBAD operon (the BAD in the mOrange-pBAD plasmid name). Arabinose in solution binds to araC protein. AraC protein is a repressor of the operon and is bound to the operator, preventing transcription. The binding of arabinose to araC causes a conformational change allowing RNA polymerase to transcribe the mOrange gene.[3] The same steps were repeated with the Puc19+ plate. Finally, the tubes were incubated overnight at 37 ℃ with shaking.

Plasmid Purification and DNA Extraction

3 mL of bacterial culture from each of the 14 mL tubes from Colony Selection was extracted and pelleted by centrifugation, 1.5 mL at a time. The pellets were then inspected/imaged for fluorescent activity in the mOrange+ group and lack of this activity in the Puc19+ group. The Puc19+ group was discarded, and the following steps pertain only to the two mOrange+ samples. All of the following steps were performed on each sample. Cells were resuspended in buffer P1, which contained LyseBlue particles to visually indicate cell lysis. Buffer P2 was then added to lyse the cells, with mixing to ensure a homogenous blue mixture. Buffer N3 was then added to neutralize the lysate, mixing until the blue color disappeared, and followed by centrifugation to collect the cell debris in a pellet. The supernatant was collected and added to a QIAprep 2.0 Spin Column to bind the plasmid DNA. This was achieved by a series of centrifugation/wash steps, discarding the flow-through each time to ensure high purity of the plasmids attached to the spin column. First, buffer PB was added and the tube was spun to bind the DNA to the column. Then buffer PE was added and the tube was spun to wash the column/increase sample purity. Finally, the plasmid DNA was eluted from the column by transferring the column to a new tube and adding buffer EB. The tube was left for 10 minutes to allow penetration of the buffer into the column membrane. Finally, the tube was centrifuged, with eluted DNA collected in the flow-through.

Nanodrop Quantification

Plasmid DNA content and purity were measured with a Nanodrop spectrophotometer. First, a baseline reading was acquired with 1 μL of ultra-pure water, followed by 1μL of the sample. After quantification, the plasmid DNA was stored at ≤-20 ℃.

Illumina Sample Preparation

eBLT and TB1 were brought to room temperature and vortexed, while the DNA samples were thawed on ice. A solution containing 500 ng of DNA with a volume of 30 μL was produced from the sample and nuclease-free water. eBLT and TB1 were then combined and mixed by

vortexing to produce a tagmentation master mix. The DNA solution was then tagmented with the master mix, by adding the mix to the DNA solution and incubating for 5 minutes at 55 ℃. In essence, the DNA present in the sample is bound to beads that attach adapter proteins while fragmenting the proteins into manageable lengths for Illumina sequencing. The adapter proteins are essential for future binding of the sequences to the flow cell used for sequencing. ST2 was then added to stop the tagmentation reaction.

After the tagmentation reaction, all non-bead-bound compounds were removed by placing the tubes with beads suspended on a magnet (the beads are magnetic). As the beads adhere to the bottom of the tube, the solution clarifies and can be discarded. This process was repeated multiple times, washing each time with TWB (a wash buffer). Finally, the purified tagmented DNA was amplified using PCR. A PCR master mix containing the necessary polymerases and nucleotides was produced from EPM and water. After adding this mix to the DNA sample and spinning down to resuspend the beads, index adapters were added. Index adapters are complementary to the previously added adapters and contain primers on the ends. Thus, these adapters are incorporated into the replicated fragments, generating strands with primer ends. Primers are essential because they allow the fragments to bind to the Illumina flow cell during sequencing.

After this amplification step, cleanup was performed to remove any remaining extraneous compounds in solution (such as unbound index adapters/polymerases/nucleotides). This involved a similar bead/magnet/wash cycle, though this time using AMPure XP beads. These are beads specifically formulated for post-PCR cleanup of barcoded products in the process of library preparation. EtOH was used for the wash steps, discarding the supernatant following magnetic collection of the beads each time (two total washes). Finally, the beads were resuspended in RSB, which also elutes the purified, fragmented DNA (with primer ends) from the beads, with an additional magnet step to collect the beads. The supernatant, containing the DNA, was collected and stored at 4 ℃ until sequencing.

Illumina Sequencing

An Illumina MiSeq was used for sequencing the purified plasmid DNA. DNA concentration was measured using a Nanodrop spectrophotometer, and library molarity was adjusted to the specifications of the sequencer. Libraries were then denatured with NaOH (to produce single strands) and diluted with HT1. The reagent cartridge and flow cell were adjusted to appropriate temperatures, followed by loading the cartridge, the sample, the flow cell, and the reagents (PR2). Finally, sequencing was run using specified parameters/protocols of the machine by the manufacturer.

Sequencing Analysis

Sequencing results (FastQ files) were analyzed with FastQC in Python to assess sequencing quality. Trimmomatic was then used to trim the reads in these files and pair the two Illumina reads together. Trimming is performed to remove any remaining adaptor sequences in the reads that may disrupt the alignment. The bacterial genome was then assembled with plasmidSPAdes and visualized with Bandage to investigate the structure of the reconstructed plasmid. Finally,

Mummer was used to to compare the assembled sequence with a reference mOrange plasmid sequence. The results from Mummer were visualized with Assemblytics to assess sequence alignment between the sample and the reference.

Results

Illumina sequencing produced high-quality reads, with a per-base quality score of over 30 throughout the sample. The mean per-sequence GC content (51%) coincided with the theoretical mean, though the distribution deviated slightly from the theoretical distribution, possessing a much more rough, jagged shape than a normal distribution, though with similar variance. Sequence length varied between 35 and 75 bp.

Colonies (Fig. 1) were observed in both the mOrange+ and the Puc19+ plates. The pellets displayed a mild orange fluorescence for the mOrange+ condition and a lack of fluorescence for the Puc19+ (control) condition (Fig. 2). Nanodrop data is present in Table 1. The A_{260}/A_{280} value for both mOrange+ samples was 1.86, very close to the expected 1.8, and the A_{260}/A_{230} values were 2.03 and 1.75. Using 2.0 as a purity benchmark, only one of the two samples had an A_{260}/A_{230} greater than 2.0, and thus only one of the samples was of high purity.

The Bandage plot (Fig. 3) depicts a relatively well-constructed circular plasmid. 28 contigs were present in the assembled genome fasta file. The Mummer plot shows good sequence alignment, with five unique alignments, three of which are significant length sequences.

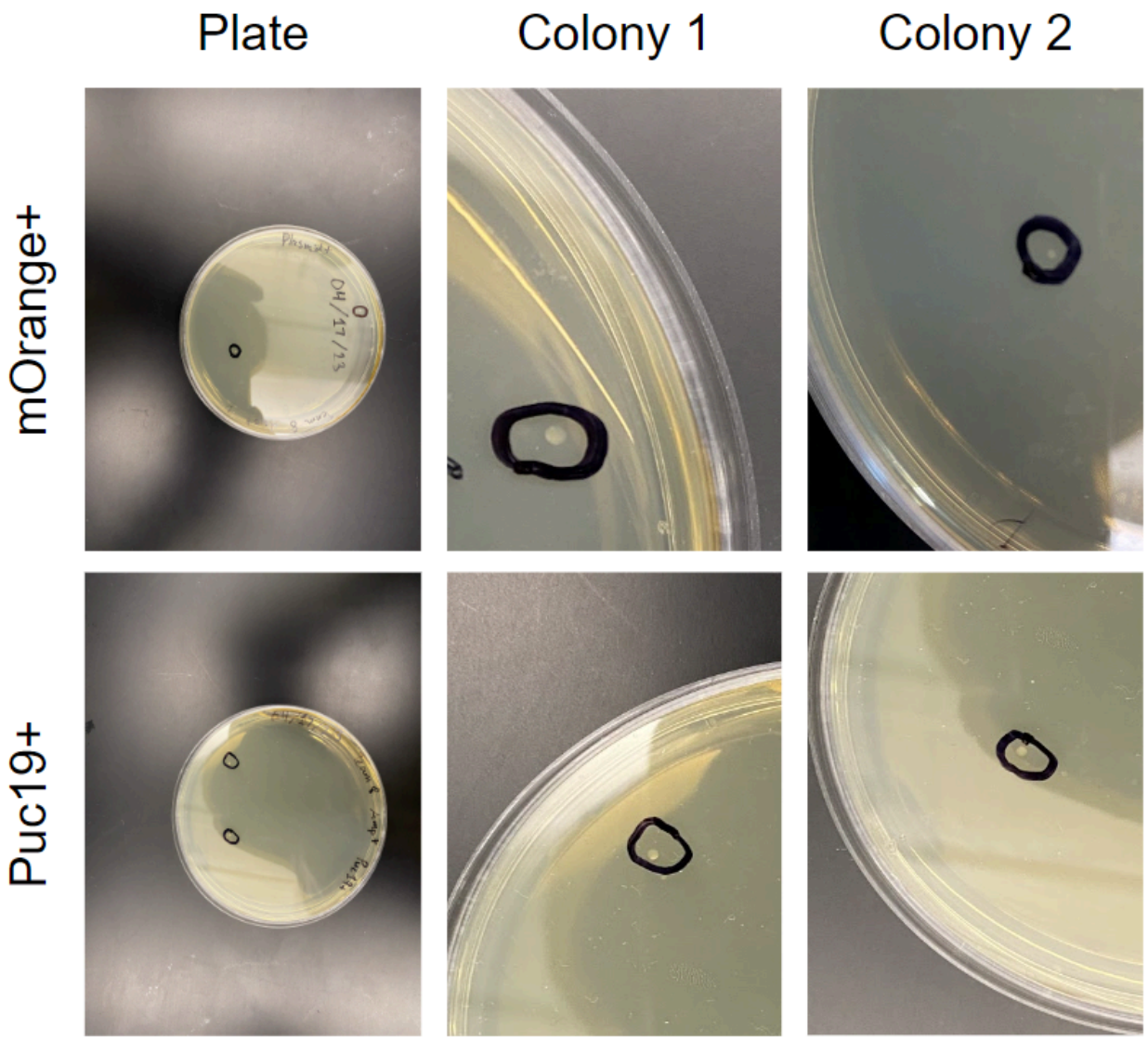

Figure 1: E. Coli culture plates and colonies. *Several colonies were observed in both the mOrange+ and Puc19+ plates (less than 5 per plate). Two isolated colonies with no microsatellites were chosen from each plate. Closeups are shown for both colonies in each plate.*

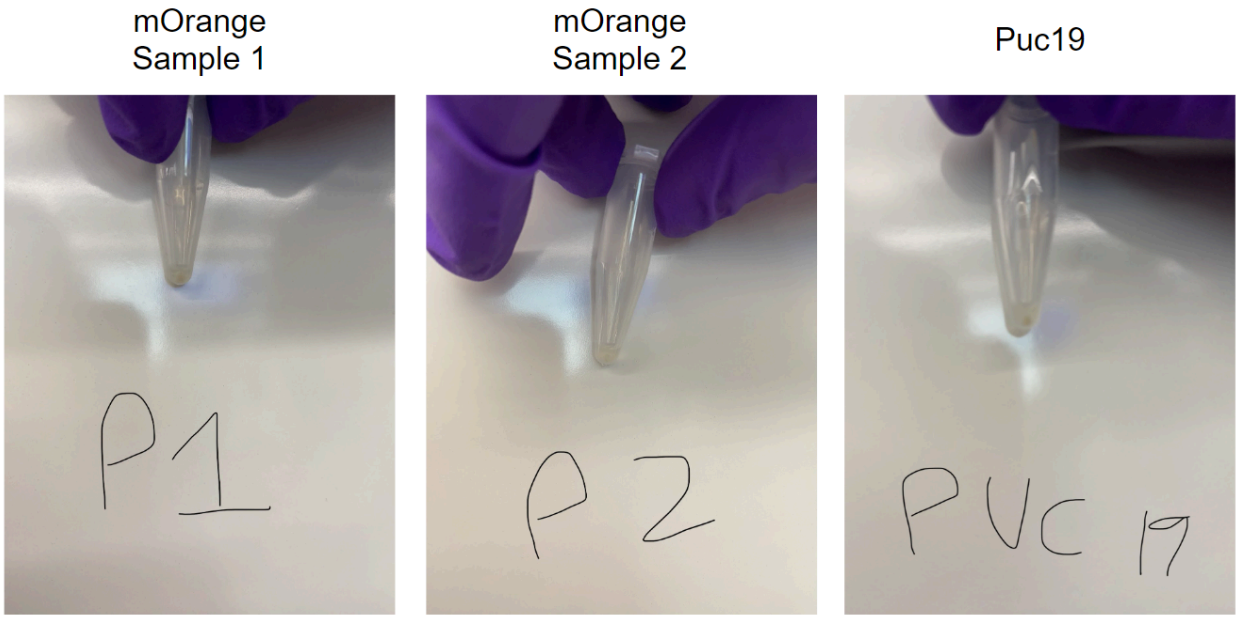

Figure 2: Liquid culture with induced fluorescence. *The fluorescence is somewhat weak, especially in sample 2; however, closely inspecting sample 1, some orange fluorescence can be observed. The Puc19+ sample contains a visible pellet, though seemingly no orange fluorescence.*

Sample:	Concentration (ng/uL):	A_{260}/A_{280}:	A_{260}/A_{230}:
Sample 1	14.9	1.86	2.03
Sample 2	9.6	1.86	1.75

Table 1: Nanodrop absorbance data. *Concentration and relevant absorbance metrics for plasmid DNA. Sample 1 contains a higher concentration and seemingly a higher purity, with its A_{260}/A_{230} over 2.0. Notice that both possess the exact same A_{260}/A_{280}.*

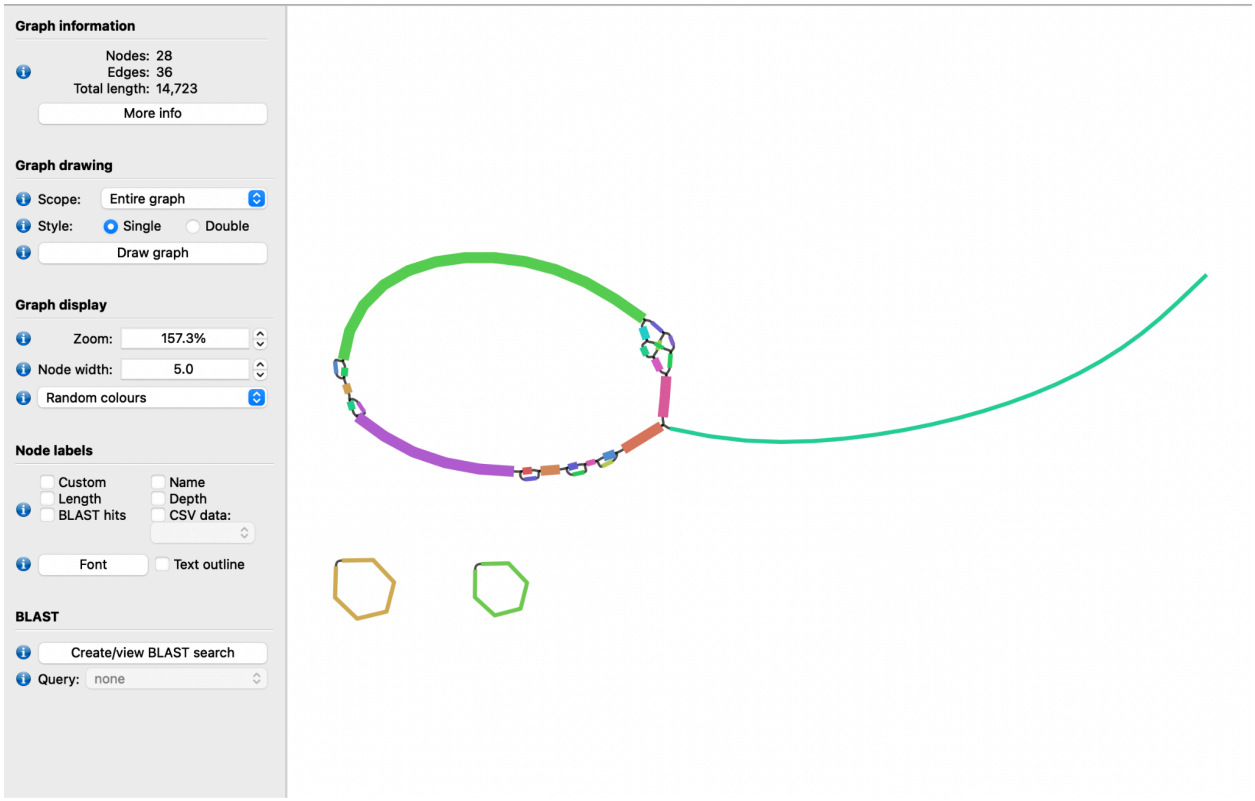

Figure 3: Bandage plasmid assembly. *Following genome assembly with plasmidSPAdes, the plasmid structure was visualized with Bandage. Notice the circular loop, indicative of a successful plasmid reconstruction from the sequencing data. There are, however, some extraneous loops and the long tail. The loops may be contaminant plasmids from the environment that were incorporated into the E. Coli. The long, disconnected tail may be indicative of a failure to reconcile the ending of the tail with the loop of DNA. This could result from the presence of too many reads leading to discrepancies in lining up the contigs (i.e. sequencing error from the large population of reads may confound the assembly, since it may seem as if multiple different contigs could fit in the same spot). This could lead the loop to be improperly closed.*

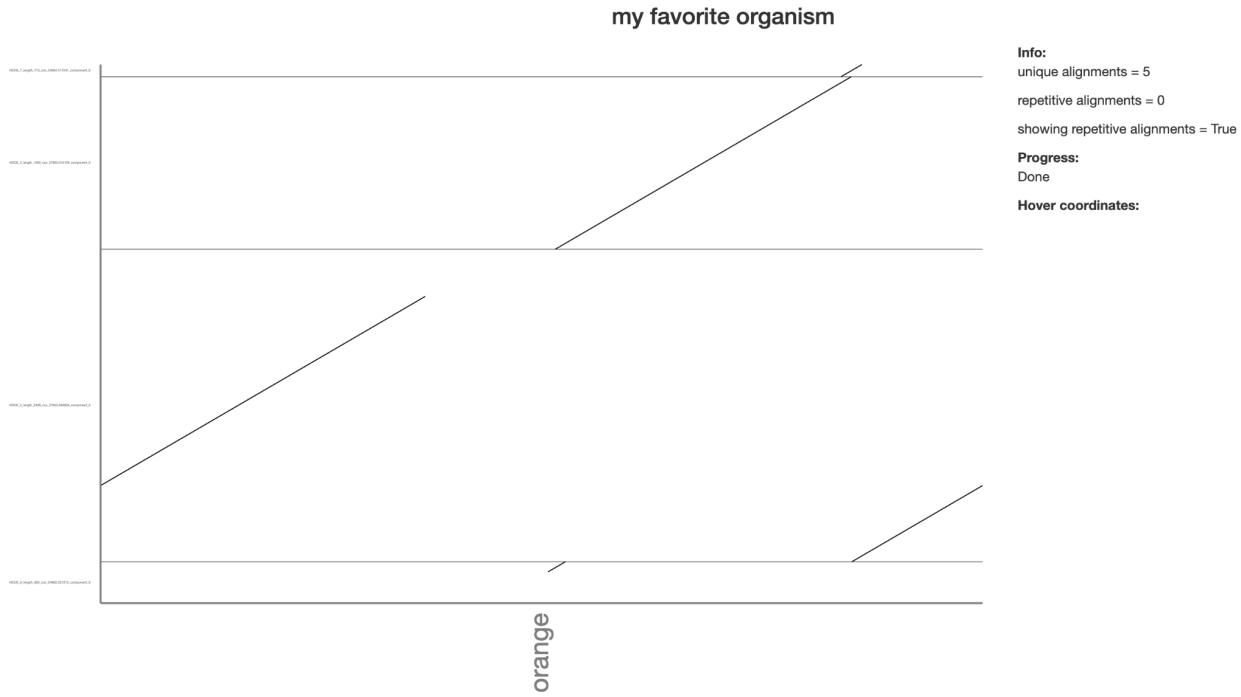

Figure 4: Mummer plot showing alignment between sample and reference plasmid sequence. *The reference plasmid is on the x-axis and the sample is on the y-axis. Notice that there are five unique alignments, three of which are of significant lengths. Also notice that the aligned sequence to the far right seems the right length to connect to the sequence on the far left. Thus, seemingly, the sequence at the beginning of the sample has been matched to the end of the reference. This may be due to the circular nature of the plasmid, which may shift the starting point between the sample and reference.*

Discussion

1. There were only a few (on the order of 1) colonies on each agar plate. This particular quantity reflects a complex mixture of conditions, including the uptake (transformation) efficiency, the viability of the cell line, the cytotoxicity of the methods used (such as the heat shock), and the balance of plasmid and bacteria concentrations. The colony count could be increased several different ways, such as increasing plasmid concentration, increasing the quantity of cells plated (i.e. performing the transformation with a larger volume of cells and plasmid), increasing the efficacy of the heat shock (by perhaps increasing the temperature differential within bounds that do not kill the cells), or using alternate transformation techniques such as electroporation (applying a high voltage to disrupt the cell membranes and make them permeable for plasmid uptake).[3] The colony count could be decreased several different ways as well, including decreasing plasmid concentration, decreasing the quantity of cells plated, decreasing the efficacy of the heat shock (by perhaps decreasing the temperature differential), and by any methods that

simply kill off the cells (i.e. adding no plasmid or using an antibiotic for which they have no resistance).

2. Bacteria were selected for by culturing in/on media containing ampicillin (an antibiotic). Both the Puc19 and mOrange plasmids contain a gene for ampicillin resistance, so any bacteria that were successfully transformed possess ampicllin resistance. Thus, only the transformed (plasmid-containing) bacteria will be able to survive while the rest of the bacteria die off from the ampicillin.
3. The culture appeared orange. The mechanism of arabinose-induced fluorescence is discussed in **Methods**. This indicates that the plasmid contains a gene coding for an orange fluorescent protein. Additionally, this observed fluorescence indicates successful uptake and expression of the plasmid by the experimental (mOrange+) group of *E. Coli*.
4. The plasmid was assembled in fragments, as seen by the many contigs that form the ring structure in Fig. 3. The presence of such a large abundance of reads for a relatively small DNA sequence may make it challenging to determine the overlap. In particular, the minimal sequencing errors on this large set of reads add up, confounding the assembly of the plasmid by suggesting multiple possible contigs that could occupy various locations. The results clearly show this, with complicated interconnections between the various fragments that compose the ring structure, rather than one contiguous, single ring structure. Downsampling could have improved the assembly by reducing the amount of reads that plasmidSPAdes is trying to reconcile, leading to fewer possible contigs and less accumulated sequencing error.
5. The aligned sequence matched the reference plasmid quite well (see Fig. 4). The majority of the sequence length matches between sample and reference, though with two small disjointed regions at the beginning and end of the sample, one larger disjointed piece at the start of the sample and the end of the reference, and a sizable gap between the two large matched regions. The larger disjointed piece that lies at the start of the sample and the end of the reference may be related to the circular structure of the plasmid. The plasmid is not a linear sequence, so the chosen starting point may be different between the sample and the reference, leading to a mismatch of regions, as seen with this larger disjointed region.at the far end of the x-axis (reference) while being at the bottom of the y-axis (sample). The gap between the two large matched regions could be attributed to poor assembly of the plasmid due to missing reads for this region, or due to sequencing errors that lead to an inaccurate assembly.

Conclusion

In the presented report, bacteria of the DH5α strain of *E. Coli* were transformed with the mOrange fluorescent-protein-encoding plasmid, using Puc19-transformed bacteria as a control. Successful plasmid uptake was demonstrated in both the mOrange and Puc19 samples by the presence of colonies on agar plates containing ampicillin. The mOrange cells were confirmed to express the fluorescent protein by the observation of mild orange fluorescence from a liquid

culture of the mOrange+ bacteria (and a lack thereof in the Puc19+ control bacteria). Additionally, the plasmid DNA was extracted from the mOrange+ group, purified, and seqeueced to confirm the presence of the mOrange plasmid. This was achieved by Illumina sequencing and plasmid assembly with plasmidSPAdes. Additionally, plasmid DNA concentrations and purity were assessed with spectrophotometry, which demonstrated high purity ($A_{260}/A_{230} > 2.0$).

The present study could be improved by increasing the number of colonies present to begin with. Very few colonies formed on the agar+ampicillin plates, suggesting a poor plasmid uptake efficiency, which could be improved by the means suggested in **Discussion**. Continuing in this vein, downsampling of the reads could assist in plasmid assembly by reducing the number of contigs to assemble, decreasing the amount of discrepancies in trying to align/assemble the contigs. Ultimately, this study explores bacterial transformation and different methods of measuring plasmid uptake, particularly sequencing and genome assembly. This work has exciting applications in academic exploration of novel molecules and industrial production of therapeutics to save lives.

References

1. "mOrange-pBAD (Plasmid #54751)." *Addgene*, www.addgene.org/54751/. Accessed 9 May 2023.
2. "DH5α Competent Cells." *Thermo Fisher Scientific - US*, www.thermofisher.com/us/en/home/life-science/cloning/competent-cells-for-transformation/competent-cells-strains/dh5a-competent-cells.html. Accessed 9 May 2023.
3. Schleif, Robert. "AraC Protein, Regulation of the L-Arabinose Operon in Escherichia Coli, and the Light Switch Mechanism of AraC Action." *FEMS Microbiology Reviews*, vol. 34, no. 5, 2010, pp. 779–796, https://doi.org/10.1111/j.1574-6976.2010.00226.x.
4. Timp, Winston *Methods in Nucleic Acid Sequencing Lab: Module 2 Bacterial Cloning,* 2023.