2024-03-13

RAPID DISCOVERY OF HOUSE-KEEPING GENES IN THE BACTERIUM KUSHNERIA, USING A SLALOM-LIBRARY-BASED CRISPRI FORWARD GENETIC SCREEN

Andrew Jenkins
Brigham Young University - Provo

Honors Thesis

RAPID DISCOVERY OF HOUSE-KEEPING GENES IN THE BACTERIUM KUSHNERIA, USING A SLALOM-LIBRARY-BASED CRISPRI FORWARD GENETIC SCREEN

by

Andrew Jenkins

Submitted to Brigham Young University in partial fulfillment

of graduation requirements for University Honors

Cell Biology and Physiology Department

Brigham Young University

April 2024

Advisor: Dr. Jonathon Hill

Honors Coordinator: Dr. Roy Silcox

Faculty Reader: Dr. Brent Nielsen

ABSTRACT

RAPID DISCOVERY OF HOUSE-KEEPING GENES IN THE BACTERIUM KUSHNERIA, USING A SLALOM-LIBRARY-BASED CRISPRI FORWARD GENETIC SCREEN

Andrew Jenkins

Department of Cell Biology and Physiology

Bachelor of Science

Rising soil salinity levels worldwide have resulted in the loss of millions of hectares of viable cropland. Several solutions to this problem have been posited, including looking into the properties of salt-tolerant plants. Recent research into salt-tolerant plants in southern Utah has found a host of microorganisms isolated from the soil, plant surfaces, or plant tissues that stimulate salt tolerance in alfalfa. Leveraging a Mobile-CRISPRi system to repress a library of genes created through the recently developed sgRNA Library Assembly by Ligation onto Magnetic Beads (SLALOM) protocol, we induced genome-wide repression of genes in *Kushneria* spp. grown under optimal and high salt selective conditions. Unfortunately, due to the inability of the CRISPRi system to function in the high-salt environment, no salt-tolerance genes were discovered. The optimal salt condition revealed eight housekeeping genes and demonstrated this method requires further refinement to gain more insights into *Kushneria* spp. salt tolerance.

ACKNOWLEDGMENTS

I am forever grateful for Dr. Jonathon Hill. He has inspired me to think deeply about what it means to use our creativity to fearlessly explore ideas and test them, sometimes against our better judgment. Thanks to Dr. Brent Nielsen for his identification and procurement of the bacteria in which we tested our new methodology. I would also like to thank Joshua Yates for his help with the SLALOM procedure and for his patient assistance in outlining the experiments detailed throughout this thesis. This study would not be possible without the support of the Honors Program and the BYU College of Life Sciences, through which grant funds made these experiments possible.

TABLE OF CONTENTS

LIST OF FIGURES

Introduction

There is a global problem of increasing agricultural soil salinity. The USDA estimates that high salinity causes 10 million hectares of soil to be lost worldwide yearly. This increase reduces yields in many standard crops such as rice, cotton, and wheat (1, 2). It is vital for a viable solution to either leech the soil of these high salt levels or modify crops to better survive these conditions. Finding salt-tolerant crop alternatives has been a current focus of study in this field, including the study of what drives salt tolerance in these plant alternatives (3). Many salt-tolerant plants host halophilic microorganisms that stimulate their plant host, often strengthening the plants halotolerance (4). Discovering the mechanisms these microorganisms use to confer salt tolerance could allow millions of hectares of land to be usable again through inoculation with halophilic bacteria.

Recent work has discovered halophilic microbes associated with three halophyte species near Goshen, UT, that can stimulate alfalfa growth when used as an inoculum in high salinity soil (1). One of these species, *Kushneria* spp., is a gram-negative halophile and still needs a large body of research associated with its function. It is closely related to the *Halomonas* species in the *Halomonadaceae* family, which are well-documented halophiles (5). As a halophile with potential use as a high-salt soil probiotic, we became interested in the genetic pathways that allow *Kushneria* to survive in high-salt environments. Recent developments in forward genetic screening have allowed us to develop a novel method that can query an entire bacterial genome for genes of interest for relatively low cost and standard laboratory techniques.

Our forward screen leverages the Mobile-CRISPRi (Mobile-Clustered Regularly Interspaced Palindromic Repeats interference) system as described by Peters et al. This system interrogated gene functionality by knocking out bacterial gene expression under selective conditions via chemical induction (6, 7). Mobile-CRISPRi takes advantage of an engineered version of the bacterial immune system machinery found in *Streptococcus pyogenes* to create an IPTG (Isopropyl ß-D-1-thiogalactopyranoside) inducible complex consisting of a catalytically dead Cas9 protein and a single guide RNA (sgRNA). This dCas9-sgRNA complex binds to a section of DNA complementary to the approximately 20 bp target nucleotide sequence in the sgRNA, blocking transcription elongation and disrupting the gene of interest (8, 9). Mobile-CRISPRi system leverages a Tn7 transcription system for inserting a kanamycin resistance gene, the sgRNA of interest, and a dCas9 gene all into the bacterium's genome at an $_{att}$Tn7 site, with the latter two segments, the sgRNA and the dCas9, under *lac* operon control (10, 11).

Applying this machinery to a genetic screen requires a representative sample of sgRNAs that repress across the entire genome of interest (12). This has typically been a high-cost process requiring the commercial construction of each sgRNA, the sequence of which would need to be known before its synthesis (13). However, a recently developed enzymatic approach to generating sgRNAs, called **s**gRNA **L**ibrary **A**ssembly by **L**igation **O**nto **M**agnetic Beads (SLALOM), significantly reduces the time and cost associated with generating a functioning sgRNA library. The SLALOM kit utilizes the HpaII enzyme cut site to fragment input DNA between the two cytosines in its palindromic recognition sequence, CCGG, on both the forward and reverse strands. This effectively prepares each fragmented strand to be located at one of the four possible CRISPR PAM

sequences, CGG, meaning the library cannot wholly represent every possible sgRNA (14). However, in our preliminary search through the *Kushneria* genome, we found that

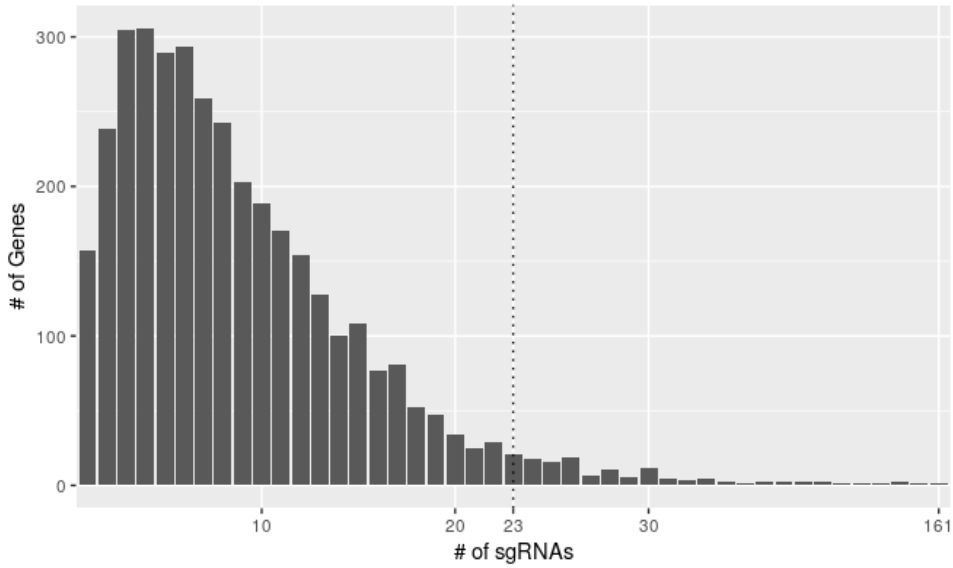


Figure 1: Predicted Guides to Genes. *Analysis of the* Kushneria *genome showing the number of theoretical sgRNAs / genes based on the HpaII cut site in protein coding regions of the genome. The median is represented in the dashed line at approximately 23 sgRNAs.*

of the 3200 gene-encoding regions, there are 28,000 HpaII enzyme cut sites. This results in an expected library size of 64,000 sgRNAs, with an average coverage of approximately 23 sgRNA per coding gene (Figure 1).

While not all predicted guides could be captured through the kit generation process, we generated an sgRNA library that sufficiently represents the *Kushneria* genome. These library segments were inserted at *Kushneria*'s $_{att}$Tn7 site using the Mobile-CRISPRi system, producing a library-bearing strain of *Kushneria*. We subjected this strain to optimal salt and high salt conditions, both with and without induction of the dCas9-sgRNA complex. After allowing these samples to grow through 16-20 generations, we extracted the genomes from each population. We PCR amplified the sgRNA area of interest, sequenced these fragments via Illumina NGS, and then aligned

the sgRNA portion to the *Kushneria* genome. These alignments could generate feature count data and identify differentially expressed genes using DESeq2 (15). While we could not identify any genes associated with salt tolerance in *Kushneria*, we discovered that high salt conditions disrupt the Mobile-CRISPRi system, removing its ability to knock down gene expression properly using dCas9. With our data, we could compare the populations grown under optimum salt concentrations and find eight potential Kushneria housekeeping genes. However, we expect to increase this amount substantially in future studies.

Materials and Methods

Optimum Salt Concentration Growth

Kushneria was streaked for single colonies from -80C stock culture onto 2.5% NaCl Luria Broth (LB) agar plates with 50 ug/ml streptomycin at 30C overnight. A colony was picked and grown in 2.5% NaCl LB broth with 50 ug/ml streptomycin in a 30C incubator at 220 rpm for 18 hours. 1 ml of this culture was then divided into 50 ml of LB broth at varying NaCl percentages (2.5%, 5%, 7%, 8%, 9%, 10%, 11%, 12%, 15%, and 17%). The growth of these cultures was tracked using a spectrophotometer, taking optical density readings at 600 nm (OD600). The growth was tracked across 32 hours. *Kushneria* doubling times were calculated by graphing the OD600 readings for each NaCl concentration.

Kushneria Doubling Time at Optimal Salt Concentration

As determined above, *Kushneria* was grown at the optimal salt concentration of 5%. The growth rate was determined by taking optical density readings on a spectrophotometer every hour for 24 hours, interspersed serial dilution, and plating of the growth at different growth phases. The standard growth rate was then determined by plotting these points and finding the doubling time of the culture.

Kushneria SLALOM DNA Library Synthesis

The original *Kushneria* strain was grown as mentioned above but at the new optimal salt concentration of 5% to the mid-exponential phase. Part of the culture was extracted and run through the Qiagen DNeasy Power Lyzer Kit for genomic DNA isolation. 2 ug of the extracted genomic DNA was then used as the input DNA for

sgRNA library synthesis via the SLALOM 1.0 kit. Successful library synthesis was verified via gel electrophoresis, with a clear band at approximately 160 bp.

Library Plasmid Growth, Digestion, and Electroporation.

The plasmids pJMP1039, pJMP2834, and pJMP2846, developed by Peters et al., were ordered from Addgene. Cultures of single colony isolates from these plasmid-bearing strains were grown overnight in average LB with 50 ug/ml carbenicillin in a 37C incubator at 220 rpm. Each plasmid was extracted from the culture using the Qiagen Midi Prep Kit (Qiagen). The library vector, pJMP2846, was digested with BsaII (New England Biolabs) overnight at 16C. This digestion was run on a 1% agarose gel, and the digested band was isolated via gel extraction using the Monarch Gel Extraction Kit (New England Biolabs), after which the sgRNA library was inserted using T4 Ligase (New England Biolabs).

250 ml of *Kushneria* was grown to exponential phase and made electrocompetent via multiple centrifugations and resuspensions in 10% glycerol, then stored on ice in 50 ul aliquots for immediate use. To generate a green fluorescent protein (GFP) encoding strain of *Kushneria,* electrocompetent *Kushneria* cells were electroporated in a 1 mm electroporation cuvette with 100 ng of pJMP1039 and 100 ng of pJMP2834. After electroporation, the cells were incubated at 30C and 220 rpm for 2 hours to allow for recovery, expression of Tn7, and transposition of the kanamycin resistance, *lacI,* GFP, inducible dCas9, and inducible GFP-targeting-sgRNA at the bacterial $_{att}$Tn7 site. This culture was then grown overnight in 5% NaCl growth media with 30 ug/ml kanamycin for selection. This method was repeated six times with 100 ng of the plasmid library to generate the library-bearing *Kushneria* strain.

GFP Knockdown

A GFP repression assay of the GFP-bearing strain was conducted to determine whether the new GFP and library-bearing strains of Kushneria responded adequately to induction via IPTG. Cultures of *Kushneria* with GFP were grown at 5% NaCl, as mentioned above, to the mid-exponential phase. A 1 ml aliquot was added to 5 ml of LB broth, either with or without 1 mM IPTG. These were allowed to grow for about 12 hours. 1 ml of each culture was extracted, centrifuged, and washed thrice in phosphate-buffered saline (PBS). The final wash left the cells resuspended in 500 ul of PBS, then divided into 3, 100 ul aliquots on a dark, clear-bottomed 96-well plate. These were then measured by a spectrophotometer at 509 nm alongside a PBS-only negative control. Using the same method, we repeated this test after the selection screen and sequencing with the 12% NaCl condition.

Salt Selection Screen and Data Analysis

The forward genetic screen for salt tolerance genes was conducted over two days. 50 ml of library-bearing Kushneria culture was prepared for three replicates of four different groups: 5% NaCl without 1 mM IPTG, 5 % NaCl with 1 mM IPTG, 12% NaCl without 1 mM IPTG, and 12% with IPTG. 1.5 ml of frozen stock library-bearing *Kushneria* was thawed on ice and added to each prepared broth. These cultures were grown in a 30C incubator at 220 rpm for approximately ten generations. Then, 1 ml was extracted from each culture, added to an additional 50 ml of broth, and grown for another ten generations.

3 ml of each culture was extracted and run through the Qiagen DNeasy Power Lyzer Kit for genomic DNA isolation, after which the library cassette was extracted from each replicate via PCR amplification. The PCR was conducted using Q5® High-Fidelity DNA Polymerase (NEB) under standard conditions from the NEB website. The PCR amplicons were then sequenced via 150 bp paired-end Illumina sequencing. The unique library portion of each read was then extracted using the *cutadapt* Python 3.0 package and then aligned to the *Kushneria* genome via *Rsubread*. These alignments were piped through DESeq2 to normalize the counts and analyze which genes were upregulated or downregulated between groups.

Results

The mechanism by which *Kushneria* increases the salt tolerance of inoculated plants remains unknown. Using various metagenomic methods to query microbial genomes for consequential pathways has proven helpful in discovering genes with novel functions. Screens with extensive libraries of gene knockouts are standard but can often take a significant amount of time and resources. These strategies involve crafting sgRNA guides by hand and KO strains for each gene of interest. This severely limits the number of genes interrogated in a single screen. With recent advancements in forward genetic screening methods, we developed a novel CRISPRi-based screen that can test the entire Kushneria genome for salt tolerance genes on a single screen.

To better understand *Kushneria*'s optimal growth rate under various salt concentrations, we tracked *Kushneria* growth in culture using OD600 values tracked over 34 hours (Figure 2A). We found that *Kushneria* had an optimal growth rate at 5% NaCl and high salt selective pressure growth at 12% NaCl, where the growth rate was reduced by one-half in the exponential phase. From this data, we grew any optimal salt concentration culture at 5% NaCl and any high salt selective screen culture at 12% NaCl. Further optimization is required to determine the precise doubling time for our bacteria at 5% NaCl, as the Mobile-CRISPRi system requires 16-20 generations to maximize the signal-to-noise ratio. We used an overnight culture of *Kushneria* at 5% NaCl to determine its doubling time to be approximately 1.5 hours (Figure 2B).

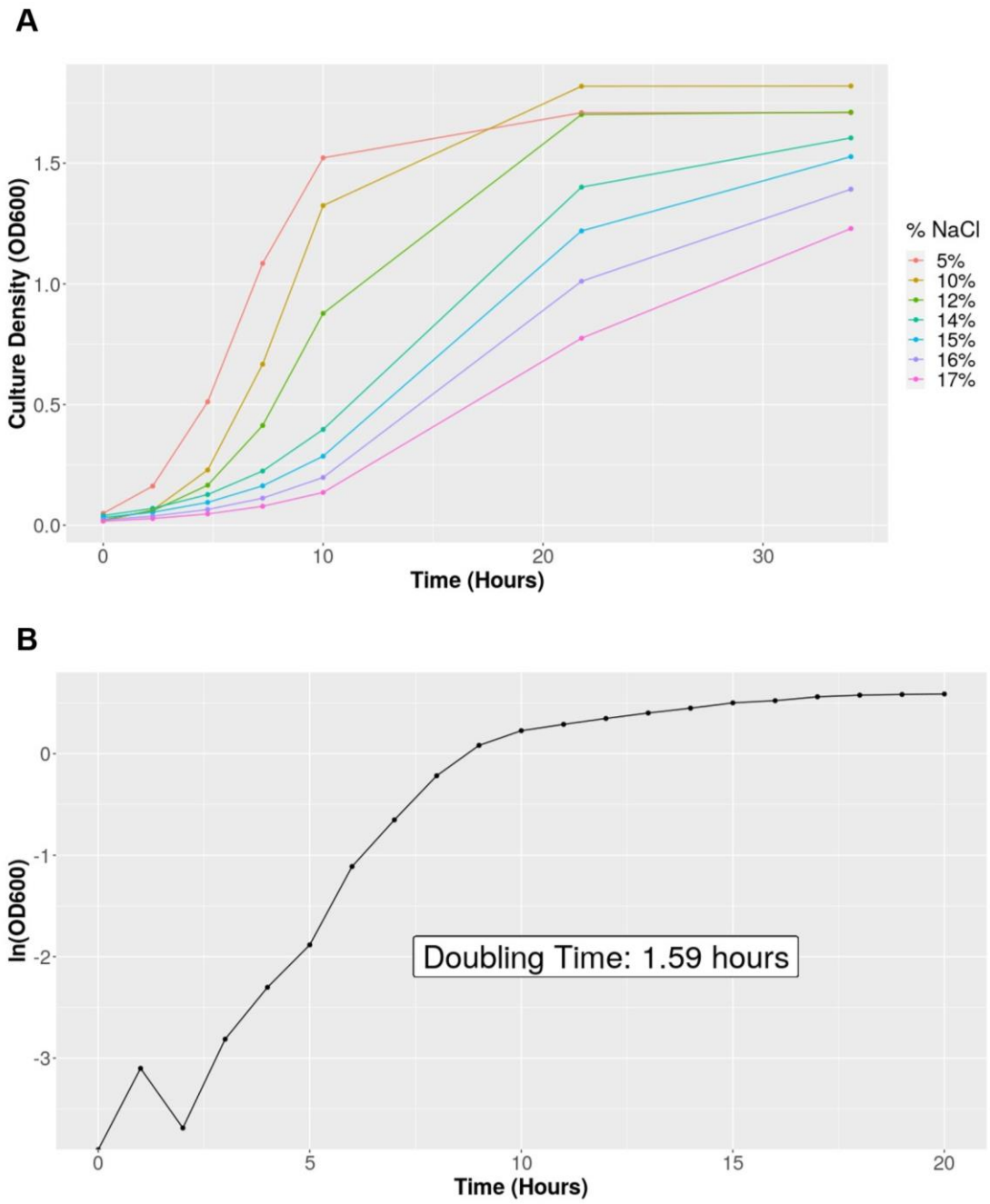


Figure 2: Kushneria Growth Optimization. *A) Optimal growth was 5%, and everything 12% and above showed slow growth and is assumed to be selective. B) The ln(OD600) is plotted here to show where the growth rate is linear, which was used to find the doubling time of 1.59 hours.*

We next determined whether the programmable Mobile-CRISPRi system in *Kushneria* would be effective. We generated a GFP-expressing *Kushneria* strain using the Tn7 transposition system and the GFP-bearing plasmid, pJMP2834, described by Banta et al. (7). After successful transformation, we conducted a test of the knockdown efficiency of *Kushneria* upon IPTG induction of the dCas-9 and the GFP-targeting-sgRNA. We initially tested the knockdown efficiency at 5% NaCl only. We achieved a

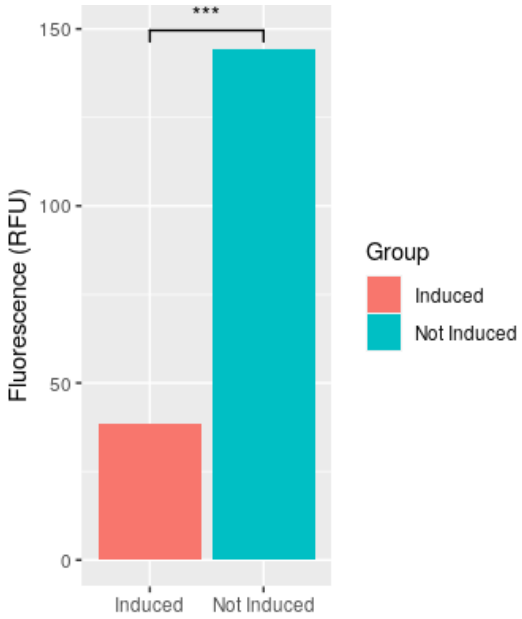


Figure 3: Initial Mobile-CRISPRi GFP Repression. *Induction repression test of GFP expressing* Kushneria *with a dCas9 and sgRNA targeting the GFP coding gene. GFP shows 4x repression in the induced vs the not induced control.*

4x knockdown of GFP expression in the induced bacteria compared to GFP-expressing controls with a p-value < 0.001 (Figure 3). The magnitude of the knockdown was lower than expected when compared to what had been described by Banta et al. in other bacterial species (8x to 40x), including *E. coli* and *S. aeruginosa*. We proceeded with the screen, determining that this level of repression would be sufficient to see adequate reductions in gene counts over 20 generations in the final screen.

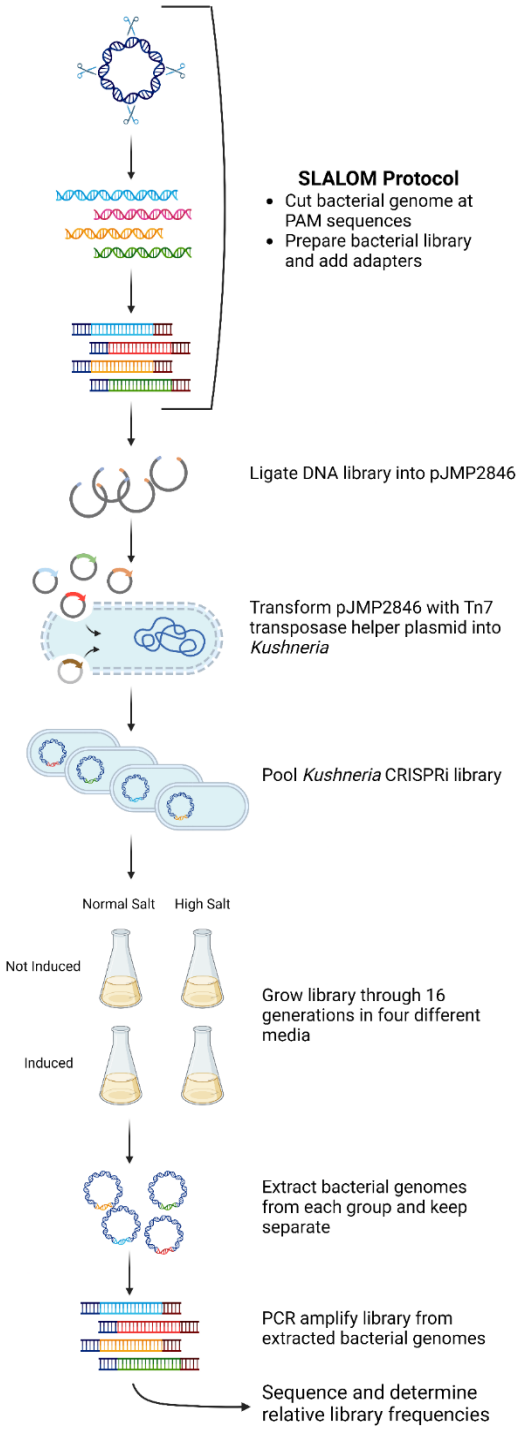


Figure 4: Bacterial Forward Genetic Screening Method. *Mobile-CRISPRi Based Method for* Kushneria *Screen. Graphic depicting our screening method starting from genomic DNA and ending with PCR amplification followed by sequencing.*

Having determined the optimal salt conditions and understanding that Mobile-CRISPRi was effective in *Kushneria*, we tested our novel forward-screening method

outlined in Figure 4. We grew a culture to its exponential phase and then extracted the *Kushneria* genome, which we used as input for the SLALOM library kit. We successfully generated a library of genomic sgRNA-coding DNA fragments, which we will use moving forward. These fragments were ligated into the modular vector pJMP2846 at the spacer sequence site. The complete library constructs were then electroporated with the Tn7 transposase helper plasmid, pJMP1039, into electrocompetent *Kushneria* cells. After a prolonged recovery to allow for transposition of the library sequence into the bacterial $_{att}$Tn7 site, aliquots of the cells were plated on kanamycin selection plates to determine the transformation efficiency. At the same time, the remaining sample was started in a kanamycin selection culture.

For the genetic screen, we thawed twelve *Kushneria* library-bearing stock cultures under four conditions: optimal salt and high salt, both with and without IPTG induction. After allowing these groups to grow to 20 generations, we extracted the genomic DNA from these groups and PCR amplified the sgRNA regions. Unfortunately, we could not capture paired reads with our amplification, primarily due to the high AT content surrounding these regions. We were forced to use primers that amplified a 500 bp fragment with the sgRNA located in the forward region. Illumina is only accurate for paired-end reads of less than 150 bp, meaning our inserted sgRNA of interest was only found in single-end reads when sequenced with their NGS platform (16). Despite this drawback, we used the Python *cutadapt* package (17) to trim our single-end adapters and align these trimmed sequences back to the *Kushneria* genome via *Rsubread* (18). Approximately 25% of these sequences aligned back to the *Kushneria* genome, aligning with over 4,000 genes of the 4,300 protein-coding sequences.

Using the *featureCounts* package and a recently annotated version of the *Kushneria* genome, we assigned these aligned reads to the *Kushneria* features (19). The counts were then passed into *DESeq2* for differential gene expression analysis. We then used this data to make a PCA plot (Figure 5A), average the replicate gene counts from each condition, calculate z-scores, and plot a sample of these genes on a heatmap (Figure 5B). We were surprised to find slight variations between each condition, with few genes showing differentiation.

Given these results, we reattempted the GFP knockdown efficiency experiments, making sure to grow the *Kushneria* in 5% and 12% NaCl. To better visualize these cells, we plated them on 5% and 12% NaCl plates with IPTG and without IPTG (Figure 6A). We measured the relative fluorescence of these plates using ImageJ2 (20). We showed that the cells grown at 12% NaCl demonstrated no significant reduction in fluorescence between induced and uninduced groups (Figure 6B). This suggests that the high salt environment somehow disables the Mobile-CRISPRi system, making it ineffective at reducing gene expression in its target region.

We completed our analysis by using our *DESeq2* data to make differential gene expression comparisons between the 5% induced and uninduced and the 12% induced and uninduced groups. Despite the noise noted earlier in Figure 5, we could see both down-regulated and up-regulated genes in the 5% comparison (Figure 7A) and only a few up-regulated genes in the 12% (Figure 7B). We have listed the down-regulated genes from the 5% comparison (Figure 7C), which we hypothesize are *Kushneria* housekeeping genes, and we hypothesize that knocking them out reduced their population after 20 generations of culturing.

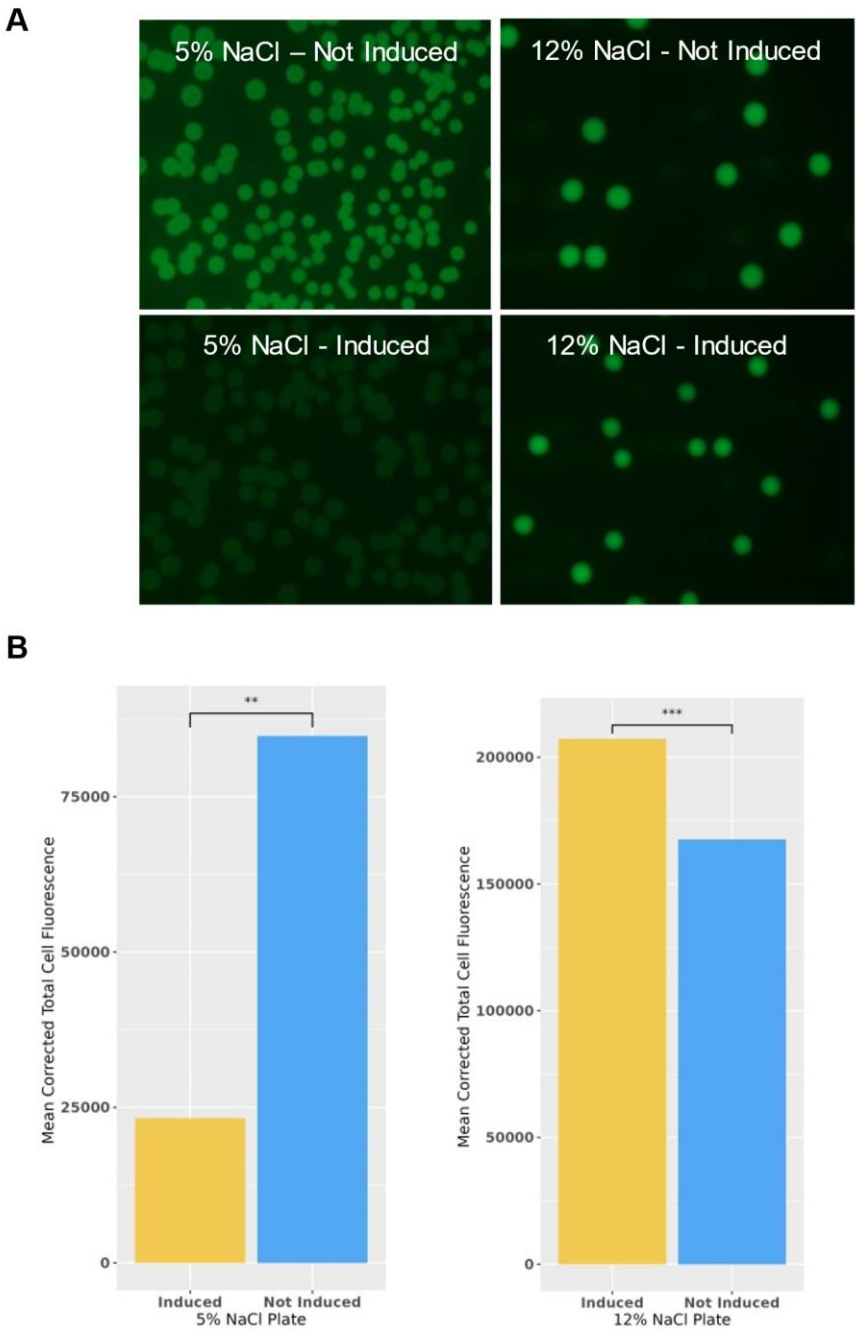
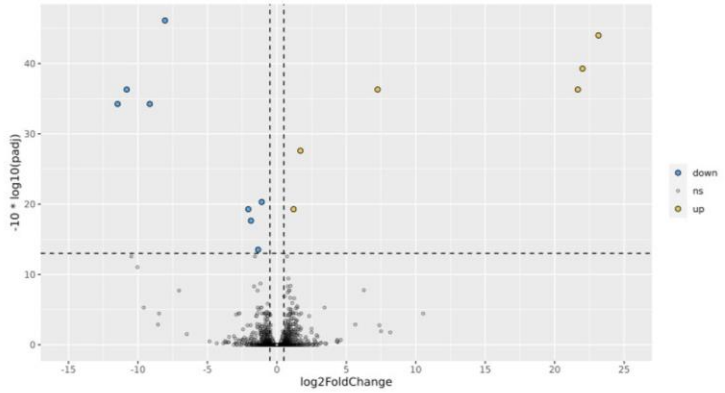
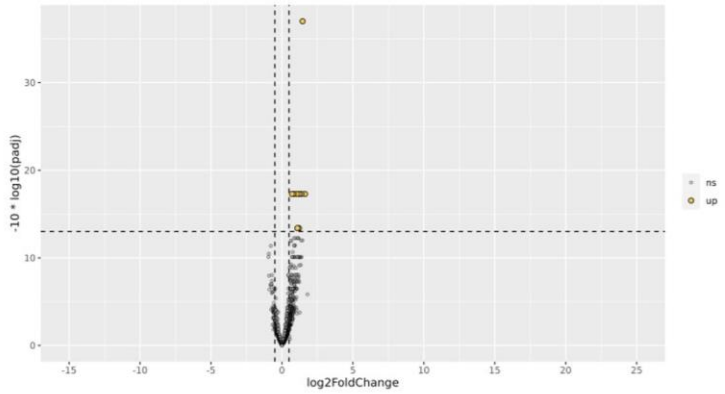


Figure 5: Post Screen Mobile-CRISPRi GFP Repression *A) Photos of GFP expressing Kushneria grown under various conditions. B) GFP fluorescence measured via ImageJ and calculated using the mean corrected total cell fluorescence between each induced and uninduced condition*

A



B



C

Gene	Protein	Description
CGEJKNPP_02396	sucC_1	Succinate--CoA ligase [ADP-forming] subunit beta
CGEJKNPP_09278	polX_3	DNA polymerase/3'-5' exonuclease
CGEJKNPP_07917	dgcE_2	putative diguanylate cyclase
CGEJKNPP_10359	ugpA	sn-glycerol-3-phosphate transport system permease protein
CGEJKNPP_07649	**UNK**	related to a purine nucleoside permease
CGEJKNPP_08932	MotA	chemotaxis protein MotA
CGEJKNPP_09591	**UNK**	related to a thiol:disulfide interchange protein DsbC
CGEJKNPP_10551	dctP_3	Solute-binding protein

Figure 6: DEG's and Potential Housekeeping Genes *A) Volcano plot of differentially expressed genes of 5% NaCl uninduced vs. induced B) Volcano plot of differentially expressed genes of 12% NaCl uninduced vs. induced C) Potential housekeeping genes, down-regulated in the 5% NaCl uninduced vs. induced gene expression plot*

Discussion

This research aimed to discover salt tolerance genes in the bacteria *Kushneria*. This study found that the *Kushneria* grows optimally at a NaCl concentration of 5%, with a reduced doubling time but reliable growth at a NaCl concentration of 12%. Our research showed that *Kushneria* receives transposable elements from the Mobile-CRISPRi system and that these elements, including dCas9 and a targeting sgRNA, can be expressed via *lac* operon induction. According to our results, we successfully generated a comprehensive sgRNA library from the *Kushneria* genome. However, we could not prove that this library effectively represses gene expression under extreme salt conditions.

In a GFP-expressing strain of *Kushneria*, we demonstrated that the Mobile-CRISPRi system can repress GFP, albeit not at the rates seen in previous papers that were attempted in better-studied model bacteria. While repression was possible through induction of our dCas9 and GFP sgRNA, this repression was relatively low at about 4-fold repression, whereas the lowest seen in Peters et al. was 8-fold and the highest was 40-fold. This loss of repression is likely due to the *lac* promoter that comes with the Mobile-CRISPRi plasmid not adequately matching *Kushneria*-specific *lac* promoters and ribosome binding sites (RBSs) (21). Through our GFP repression assay, we determined that even 4-fold repression was enough to see significant population changes over 16-20 generations. In our selective salt scenario, however, we did not see consistent reductions in as many of the genes as we expected, likely due to this promoter mismatch. This optimization will significantly improve the screen quality and reveal new genes not observable through our current method.

The inability of the Mobile-CRISPRi system to function appropriately in bacterial populations undergoing environmental stress, such as our high-salt condition, may be related to *Kushneria*'s stress response. *E. coli* and other halophilic bacteria have been found to demonstrate more antibiotic resistance as NaCl concentrations rise, likely due to increased efflux pump expression (22, 23). *Kushneria*'s salt tolerance is due to an increased capacity to regulate its internal osmolarity, which may mean it increases drug efflux and decreases drug influx as the salt concentration rises. At the 12% NaCl concentration, when *Kushneria* is induced to make dCas9 via IPTG, these cells are likely being outcompeted by populations that may not have retained expression of the recombinant protein.

Despite these limitations, our screening method between induced and uninduced *Kushneria* library-bearing strains grown at 5% NaCl revealed eight potential housekeeping genes that appear to be necessary for *Kushneria* survival. Further validation by knocking out these genes and confirming they are required for *Kushneria* growth will demonstrate that this forward genetic screening method works but needs optimization to be more effective in *Kushneria* and other nonmodel organisms.

Conclusion

While our study did not elucidate any substantial genetic pathways contributing to *Kushneria's* halotolerance, we discovered eight potential housekeeping genes necessary for *Kushneria's* survival. We also found that high salt environments can provoke unexpected responses to inducible protein systems and that great care must be taken in future CRISPRi screens to ensure that environmental selection does not interfere with dCas9's ability to repress at the sgRNA of interest. Future studies will continue to optimize this forward genetic screening method, and its use will enable thousands of nonmodel bacterial organisms to be rapidly screened for valuable traits, like the halotolerance of *Kushneria.* Understanding how this and other microbiota both survive and confer survival to plants in high-salt soils could be the key to resolving problems associated with rising levels of soil salinity throughout the globe.

References

1. Kearl J, McNary C, Lowman JS, Mei C, Aanderud ZT, Smith ST, West J, Colton E, Hamson M, Nielsen BL. 2019. Salt-Tolerant Halophyte Rhizosphere Bacteria Stimulate Growth of Alfalfa in Salty Soil. Front Microbiol 10.

2. Gul B, Ansari R, Ali H, Yousuf M, Weber D, Nielsen B, Koyro H-W, Khan M. 2014. The sustainable utilization of saline resources for livestock feed production in arid and semi-arid regions: A model from Pakistan. Emir J Food Agric https://doi.org/10.9755/ejfa.xxx.xxx.

3. Egea I, Estrada Y, Faura C, Egea-Fernández JM, Bolarin MC, Flores FB. 2023. Salt-tolerant alternative crops as sources of quality food to mitigate the negative impact of salinity on agricultural production. Front Plant Sci 14.

4. Meinzer M, Ahmad N, Nielsen BL. 2023. Halophilic Plant-Associated Bacteria with Plant-Growth-Promoting Potential. Microorganisms 11:2910.

5. Sánchez-Porro C, de la Haba RR, Soto-Ramírez N, Márquez MC, Montalvo-Rodríguez R, Ventosa A. 2009. Description of Kushneria aurantia gen. nov., sp. nov., a novel member of the family Halomonadaceae, and a proposal for reclassification of Halomonas marisflavi as Kushneria marisflavi comb. nov., of Halomonas indalinina as Kushneria indalinina comb. nov. and of Halomonas avicenniae as Kushneria avicenniae comb. nov. Int J Syst Evol Microbiol 59:397–405.

6. Peters JM, Koo B-M, Patino R, Heussler GE, Hearne CC, Qu J, Inclan YF, Hawkins JS, Lu CHS, Silvis MR, Harden MM, Osadnik H, Peters JE, Engel JN, Dutton RJ, Grossman AD, Gross CA, Rosenberg OS. 2019. Enabling genetic analysis of diverse bacteria with Mobile-CRISPRi. 2. Nat Microbiol 4:244–250.

7. Banta AB, Ward RD, Tran JS, Bacon EE, Peters JM. 2020. Programmable Gene Knockdown in Diverse Bacteria Using Mobile-CRISPRi. Curr Protoc Microbiol 59:e130.

8. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. 2013. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell 152:1173–1183.

9. Pellegrino GM, Browne TS, Sharath K, Bari KA, Vancuren SJ, Allen-Vercoe E, Gloor GB, Edgell DR. 2023. Metabolically-targeted dCas9 expression in bacteria. Nucleic Acids Res 51:982–996.

10. Hickman AB, Dyda F. 2016. DNA Transposition at Work. Chem Rev 116:12758–12784.

11. Shkumatov AV, Aryanpour N, Oger CA, Goossens G, Hallet BF, Efremov RG. 2022. Structural insight into Tn3 family transposition mechanism. 1. Nat Commun 13:6155.

12. Todor H, Silvis MR, Osadnik H, Gross CA. 2021. Bacterial CRISPR screens for gene function. Curr Opin Microbiol 59:102–109.

13. Yates JD, Russell RC, Barton NJ, Yost HJ, Hill JT. 2021. A simple and rapid method for enzymatic synthesis of CRISPR-Cas9 sgRNA libraries. Nucleic Acids Res 49:e131.

14. Lane AB, Strzelecka M, Ettinger A, Grenfell AW, Wittmann T, Heald R. 2015. Enzymatically Generated CRISPR Libraries for Genome Labeling and Screening. Dev Cell 34:373–378.

15. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550.

16. Tan G, Opitz L, Schlapbach R, Rehrauer H. 2019. Long fragments achieve lower base quality in Illumina paired-end sequencing. 1. Sci Rep 9:2856.

17. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. 1. EMBnet.journal 17:10–12.

18. Liao Y, Smyth GK, Shi W. 2019. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res 47:e47.

19. Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30:923–930.

20. Rueden CT, Schindelin J, Hiner MC, DeZonia BE, Walter AE, Arena ET, Eliceiri KW. 2017. ImageJ2: ImageJ for the next generation of scientific image data. BMC Bioinformatics 18:529.

21. Shine J, Dalgarno L. 1975. Determinant of cistron specificity in bacterial ribosomes. Nature 254:34–38.

22. Zhu M, Dai X. 2018. High Salt Cross-Protects Escherichia coli from Antibiotic Treatment through Increasing Efflux Pump Expression. mSphere 3:10.1128/msphere.00095-18.

23. Park Y-L, Choi T-R, Kim HJ, Song H-S, Lee HS, Park SL, Lee SM, Kim SH, Park S, Bhatia SK, Gurav R, Sung C, Seo S-O, Yang Y-H. 2021. NaCl Concentration-Dependent Aminoglycoside Resistance of Halomonas socia CKY01 and Identification of Related Genes. J Microbiol Biotechnol 31:250–258.

Appendix I

Python code for trimming each single-end read, aligning it to the Kushneria genome, and assigning these alignments to a Kushneria annotation file.

```
### MERGE FASTQ READS ###
### Because we only care about single end reads and not paired end we merge the
two read files for each sample into a single fastq file

import os
import subprocess

def merge_fastq_files(directory):
    # Get a list of files in the directory
    files = sorted(os.listdir(directory))

    print(files)

    # Filter files to get only those ending with '.fastq'
    fastq_files = [file for file in files if file.endswith('.fastq.gz')]

    # Iterate over file pairs and merge them
    for i in range(0, len(fastq_files), 2):
        file1 = fastq_files[i]
        file2 = fastq_files[i + 1] if i + 1 < len(fastq_files) else None

        if file2:
            merged_file = f"merged_{file1.split('_')[0]}.fastq.gz"

            # Execute the command to merge files
            command = f"cat {directory_path}{file1} {directory_path}{file2} >
                        {directory_path}{merged_file}"

            subprocess.call(["touch", "{directory_path}{merged_file}"],
                            shell=True)
            subprocess.call(command, shell=True)

            print(f"Merged {file1} and {file2} into {merged_file}")
        else:
            print(f"Ignoring single file: {file1}")

# Specify the directory containing your fastq files
directory_path = "../kushneria/data/raw_reads/K_raw_fastq/"

# Call the function to merge fastq files
merge_fastq_files(directory_path)
```

```
### TRIM ALL FASTQ READS ###
### Searches for adapter sequences that surround the sgRNA sequence and saves
them to a new trimmed fastq file
def trim_all_files(directory_path):
    filenames = sorted(os.listdir(directory_path))
    for file in filenames:
        print(file)
        if file.split('_')[0] == "merged":
             command = f"cutadapt -g CTTCTAGT...CGGTTGG -o
             ./data/trimmed_reads/{file.split('.')[0]}_trimmed.fastq.gz
             ./data/raw_reads/K_raw_fastq/{file}"
            subprocess.call(command, shell=True)
trim_all_files("../kushneria/data/raw_reads/K_raw_fastq/")


### Align reads to target genome using subread ###
### Takes in every trimmed read file and aligns to the K. B5 genome

import subprocess
import os

def subread_align(genome, input_file, output_file):
  """Aligns reads to genome
  Args:
    input_file: The path to the fastQ files
    genome: The path to the K.B5 fasta file.
    output_file: The path to the output BAM file.
  """
  filenames = sorted(os.listdir(input_file))
  for file in filenames:
    command = f"subread-align -i {genome} -r {input_file}{file} -o
    {output_file}{file.split('_')[0]}_{file.split('_')[1]}.bam -t 1
    -- multiMapping"
    subprocess.call(command, shell=True)
genome = "./data/genome/B5_assembly.fa.gz "
input_file = "../kushneria/data/cleaned/"
output_file = "./data/alignments/subread_new/"

subread_align(genome, input_file, output_file)
```

```
### Count Genomic Features ###
### Takes in the aligned bam file and counts features as outlined by the K. B5
annotated genome file

import pandas as pd
import subprocess

def featureCounts(bam_file, gtf_file, output_file):
  """Counts reads by genomic features using featureCounts.
  Args:
    bam_file: The path to the BAM file.
    gtf_file: The path to the GTF file.
    output_file: The path to the output file.
  """
  filenames = sorted(os.listdir(bam_file))
  for file in filenames:
      command = f"featureCounts -T 8 -t CDS -g transcript_id -d 10 -a {gtf_file}
-o {output_file}{file.split('_')[1]}.counts {bam_file}{file}"
      subprocess.call(command, shell=True)
bam_file = "../kushneria/data/alignments/subread_new/"
gtf_file = "../kushneria/data/genome/B5.gtf"
output_file = "../kushneria/data/counts/"

featureCounts(bam_file, gtf_file, output_file)
```

```
### Cleanup Counts Data ###
### Makes count file into a dataframe and removes any gene where the total counts
across all replicates is less than 10
import os
import pandas as pd

OUTPUT_PATH = ".data/output_files/"
os.makedirs(OUTPUT_PATH, exist_ok=True)

counts_file = "../kushneria/data/counts/all_count.csv"
meta_file = "../kushneria/data/counts/count_metadata.csv"

#read genome counts csv into pandas dataframe
counts_df = pd.read_csv(counts_file,sep='\t',header=(1))
#remove start, end, strand, length, and chr columns
counts_df = counts_df.drop(["Chr", "Start", "End", "Strand", "Length"], axis=1)
#set first column values as rownames
counts_df = counts_df.set_index('Geneid')
#change column names
counts_df = counts_df.set_axis(["NS_NI_1", "S_I_1", "S_I_2", "S_I_3", "NS_NI_2",
"NS_NI_3", "NS_I_1", "NS_I_2", "NS_I_3", "S_NI_1", "S_NI_2", "S_NI_3"], axis=1)

#swap rows and columns
counts_df = counts_df.transpose().rename({'Geneid': 'sample'}, axis=0)

#read in the meta_data
meta_df = pd.read_csv(meta_file, sep=";").set_index('sample')
meta_df.columns.name = ''

#remove genes that have total counts across samples less than 10
genes_more_than_10 = counts_df.columns[counts_df.sum(axis=0) >= 10]
counts_df = counts_df[genes_more_than_10]
counts_df.columns.name = ''

counts_df.to_csv("../kushneria/data/counts/cleaned_counts.csv")
```

Appendix II

R code for analyzing the feature counts data with DESeq2, PCA, and volcano plots.

```
#DESEQ from counts table generated using subread in python environment with new
genome
counts_df <- read.csv("~/CRISPR Screen Analysis/kushneria/Kushneria Sequence
Analysis/counts_python/cleaned_counts.csv")

row.names(counts_df) <- counts_df$X

counts_df <- data.frame(t(subset(counts_df, select = -c(X))))

counts_df <- counts_df[ ,sort(colnames(counts_df))]

meta <- data.frame(Conditions = c("NS_I", "NS_I", "NS_I", "NS_NI", "NS_NI",
"NS_NI", "S_I", "S_I", "S_I", "S_NI", "S_NI", "S_NI"), row.names = samples)

dds2 <- DESeqDataSetFromMatrix(countData = counts_df,
                               colData = meta,
                               design = ~ Conditions)
dds2 <- DESeq(dds2)


# Make PCA Plot

pcaPlot <- plotPCA(rlog(dds2), intgroup = "Conditions", ntop = nrow(dds2),
returnData = TRUE)

pcaPlot

ggPCA <- ggplot(pcaPlot, aes(PC1, PC2, color=Conditions)) +
  geom_point(size=3) +
  theme(axis.title = element_text(size=16),
        axis.text = element_text(size=14),
        legend.text = element_text(size=16),
        legend.title = element_text(size=16)) +
  scale_color_discrete(name = "Condition", labels=c('5%, Induced',
                                                    '5%, Uninduced',
                                                    '12%, Induced',
                                                    '12%, Uninduced'))

ggPCA
```

```r
# Make Volcano Plot for normal salt condition with uninduced and induced groups
dds_NS_NI_vs_NS_I <- results(object = dds2,
                      contrast = c("Conditions", "NS_NI", "NS_I"),
                      lfcThreshold = 0)
dds_NS_NI_vs_NS_I <-
dds_NS_NI_vs_NS_I[order(abs(dds_NS_NI_vs_NS_I$log2FoldChange), decreasing =
TRUE), ]
write.csv(dds_NS_NI_vs_NS_I, file = "dds_NS_NI_vs_NS_I.csv")
#Make the dds_results into a data frame object
dds_NS_NI_vs_NS_I <- data.frame(dds_NS_NI_vs_NS_I)
dds_NS_NI_vs_NS_I <- na.omit(dds_NS_NI_vs_NS_I)
#add a gene_type column with the values of "up" if it's fold change is >= 2 and
if it's padj value is <= 0.05, and down regulated if it's fold change is <= 0.05
and it's padj is <= 0.05, otherwise make it "ns"
dds_NS_NI_vs_NS_I <- dds_NS_NI_vs_NS_I %>%
  mutate(gene_type = case_when(log2FoldChange >= 0.5 & padj <= 0.05 ~ "up",
                               log2FoldChange <= -0.5 & padj <= 0.05 ~ "down",
                               TRUE ~ "ns"))
down_genes <- row.names(dds_NS_NI_vs_NS_I[dds_NS_NI_vs_NS_I$gene_type ==
"down",])
dds_NS_NI_vs_NS_I <- cbind(gene=rownames(dds_NS_NI_vs_NS_I),
dds_NS_NI_vs_NS_I)
#set where vertical and horizontal significance lines will be placed
vlines <- c(-0.5, 0.5)
hline <- -10*log10(0.05)
#Aesthetics for the dots on the plot
cols <- c("up" = "#F2CA52", "down" = "#52A8F2", "ns" = "grey")
sizes <- c("up" = 2, "down" = 2, "ns" = 1)
alphas <- c("up" = 1, "down" = 1, "ns" = 0.5)
#ggplot function
vp_NS_NI_vs_NS_I <- ggplot(data = dds_NS_NI_vs_NS_I,
       aes(x = log2FoldChange,
           y = -10*log10(padj),
             fill = gene_type,
             size = gene_type,
             alpha = gene_type)) +
  geom_point(shape = 21) +
  geom_hline(yintercept = hline, linetype = "dashed") +
  geom_vline(xintercept = vlines, linetype = "dashed") +
  scale_fill_manual(values = cols, name = "") + # Modify point colour
  scale_size_manual(values = sizes, name = "") + # Modify point size
  scale_alpha_manual(values = alphas, name = "") + # Modify point transparency
  scale_x_continuous(breaks = c(seq(-15, 25, 5)),
                     limits = c(-15, 25))
```

```
# Make Volcano Plot for high salt condition with uninduced and induced groups
dds_S_NI_vs_S_I <- results(object = dds2,
                      contrast = c("Conditions", "S_NI", "S_I"),
                      lfcThreshold = 0)
dds_S_NI_vs_S_I <- dds_S_NI_vs_S_I[order(abs(dds_S_NI_vs_S_I$log2FoldChange),
decreasing = TRUE), ]
write.csv(dds_S_NI_vs_S_I, file = "dds_S_NI_vs_S_I.csv")
#Make the dds_results into a data frame object
dds_S_NI_vs_S_I <- data.frame(dds_S_NI_vs_S_I)
dds_S_NI_vs_S_I <- na.omit(dds_S_NI_vs_S_I)
#add a gene_type column with the values of "up" if it's fold change is >= 2 and
if it's padj value is <= 0.05, and down regulated if it's fold change is <= 0.05
and it's padj is <= 0.05, otherwise make it "ns"
dds_S_NI_vs_S_I <- dds_S_NI_vs_S_I %>%
  mutate(gene_type = case_when(log2FoldChange >= 0.5 & padj <= 0.05 ~ "up",
                               log2FoldChange <= -0.5 & padj <= 0.05 ~ "down",
                               TRUE ~ "ns"))

down_genes <- row.names(dds_S_NI_vs_S_I[dds_S_NI_vs_S_I$gene_type == "down",])
dds_S_NI_vs_S_I <- cbind(gene=rownames(dds_S_NI_vs_S_I),
dds_S_NI_vs_S_I)
#set where vertical and horizontal significance lines will be placed
vlines <- c(-0.5, 0.5)
hline <- -10*log10(0.05)
#Aesthetics for the dots on the plot
cols <- c("up" = "#F2CA52", "down" = "#52A8F2", "ns" = "grey")
sizes <- c("up" = 2, "down" = 2, "ns" = 1)
alphas <- c("up" = 1, "down" = 1, "ns" = 0.5)
#ggplot function
vp_S_NI_vs_S_I <- ggplot(data = dds_S_NI_vs_S_I,
       aes(x = log2FoldChange,
           y = -10*log10(padj),
             fill = gene_type,
             size = gene_type,
             alpha = gene_type)) +
  geom_point(shape = 21) +
  geom_hline(yintercept = hline, linetype = "dashed") +
  geom_vline(xintercept = vlines, linetype = "dashed") +
  scale_fill_manual(values = cols, name = "") + # Modify point colour
  scale_size_manual(values = sizes, name = "") + # Modify point size
  scale_alpha_manual(values = alphas, name = "") + # Modify point transparency
  scale_x_continuous(breaks = c(seq(-15, 25, 5)),
                     limits = c(-15, 25))
```