12-2021

# Analyzing Gene Expression Profiles of a Virus and its Host During Infection

Elizabeth Reisher
erussman@unomaha.edu

# Analyzing Gene Expression Profiles of a Virus and its Host During Infection

University Honors Program Thesis

University of Nebraska at Omaha

Elizabeth (Russman) Reisher

BIOL4560

December 2021

William Tapprich, PhD

wtapprich@unomaha.edu

**UNIVERSITY OF NEBRASKA AT OMAHA**

**HONORS THESIS PROPOSAL ABSTRACT**

PROGRAM: UNIVERSITY HONORS PROGRAM

DIRECTOR: LUCY MORRISON

ADDRESS: KH 208

UNIVERSITY OF NEBRASKA AT OMAHA

OMAHA, NE 68182

TELEPHONE: (402) 554-2696

**Abstract of thesis:**

In recent years, RNA sequencing has become an important part of gene expression analysis. RNA sequencing applications are used to study many aspects of RNA structure, expression, and translation. With developing technologies, RNA sequencing is used to learn more about the biology of RNA, helping to understand more about what RNA does under different conditions, such as when the RNA's host is under attack from a virus. New RNA sequencing technologies allow researchers to learn more about what happens to the host when there is an attack from an outside source, such as a virus. RNA sequencing also tells the researcher how the attacker successfully attacks the host and how the host responds. In this study, RNA sequencing is used to understand how a bacterial virus attacks its bacterial host.

## Table of Contents

## Introduction

Bacteriophages (also referred to as “phages") are a specific type of virus that infects and kills bacteria (*Bacteriophage / Phage | Learn Science at Scitable*, n.d.). They are also the most prevalent biological entity in the biosphere (Clokie et al., 2011). One such phage, Glacier Creek Virus (GC) was discovered by UNO students in a water sample from Glacier Creek in Omaha. GC is a member of the *Myoviridae* family in the order *Caudovirales*. This order of viruses is characterized by non-enveloped viral particles composed of a head and tail structure. The tubular tail structure uses a contractile mechanism to inject the viral DNA genome into the host bacteria. *Caudovirales* compose around 96% of all discovered phages (Fokine & Rossmann, 2014). Their genomes consist of double stranded DNA that can range from 18,000 base pairs (18kB) to 500,000 base pairs (500kB) (Maniloff & Ackermann, 1998). For Caudoviruses and for that matter bacteriophages in general, the majority of genes identified by bioinformatics analysis do not have sequences that produce proteins with a known function. These unknown proteins are referred to as “hypothetical proteins". Due to the large number of “hypothetical proteins", it is difficult to determine what proteins are occurring during the different stages of infection and what is being expressed during the process. By exploring the gene expression profile throughout the infection cycle of the myovirus, GC, this project hopes to determine which genes are being expressed before, during, and after infection of the host. This profiling will analyze expression of the viral genome and the expression of the host genome simultaneously.

*Pseudomonas fluorescens* is the host bacteria used in this study. *Pseudomonas fluorescens* is one of the most abundant species of bacteria and is easy to use. There are not many diseases and infections attributed to *Pseudomonas fluorescens* in humans, making using *Pseudomonas fluorescens* relatively safe for laboratory study (Scales et al., 2014). Many

different metabolic capabilities allow *Pseudomonas fluorescens* to survive in diverse host environments, even outside of a mammal (Scales et al., 2014). Having diverse host environments helps the bacteria to be resilient, which also makes it a good candidate for use in laboratory studies. The *Pseudomonas fluorescens* bacteria are motile rods, generally aerobic, cannot ferment glucose, and are chemoorganotrophs, which are organisms that use organic materials or oxygen as their primary energy source. In the laboratory, the bacteria grow best in a nutrient media that contains agar (Scales et al., 2014). *Pseudomonas fluorescens* colonies in this study were grown in 691 media plates that contain agar to promote robust growth.

Understanding the infection process is important when characterizing a virus. Virus particles can enter the cell by attaching to the host cell's membrane or wall and inserting their genetic material into the cell (*Virus | Learn Science at Scitable*, n.d.). The virus is then able to replicate its genome by attacking the host cell to make new viral capsids, which then break out of the cell (*Virus | Learn Science at Scitable*, n.d.). Viral infection usually causes the host cells to die, leading to the virus overtaking portions of the host organism. During this process, the virus causes the host to produce proteins necessary for the virus to survive and multiply (*Understanding Viral Gene Expression*, 2017).

For gene expression purposes, analyzing a sample of purified RNA from the infected host cell provides a clear understanding of what proteins are expressed in the cell at any given time during infection. Genes code for proteins in the cell and proteins give insight on the function of the cell at any given time (*Gene Expression | Learn Science at Scitable*, n.d.). RNA sequencing is useful to provide insight and knowledge of the transcriptome, which is the entire constellation of genes expressed in the cell (Kukurba & Montgomery, 2015). RNA sequencing data is useful for researchers to understand more about transcription and potentially identify novel transcripts in

the cell (Kukurba & Montgomery, 2015). RNA sequencing has been used to characterize RNA as a biological molecule through gene expression when exposed under various conditions. Gene expression analysis of the RNA sequencing data allows us to get a visual insight of what the cell is doing, what viral and host proteins it is making, and the various pathways for the transfer of information in the cell.

KBase is an open source software that houses many Bioinformatics tools that aid in analysis of sequencing data (Arkin et al., 2018). KBase is maintained by The US Department of Energy and is unique in that no other platform is able to integrate all of the features that KBase has available on its server (Arkin et al., 2018). KBase is user friendly and helps to maintain a clean workspace and environment since it stores the outputs without needing them to be downloaded to the device for further analysis. This helps to maintain the entirety of the output and minimizes any loss of data due to downloading between steps. In addition to being user friendly, KBase also allows for its users to search for algorithms based on the input or output types, allowing for easy comparison and validation of the results from another output.

The goal of this proposed project is to determine what the bacterial cell and the virus is doing during varying points during the infection process. This includes steps the bacteria is taking in its attempt to protect itself from infection and what the virus is doing to advance the infection process in the bacteria. Understanding what is going on in the cells at various points during the infection process is vital for researchers to better characterize the virus and the infection process. Researchers can also learn more about the transcriptome of the host as it is being attacked by a virus and compare the results to a normal functioning host.

## Methodology

The host, *Pseudomonas fluorescens,* was prepared by growing isolated colonies from glycerol storage stock on 691 media plates and incubated for 12-16 hours at 26°C. A resulting bacterial colony was lifted from the plate using a toothpick then resuspended in 200 μL of 691 media in a microcentrifuge tube. This mixture was streaked on a fresh 691 media plate and again grown for 12-16 hours at 26°C. This ensured that bacterial colonies were uniform and well isolated.

An overnight culture of *Pseudomonas fluorescens* was set up by resuspending a single *Pseudomonas fluorescens* colony in 200 μL of 691 media and added to 10 mL of 691 media in a 125 mL flask. The overnight culture was placed in a 26°C shaker for 14 hours. A log phase culture was prepared by adding a final volume of 2% of the overnight culture to fresh 691 media (50 mL fresh 691 and 1 mL overnight). Once the culture hit early log phase ($A_{600}\approx0.4$), the culture was split into 2 flasks and the first sample collection before the infection occurred. One culture was inoculated with a ratio of phage particles to host cells (MOI) of 100:1. SM media was added to the control flask to allow the volume of both the control and infected flasks to be equal. RNA was extracted at various time stamps after the infection occurred. Samples were extracted at the 0-, 5-, 10-, 20-, 30-, 40-, 50-, 60-, 90-, and 120-minute mark after infection for both the infected and control culture and one sample was extracted before infection. Each sample collected was 1 mL and cell density was measured using the spectrophotometer. The absorbance (bacterial concentration) was recorded for each sample of the infected and the control culture. Once the sample was collected and measured, bacterial cells were pelleted by centrifugation in a microcentrifuge tube at maximum speed for 1 minute. The supernatant was carefully decanted to

a fresh microcentrifuge tube. The microcentrifuge tubes containing bacterial pellets were then placed in a -20°C freezer before nucleic acid extraction would occur.

The RNA extraction kit was obtained from VWR Life Science. To prepare the lysate from bacteria, it is important to note that frozen pellets should not be thawed before conducting this portion of the experiment. A solution of TE-Lysozyme was prepared at 1 mg/mL and was prepared with a sterile, RNase-free TE buffer at pH 8.0. Once prepared, the solution was kept on ice until needed. The TE-Lysozyme was used to resuspend the bacteria by adding 100 µL to each frozen pellet. Once added, the pellet was vortexed for 30 seconds and then incubated at room temperature for 5 minutes. After incubation, 300 µL of the Lysis solution that was provided in the kit was added to the pellet and vortexed for 10 seconds. 200 µL of 100% ethanol was added to the lysate and vortexed for another 10 seconds.

After the lysate was prepared, the next step was to bind the RNA to the column. A column for each sample was assembled using the collection tubes that were provided in the kit. 600 µL of the lysate with the ethanol was added to the column and centrifuged at 14,000 x g for 1 minute. The flow through was discarded and 400 µL of the provided wash solution was added to the column and centrifuged at 14,000 x g for 2 minutes. A working solution of 0.25 Kunitz unit/µL RNase-free DNase 1 in 20 mM Tris, pH 8.3, 2 mM $MgCl_2$ was prepared. Each column needed 100 µL. Each column received 100 µL of the RNase-free DNase 1 solution and centrifuged at 14,000 x g for 1 minute. The eluate that was present in the collection tube was pipetted back onto the same column and incubated for 15 minutes at 25-30°C.

After incubation, 400 µL of the wash buffer was added to the column and centrifuged at 14,000 x g for 1 minute. Once done, the flow through was discarded and the wash step was

repeated. After a total of 3 washes, the column was placed into a collection tube and centrifuged at 14,000 x g for 2 minutes to dry the resin. For RNA elution to occur, the column was placed in a new RNase-Free microcentrifuge tube and 50 μL of the elution solution was added to the column. The column was centrifuged at 200 x g for 2 minutes followed by another 2 minutes at 14,000 x g. Once RNA is extracted, store the samples were stored at -80°C.

The samples were then measured for RNA using a Qubit Fluorometer. To confirm the presence of mRNA in order for proper sequencing to occur, gel electrophoresis was done on the samples. In a 125 mL flask add 0.45 g agarose was added to 27 mL of RNase Free dd$H_2O$ and microwaved until boiling and the solution was clear. The solution was cooled until the flask could be held without burning the hand. 3 mL of 10X TAE with 1 μg/mL Ethidium Bromide was added to the flask and the mixture was poured into a casting tray. An RNA ladder was prepared by adding 5 μL of each HR RiboRuler and RNA gel dye. Samples were prepared by adding 2 μL of the sample, 2 μL of RNA gel dye, and RNase Free dd$H_2O$ to make a total of 10 μL. Samples were mixed well and heated at 90°C for 2 minutes then placed on ice. Once the gel solidified, samples were centrifuged then loaded onto the gel. The gel was electrophoresed at 70 mV for 90 minutes.

The selected Bioinformatics pipeline was done on the sequencing results to analyze the infection process (Figure 1). Once the sequencing results from Illumina are ready for analysis, the pipeline is used to characterize the expression of the infection. First, an RNA-seq sample set needs to be created using the "Create RNA-seq Sampleset" tool on KBase. Once created, the sample set is then aligned to the *Pseudomonas fluorescens* reference genome using HISAT2. It is important to note that in order to properly receive results for this pipeline, a genome object must be used when aligning with HISAT2 The alignments then are assembled using StringTie and an

expression matrix was created using DESEq2. DESEq2 allows for the gene to be viewed as well as significance, fold change, p-value, and q-value for the data (Figure 5). The TPM expression matrix was filtered using an ANOVA filtering tool on KBase. When adding the parameters for the filter expression matrix step, add a string label in the filtered expression matrix parameter and a different string label for filtered feature set parameter. These labels will help store your results for easy access. The results were then clustered using a k-means clustering k determining algorithm to produce the most ideal k for the clustering. Once done, use the filtered expression matrix in the k-means clustering algorithm to produce results of gene clustering (Figure 6). To determine the proteins involved on the phage genome alignment, the pipeline was followed through assembly to see the proteins involved from the phage genome (Figure 8).

**Figure 1**

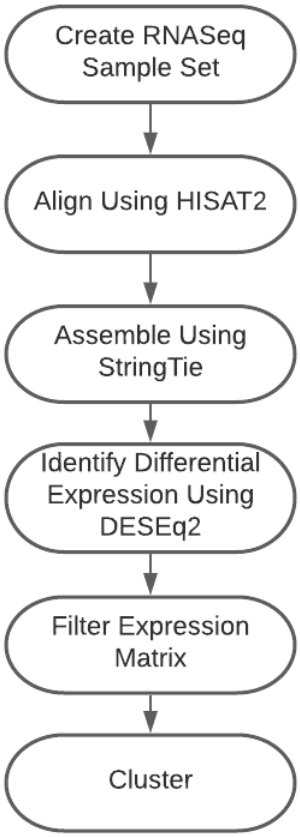


**Figure 1: Bioinformatics Pipeline for Analysis of the Infection Process** This figure shows the pipeline that was used on the sequences to generate information on the infection process. All algorithms used are free.

**Figure 2**

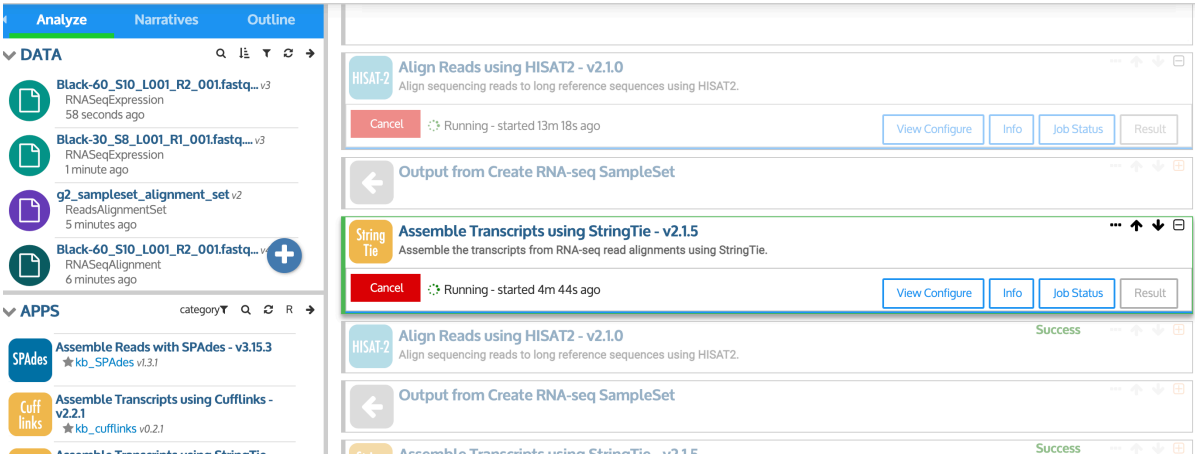


**Figure 2: KBase Narrative Image** This figure shows an image of the KBase narrative that was used to analyze the expression of the RNA-Seq sampleset.

## Results

A successful infection occurred in the results from the September 27, 2021 run, however, when analyzed for total RNA extraction, the concentration RNA measured via the Qubit Fluorometer was low and sequencing did not produce any results (Figure 3). Another successful infection occurred from an experiment conducted on October 27, 2021, shown in Figure 4 Unfortunately, due to time constraints required of sequencing the experiment conducted on October 27, 2021 could not be sequenced in time for development and analysis of the pipeline. A research colleague's sequencing data was used to develop the pipeline. The pipeline assembles and aligns the RNA-seq sample set and produces results for analysis of the infection process. Figure 5 shows the results from the DESEq2 algorithm, which shows statistical analysis. The data labels with black are the control samples and the ones labeled with blue are the infected samples. Figure 6 shows the heat map results after clustering the samples. Each blue, filled in square represents that the gene was expressed in the sample. Figure 7 shows each feature ID and

function of the squares in the heat map. Most of the gene functions listed are involved in translation and the others are involved in energy transfer in the cell. Figure 8 shows the results of aligning and assembling the sample set with the phage genome. The proteins that are being expressed with this alignment are phage attaching to host proteins and proteins that help the phage take over the host to properly infect the host.

**Figure 3**

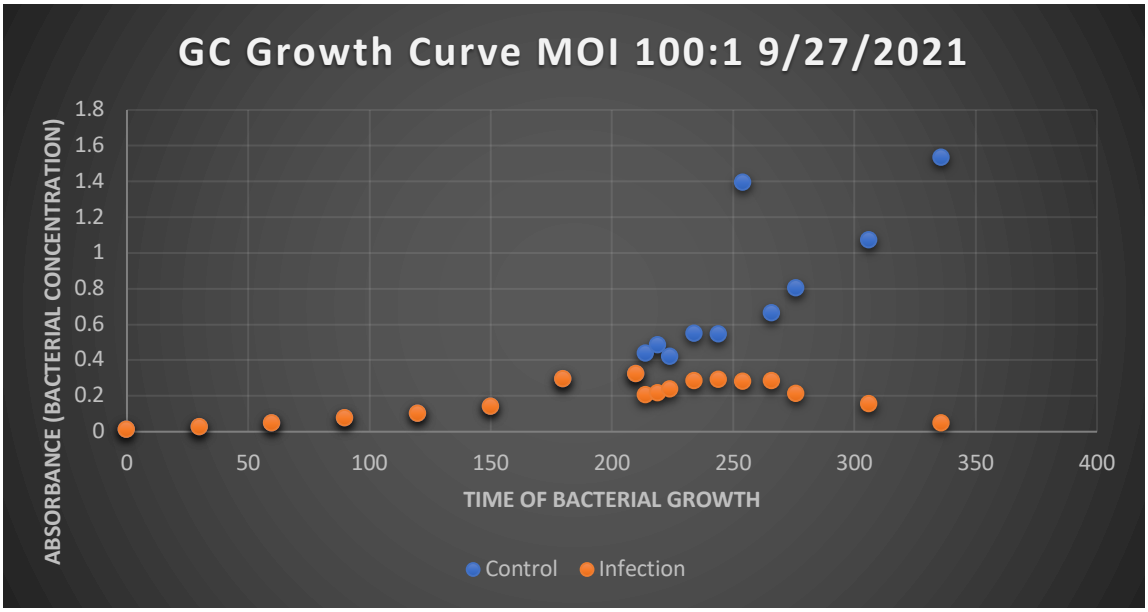


**Figure 3: Growth Curve of GC with MOI of 100:1** Growth Curve of GC with an MOI of 100:1 produces results to conclude that an infection has occurred. The control continued to grow exponentially, while the infection produced a bell-shaped like graph.

**Figure 4**

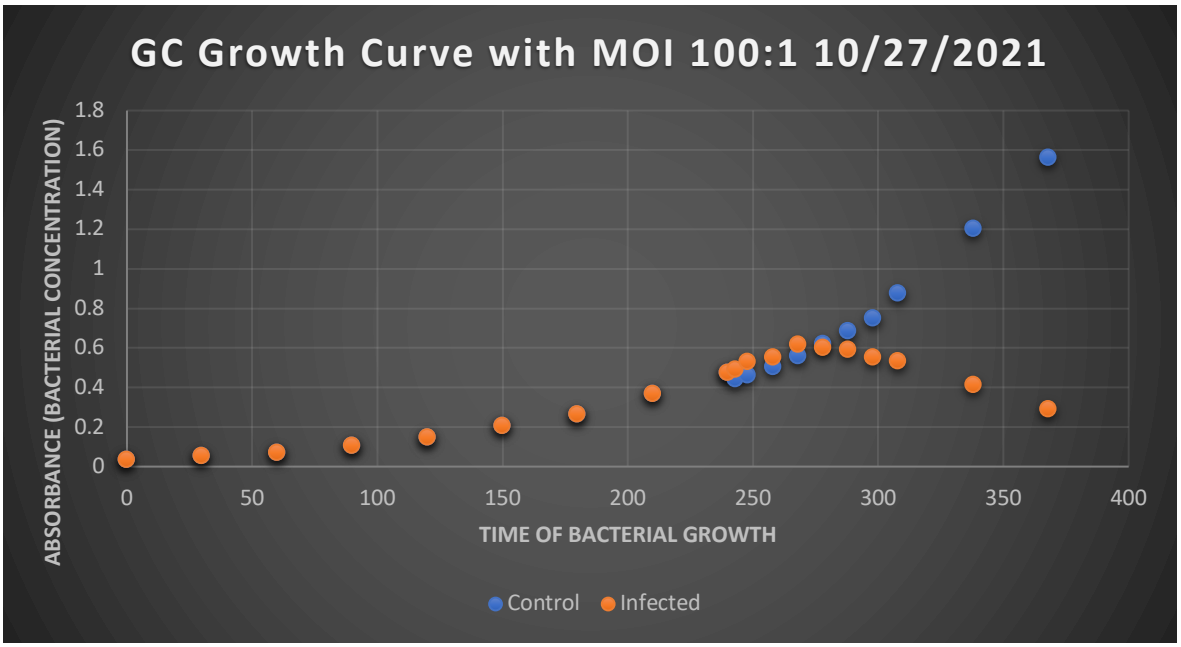


**Figure 4: Growth Curve of GC with MOI of 100:1** Growth Curve of GC with an MOI of 100:1 produces results to conclude that an infection has occurred. The control continued to grow exponentially, while the infection produced a bell-shaped like graph.

**Figure 5**

Show 10 entries Search:

| Gene | p-value | q-value | Significance (-Log10) | Fold Change (Log2) |
|---|---|---|---|---|
| DGGOJLHL_00019 | 5.51885E-1 | 9.73139E-1 | 1.18252E-2 | -1.19480E0 |
| DGGOJLHL_00183 | 2.67932E-1 | 9.73139E-1 | 1.18252E-2 | -1.89089E0 |
| DGGOJLHL_00185 | 6.22871E-2 | 9.73139E-1 | 1.18252E-2 | -2.75914E0 |
| DGGOJLHL_00188 | 5.75485E-1 | 9.73139E-1 | 1.18252E-2 | -1.74595E0 |
| DGGOJLHL_00204_rRNA_3 | 6.76349E-1 | 9.73139E-1 | 1.18252E-2 | 1.31364E0 |
| DGGOJLHL_00321_rRNA_6 | 4.32132E-1 | 9.73139E-1 | 1.18252E-2 | 2.15905E0 |
| DGGOJLHL_00322_rRNA_7 | 1.90378E-1 | 9.73139E-1 | 1.18252E-2 | -1.16075E0 |
| DGGOJLHL_00577 | 6.36823E-1 | 9.73139E-1 | 1.18252E-2 | 1.48154E0 |
| DGGOJLHL_00762 | 2.19156E-1 | 9.73139E-1 | 1.18252E-2 | 1.78678E0 |
| DGGOJLHL_00763 | 4.56353E-1 | 9.73139E-1 | 1.18252E-2 | -2.09968E0 |

Showing 1 to 10 of 41 entries Previous 1 2 3 4 5 Next

**Figure 5: DESEq2 Results** This figure shows the results of the DESEq2 run that shows the significance, p-value, q-value, and fold change for each gene in the expression matrix.

**Figure 6**

| Condition ID | Min | Max | Average | Std. Dev. | Expression Values |
|---|---|---|---|---|---|
| Black-0_S6_L001_R2_001.fastq.gz_reads_expression | 10.62 | 14.30 | 12.44 | 1.84 | |
| Black-10_S7_L001_R1_001.fastq.gz_reads_expression | 8.71 | 14.26 | 11.69 | 2.72 | |
| Black-30_S8_L001_R1_001.fastq.gz_reads_expression | 11.00 | 15.23 | 13.30 | 1.80 | |
| Black-40_S9_L001_R1_001.fastq.gz_reads_expression | 0.00 | 14.72 | 9.61 | 6.63 | |
| Black-60_S10_L001_R2_001.fastq.gz_reads_expression | 0.00 | 15.02 | 10.47 | 7.04 | |
| Blue-0_S1_L001_R1_001.fastq.gz_reads_expression | 0.00 | 9.67 | 2.42 | 4.83 | |
| Blue-10_S2_L001_R2_001.fastq.gz_reads_expression | 0.00 | 14.44 | 6.75 | 7.83 | |
| Blue-30_S3_L001_R2_001.fastq.gz_reads_expression | 0.00 | 13.79 | 9.63 | 6.52 | |
| Blue-40_S4_L001_R1_001.fastq.gz_reads_expression | 0.00 | 14.52 | 3.63 | 7.26 | |
| Blue-60_S5_L001_R2_001.fastq.gz_reads_expression | 9.13 | 14.42 | 11.76 | 2.21 | |

| Condition ID | Min | Max | Average | Std. Dev. | Expression Values |
|---|---|---|---|---|---|
| Black-0_S6_L001_R2_001.fastq.gz_reads_expression | 0.00 | 11.43 | 1.49 | 3.80 | |
| Black-10_S7_L001_R1_001.fastq.gz_reads_expression | 0.00 | 0.00 | 0.00 | 0.00 | |
| Black-30_S8_L001_R1_001.fastq.gz_reads_expression | 0.00 | 11.67 | 0.83 | 3.12 | |
| Black-40_S9_L001_R1_001.fastq.gz_reads_expression | 0.00 | 11.39 | 0.81 | 3.04 | |
| Black-60_S10_L001_R2_001.fastq.gz_reads_expression | 0.00 | 12.55 | 2.28 | 4.60 | |
| Blue-0_S1_L001_R1_001.fastq.gz_reads_expression | 0.00 | 12.04 | 7.93 | 4.45 | |
| Blue-10_S2_L001_R2_001.fastq.gz_reads_expression | 0.00 | 12.84 | 1.64 | 4.21 | |
| Blue-30_S3_L001_R2_001.fastq.gz_reads_expression | 0.00 | 11.62 | 2.27 | 4.54 | |
| Blue-40_S4_L001_R1_001.fastq.gz_reads_expression | 0.00 | 14.43 | 1.03 | 3.86 | |
| Blue-60_S5_L001_R2_001.fastq.gz_reads_expression | 0.00 | 7.96 | 1.62 | 3.23 | |

Figure 6: Heat Map Clustering Results This figure shows the heat map clustering result of the k-means clustering algorithm. The blue represents that the specific gene is expressed in that sample. There are two heat maps, one for each of the 2 clusters. Figure 7 shows the genes in order of the heat map and gives the function.

**Figure 7**

| Cluster 1 Feature ID | Function |
|---|---|
| DGGOJLHL_00185 | ATP synthase subunit alpha |
| DGGOJLHL_00322_rRNA_7 | No Known Function |
| DGGOJLHL_05300_tmRNA_1 | No Known Function |
| DGGOJLHL_05524 | 50S ribosomal protein L22 |

| Cluster 2 Feature ID | Function |
|---|---|
| DGGOJLHL_00321_rRNA_6 | No Known Function |
| DGGOJLHL_00681 | Bifunctional protein PutA |
| DGGOJLHL_00762 | 30S ribosomal protein S6 |
| DGGOJLHL_00907_rRNA_10 | No Known Function |
| DGGOJLHL_01006 | 30S ribosomal protein S20 |
| DGGOJLHL_02077 | Hypothetical Protein |
| DGGOJLHL_02103 | Glutamate-pyruvate aminotransferase AlaA |
| DGGOJLHL_02202 | 3-hydroxydecanoyl-[acyl-carrier-protein] dehydratase |
| DGGOJLHL_03846 | Isocitrate dehydrogenase [NADP] |
| DGGOJLHL_05229_rRNA_14 | No Known Function |
| DGGOJLHL_05527 | 50S ribosomal protein L23 |
| DGGOJLHL_05532 | Elongation factor G 1 |
| DGGOJLHL_05533 | 30S ribosomal protein S7 |
| DGGOJLHL_05550_rRNA_17 | No Known Function |

**Figure 7: Cluster Features and Functions** This figure shows the feature ID of each gene and its function from the heat map results. The features are shown in order with their corresponding location in Figure 6.

**Figure 8**

| **Phage Proteins Expressed in Expression Matrix** |
|---|
| 27 hypothetical |
| DNA packaging protein A, T7-like gp18 |
| Phage capsid assembly scaffolding protein Gp9 |
| Phage collar, head-to-tail connector protein Gp8 |
| Phage DNA ejectosome component Gp16, peptidoglycan lytic exotransglycosylase (EC 4.2.2.n1) |
| Phage DNA ejectosome component, internal virion protein Gp15 |
| Phage DNA-directed DNA polymerase (EC 2.7.7.7) |
| Phage DNA-directed RNA polymerase (EC 2.7.7.6) |
| Phage endolysin |
| Phage endonuclease I (EC 3.1.21.2), four-way DNA junctions resolving |
| Phage exonuclease (ACLAME 274) |
| Phage internal (core) protein |
| Phage major capsid protein Gp10A |
| Phage non-contractile tail fiber protein Gp17 |
| Phage non-contractile tail tubular protein Gp11 |
| Phage non-contractile tail tubular protein Gp12 |
| Phage primase/helicase protein Gp4A |
| 2 phage protein |
| Phage protein Gp5.7 |
| Phage single-stranded DNA-binding protein Gp2.5 |
| 2 Phage tail fiber protein / T7-like tail tubular protein A |
| Phage terminase large subunit Gp19, DNA packaging |

**Figure 8: The phage proteins expressed when aligned to the phage genome** This figure shows the proteins expressed when the RNA-seq sample set is aligned and assembled using the phage genome.

## Discussion

During the infection process, a MOI ratio of 10:1 was conducted as an original infection basis. However, upon infection, results showed that there was not enough virus present in the culture to infect the cells (Figure 9). Upon obtaining these results, a conclusion was drawn that the MOI ratio would need to change in order for a successful infection of *Pseudomonas fluorescens* with GC to occur. A successful infection results in the growth curve graph to be bell-shaped for the infected culture (Figures 3 & 4), while the control remains exponential growth. A possible reason of GC not being able to successfully infect *Pseudomonas fluorescens* is that GC is a weaker virus. For a successful infection to occur, there would need to be many more viral particles in the culture for the bacteria to be feeling overwhelmed during the infection. In addition to a change in the MOI ratio, a conclusion was made that infection would need to occur at the earlier end of early log phase. Once the bacteria are established in log phase, a successful infection to occur is more difficult to obtain because of the exponential growth of the bacteria during log phase. Infecting during the exponential growth of log phase would give the virus a harder environment to attack successfully since the number of bacteria is growing rapidly as the virus is trying to infect.

Even though the Qubit Fluorometer gave results that showed RNA in the samples from the run on September 27, there must not have been mRNA in the sample to produce sequencing results. A possible reason that there was not mRNA in the sample was that the sample was left out at room temperature too long or an issue in the extraction process. In order to maintain mRNA integrity until extraction, it is important to keep the pelleted sample on ice or in a freezer until it is ready to be used in a future experiment since mRNA is highly degradable at temperatures higher than -20°C (Fabre et al., 2014). A way to confirm this would be to run an

agarose gel to confirm the presence of mRNA. Gel electrophoresis was done on the October 27 run, which confirmed the presence of mRNA (Figure 10).

The pipeline produced results that are similar to what should be going on in the cell during infection. The virus would come in and attack the host and begin to make their own proteins in the host, which is why there are many translational related genes showing in the results for the infected sample (blue). However, the statistical analysis done by DESEq2 is beneficial to understand the significance of the results, it would be difficult to understand true significance unless multiple samples of the same time point were collected and analyzed. The results from aligning and assembly using the phage genome are what was expected with the proteins to help bind and take over the host are being expressed as well as the bulk being hypothetical proteins.

Future applications of this project could be collecting multiple samples at each time point to retrieve better statistical analysis to confirm results. Collecting multiple samples at each time point would allow for a more complete statistical analysis to help confirm the actual presence of the proteins during that time point in the infection. The developed pipeline is useful in identifying genes expressed from each time point sample to have a better understanding of the infection process of a bacteriophage and its host.

**Figure 9**

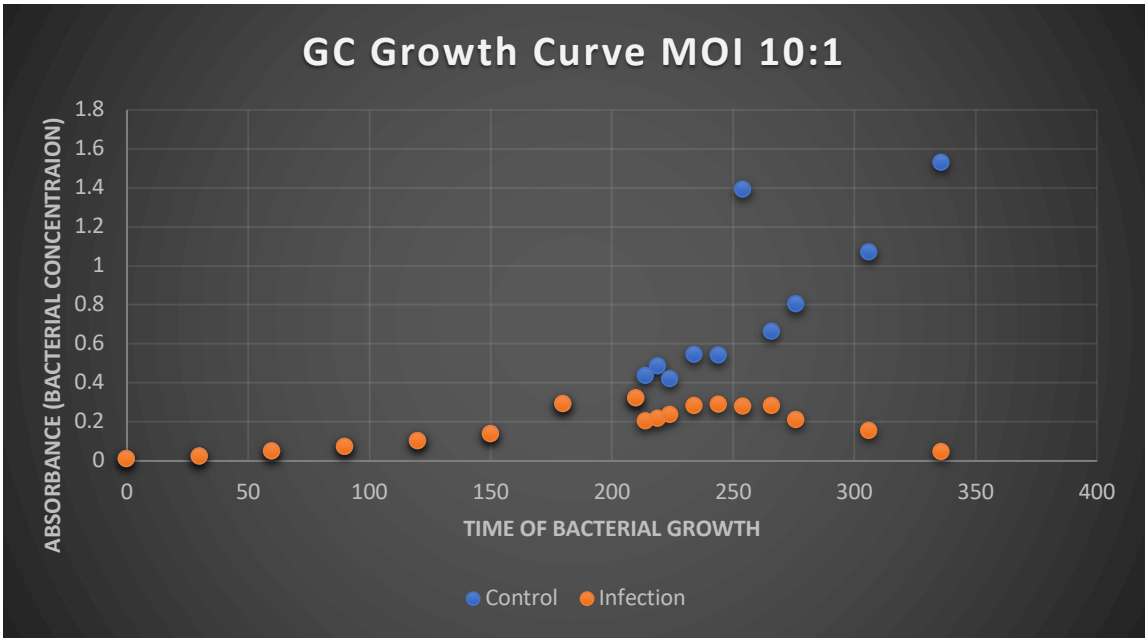


**Figure 9: Growth Curve of GC with MOI of 10:1** Growth curve of GC at MOI of 10:1 shows that there was not enough viral particles present to perform a successful infection. When a successful infection occurs, the infected bacterial concentration is supposed to die off, leaving a bell-shaped graph.

**Figure 10**

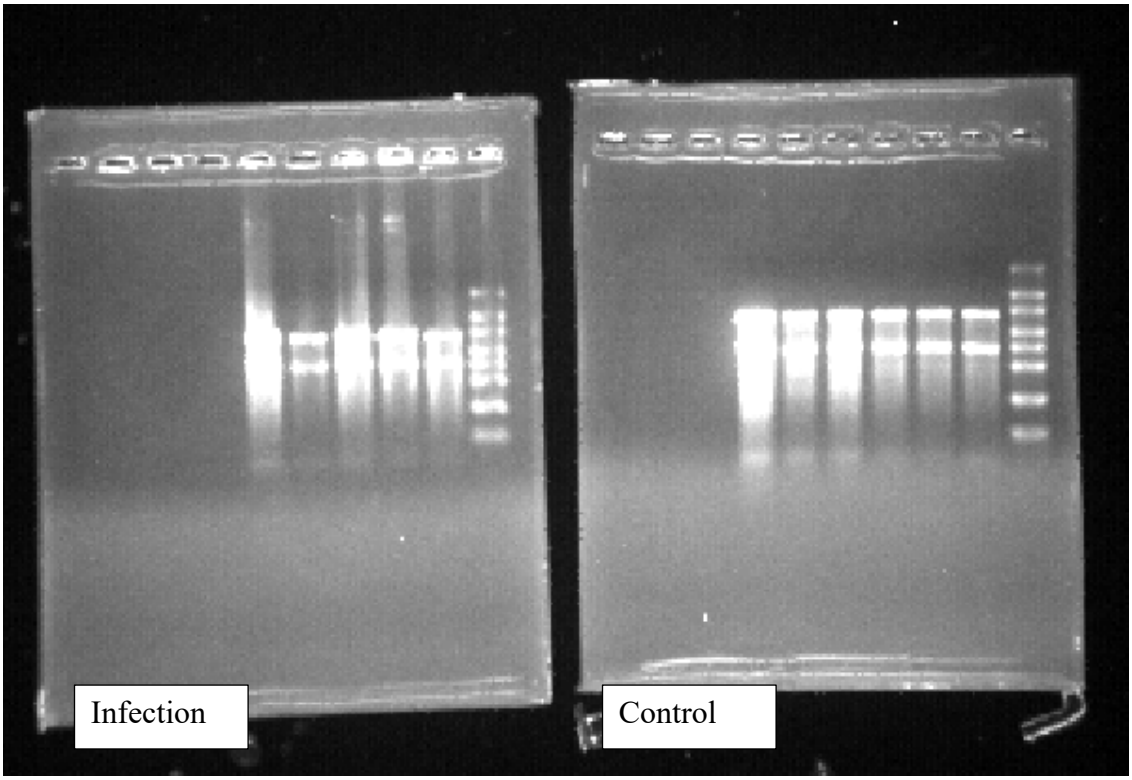


**Figure 10: Gel Electrophoresis Run of GC Samples.** This figure shows the agarose gels of the experiment conducted on October 27, confirming the presence of mRNA in the sample. The gel on the left are the infection samples and the gel on the right are the control samples. The smearing of the gel provides an insight to there being mRNA in the sample.

## References

Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., Dehal, P., Ware, D., Perez, F., Canon, S., Sneddon, M. W., Henderson, M. L., Riehl, W. J., Murphy-Olson, D., Chan, S. Y., Kamimura, R. T., Kumari, S., Drake, M. M., Brettin, T. S., … Yu, D. (2018). KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nature Biotechnology*, *36*(7), 566–569. https://doi.org/10.1038/nbt.4163

*Bacteriophage / phage | Learn Science at Scitable*. (n.d.). Retrieved November 10, 2021, from https://www.nature.com/scitable/definition/bacteriophage-phage-293/

Clokie, M. R., Millard, A. D., Letarov, A. V., & Heaphy, S. (2011). Phages in nature. *Bacteriophage*, *1*(1), 31–45. https://doi.org/10.4161/bact.1.1.14942

Fabre, A.-L., Colotte, M., Luis, A., Tuffet, S., & Bonnet, J. (2014). An efficient method for long-term room temperature storage of RNA. *European Journal of Human Genetics*, *22*(3), 379–385. https://doi.org/10.1038/ejhg.2013.145

Fokine, A., & Rossmann, M. (2014). Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage*. https://doi.org/10.4161/bact.28281

*Gene Expression | Learn Science at Scitable*. (n.d.). Retrieved November 10, 2021, from https://www.nature.com/scitable/topicpage/gene-expression-14121669/

Kukurba, K. R., & Montgomery, S. B. (2015). RNA Sequencing and Analysis. *Cold Spring Harbor Protocols*, *2015*(11), 951–969. https://doi.org/10.1101/pdb.top084970

Maniloff, J., & Ackermann, H. W. (1998). Taxonomy of bacterial viruses: Establishment of tailed virus genera and the order Caudovirales. *Archives of Virology*, *143*(10), 2051–2063. https://doi.org/10.1007/s007050050442

Scales, B. S., Dickson, R. P., LiPuma, J. J., & Huffnagle, G. B. (2014). Microbiology, genomics, and clinical significance of the Pseudomonas fluorescens species complex, an unappreciated colonizer of humans. *Clinical Microbiology Reviews*, *27*(4), 927–948. https://doi.org/10.1128/CMR.00044-14

*Understanding Viral Gene Expression*. (2017, August 15). UMass Chan Medical School. https://www.umassmed.edu/korostelev-lab/research/understanding-viral-gene-expression/

*Virus | Learn Science at Scitable*. (n.d.). Retrieved November 10, 2021, from https://www.nature.com/scitable/definition/virus-308/