Identification of Bacterial Operons Required During The Plant Immune Response using a RB-TnSeq Approach

Emily Jordan Getzen

Honors Thesis
University of North Carolina at Chapel Hill

April 2019

Advisors:
Jonathan M. Conway and Jeffery L. Dangl

Approved By:

_____
Jeffery Dangl
Thesis Advisor

_____
Lianne Cartee

_____
Kenneth Donnelly

Abstract:

Today, the majority of agrochemicals used to increase agricultural output have adverse effects on the environment, and the developments of more sustainable approaches that utilize plant-growth promoting bacteria are needed [1]. In order to benefit the plants, these microbes must be able to colonize the plant root in mixed environmental microbial communities that induce immune responses from the plant. Toward understanding the genetic basis of root colonization by plant-growth promoting microbes, a randomly barcoded transposon insertion library sequencing method (Bar-Seq) was utilized to identify genes that protect bacteria from the plant-immune response. *Arabidopsis thaliana* ecotype Col-0 and two mutants deficient in immune responses were grown for 5 weeks in liquid media and then an immune response was induced for 1 day using flg22 flagellin peptide. Exudates containing anti-microbial chemicals produced by the plants were collected. Bar-Seq libraries of four different plant-associated bacteria were grown in each exudate and the transposon barcodes corresponding to interrupted genes were sequenced. First, the genes are assigned fitness scores by comparing barcode abundance to input. Then, the fitness scores are compared (mutant vs. wild-type, + flg22 vs. – flg22) to identify genes required for growth during the immune response. These results will assist in the efforts to achieve higher crop yields in a more sustainable way, and provide insight into the genetic interactions between bacteria and their hosts.

Introduction:

With the human population increasing by tens of millions of people each year, the need for high agricultural yields continues to be of utmost importance. Today, hundreds of millions are chronically malnourished. In previous decades, growth in food production was achieved by the increased use of fertilizers, pesticides, and expansion of land used for irrigated crops [2]. However, the same farms that rely on these methods are associated with poor soil health and deterioration of the environment [1]. One of the most utilized chemical fertilizers is nitrogen; it is not entirely absorbed by plants and interferes with both ground and surface water. Pesticides, another abundantly used agrochemical, are hazardous toward non-target organisms and pose a risk for farmworkers exposed to high concentrations [3]. An alternative approach to the use of these fertilizers and pesticides is to utilize growth-promoting microbes to generate an increase in plant productivity. Currently, the Organic Materials Review Institute lists 174 products named 'microbial inoculants', but few of these products are backed by mechanistic studies [7].

Two heavily studied plant-growth promoting bacteria are *Pseudomonas simiae* strain WCS417, and *Burkholderia bryophila* strain MF376. Plant-growth promoting activities, including lateral root growth stimulation and activation of auxin signaling pathways, are displayed by *P. simiae* WCS417 [4]. *B. bryophila* MF376 is known to increase the surface area of plant shoots, and is associated with bioremediation and plant-growth promotion [8]. Both are known to be effective colonizers of plants [4] [8].

In order for plant growth-promoters (PGPs) to exert this effect on plants, they need to be able to efficiently colonize the plant roots and rhizosphere in mixed environmental microbial communities that induce immune responses from the plant. The

first line of defense of plant immunity utilizes transmembrane pattern recognition receptors that respond to microbe- or pathogen-associated molecular patterns (MAMPs or PAMPs) [9]. PAMP-triggered immunity (PTI) serves to halt further colonization upon recognition of a PTI elicitor [9]. A common PTI elicitor is bacterial flagellin, the principal component of the flagellum filament, which is used by bacteria for flagellum-based motility. A 22-amino acid epitope from bacterial flagellin (flg22) is recognized by the conserved FLS2 receptor in the plant, triggering PTI [9]. The flg22 response is restricted to the elongation zone of the root tip, and also induces long-root inhibition in *Arabidopsis thaliana* [10]. Plants also have the ability to synthesize hormones and defense compounds that play an important role in defense against pathogens. One such hormone is salicylic acid, a common plant phenolic that has been confirmed by many studies to be a signal in PTI [11]. Plants that are unable to accumulate the hormone have enhanced susceptibility to viral, fungal, and bacterial pathogens [12]. Plants also utilize secondary metabolites synthesized following a pathogenic attack, such as phytoalexins and coumarins [15]. Phytoalexins accumulate rapidly at the site of a pathogen infection [15], and coumarins have been confirmed by many studies to have antibacterial, fungitoxic, and insecticide properties [16].

The genetic mechanisms behind successful microbe-plant colonization may be generalizable across different plant growth-promoting bacterial strains. In order to identify genes required for root and rhizosphere colonization, the randomly barcoded transposon mutagenesis sequencing (RB-TnSeq) method was utilized. The RB-TnSeq method combines the advantages of TnSeq (the generation of large numbers of gene disruption mutant strains) and the high throughput quantification of DNA barcodes with

next generation sequencing [5]. In this method, a pool of bacterial mutants are constructed in which each carries a randomly barcoded transposon that disrupts a single gene. Each barcode is mapped to the gene that was disrupted, generating a large map of barcoded gene disruptions within different mutants within the population of a single bacterial strain. The library of mutants can be utilized to assess genetic mechanisms behind bacterial fitness in different environments based on the relative survival rates of mutants. This assessment is accomplished by high-throughput sequencing and quantification of the change in barcode abundance between two culture conditions.

The utilization of the RB-TnSeq method has led to the successful identification of genes required for fitness in various conditions for different bacteria. In a previous study, RB-TnSeq was used in 387 mutant fitness assays for 130 different bacteria-carbon sources [5]. As a result, 5,196 genes with significant phenotypes were annotated for five bacteria [5]. In a more recent study, 115 *P. simiae* WCS417 genes were identified as required for colonization of the *A. thaliana* root system related to motility and carbon metabolism [4].

For this experiment, *Arabidopsis thaliana* ecotype Col-0 and the fls2 (lacking flg22 perception) and sid2 (lacking salicylic acid accumulation) mutants were grown in liquid media for five weeks. Then an immune response was induced in half the plants for each condition by using flg22 peptide. Exudates containing anti-microbial chemicals produced by the flg22 response were collected. Bar-Seq libraries of MF376 and WCS417 were grown in each exudate condition and fitness scores for the genes in each strain were determined based on barcode abundance. Using these fitness scores, we confirmed

known bacterial mechanisms for survival against the plant immune response, and identified new mechanisms with the potential for further study.

Methods:

Seed Surface Sterilization and Germination

Arabidopsis thaliana Col-0, fls2, and sid2 seeds were surface sterilized by suspending seeds in 70% ethanol and 0.1% Tween 20 for 1 minute, washing once with distilled water, re-suspending in 10% bleach and 0.1% Tween 20 for 10 minutes, and washing four times with distilled water. Seeds were finally re-suspended in distilled water, wrapped in aluminum foil and stratified in the dark at 4°C for at least 3 days. Stratified seeds were transferred onto square petri dishes containing ½ X MS medium, pH 5.8 (2.22 g/L Murashige & Skoog modified basal medium with Gamborg Vitamins - PhytoTechnology Laboratories product # M404, 0.02% MES hydrate) with 1% agar (Bacto). Petri dishes with seeds were placed in a growth chamber with a 9h-light/15h-dark photoperiod at 21/18 °C for three days.

Bacterial Growth Conditions

1 mL aliquots of the Bar-Seq libraries (MF376, WCS417) were added to 50 mL of media 2xYT media with 50 µg/mL kanamycin and grown 6 hours (into the exponential growth phase). They were then inoculated in 2xYT liquid medium, and placed in a shaking incubator at 28 °C for two days.

Exudate Collection System

One system is comprised of one glass 16-oz mason jar filled with 250 g of 3mm glass beads and a 2.875” 304 stainless steel 60’’ mesh disk (McMaster-Carr part# 9317T197). The mason jars and glass beads were washed with 10% HCl before adding the mesh disk and autoclaving. After the devices were sterilized, 70 mL of liquid ½ X MS media, pH 5.8 was added. Approximately 50 germinated seedlings (grown for 4 days) were transferred under sterile conditions to the stainless steel mesh and placed into the exudate collection system jars. The mesh floats slightly above the glass beads, and the plant roots eventually find their way into the holes of the mesh and into the liquid media and glass beads. Fourteen devices were used for each plant genotype (Col-0, fls2, sid2). The Mason jars were sealed with Breathe-Easy sealing film and were placed in a short-day growth chamber (9h-light/15h-dark photoperiod at 21/18 °C) for five weeks before root exudate is collected.

Figure 1: Exudate collection system.

Exudate Collection

After five weeks of growth, the Mason jars were taken to sterile hoods. All of the liquid was removed from the jar with a sterile 60mL syringe, and replaced with 70 mL of new ½ X MS liquid media. In half of the replicates (7 Mason jars out of 14) for each plant genotype, 1 µM flg22 peptide was added to induce the flg22 response in these plants. The mason jars were returned to the short-day growth chamber for 24 hours. Then, the media containing exudates released over the 24 hour period was removed, pooled for each condition, and frozen at -80°C. The exudate containing media was concentrated by freeze-drying approximately 5-fold.

Figure 2: Freeze-drying the exudate.

In order to normalize the carbon content of the various concentrated exudates, non-purgeable organic carbon (NPOC) was quantified in each condition's pooled concentrated exudate sample. The concentrated exudate was sterile filtered and then diluted 1/10 and 1/100. Samples were acidified to 0.5% phosphoric acid and run on a TOC-L Total Organic Carbon Analyzer (Shimadzu).

RB-TnSeq library growth in Plant Exudates

Each exudate condition was normalized to 700 ppm NPOC, and 0.1% glucose was added. Cultures for each library were washed 3x with 10 mM $MgCl_2$ and normalized OD600 to 0.4. 330 µL of the washed libraries were added to 4 mL of the exudates to achieve OD600 = 0.3 inoculation. Each condition was 4 mL of the bacteria-exudate mixture in triplicate. The samples were incubated at 28C to early stationary phase (~7-12 hours). After growth, the samples were centrifuged at 4000G, supernatant was removed, and samples were frozen at -80 °C.

Barcode amplification and Library Preparation

DNA was extracted using the Quick-DNA Miniprep Kit (ZymoResearch) according to the manufacturer's protocol [8]. DNA was quantified on the NanoDrop and frozen at -20C. 20ng DNA harvested from each condition was used as the template for PCR amplification of the transposon barcode for Bar-Seq analysis. Truseq-derived indexed Bar-Seq primers (already with the illumina adapters) flanked the transposon barcode region, and PCR was performed using a Q5 polymerase with Q5 GC enhancer for 25 cycles of 30 seconds at 98 °C, 30 seconds at 55 °C, and 30 seconds at 72 °C, then a final extension of 72 °C for five minutes [5].

After PCR, 3 µL of each reaction was pooled and purified using AMPureXP beads according to the manufacturer's protocol [37]. The clean library pool of PCR products from 96 multiplexed samples was sequenced on one lane of the Illumina HiSeq 2500 HighOutput using the 2T paired-end 50-cycle protocol.

Computational Bar-Seq Pipeline

Scripts from the BarSeqR pipeline [5] and FastQ files from the Illumina HiSeq 2500 run were downloaded and placed in a newly created Bar-Seq Analysis working directory on a Linux computer. MultiCodes.pl identifies the barcode in each read and makes a table of barcode counts. CombineBarSeq.pl used the pool definition file and table of counts for each individual condition/replicate to combine all the information and generate a table that describes the strain count. The genome assembly (.fna) and annotation (.gff) files for WCS417 (IMG Genome ID 2585427642) and MF376 (IMG Genome ID 2546825544) were downloaded into the working directory and used in SetupOrg.pl to create an output directory for a genome sequence file, genes tab file, and a file describing the GC content of the genes. From there, BarSeqR.pl script combined the individual sets from the combineBarSeq.pl and the genes table from SetupOrg.pl. This script also invoked the FEBA.R script to produce the gene fitness scores. Fitness is calculated using the equation: $fitness = \frac{\log(Abundance\ in\ Condition)}{\log(Abundance\ in\ 2XYT)}$ where IPT refers to the washed library initially grown in 2XYT, and Exp refers to mutants used to inoculate any of the exudate selective growth conditions. See exact instructions in Appendix.

To obtain bacterial fitness scores related to the plant immune response, fitness scores in wild-type (WT) without flg22, mutant exudates +/- flg22, and MS were then compared to mutant survival in WT Col-0 exudate with flg22 using the equation $fitness = ExpScore - ControlScore$ where *ControlScore* refers to fitness change from the IPT condition to WT+flg22 and *ExpScore* refers to the fitness change from the IPT condition to any of the other exudate conditions.

In order to determine significant fitness scores, two approaches were utilized. The first approach involved looking for significant differences based on a *P* value of 0.05 (Student *t* test) for each gene by comparing the three replicates that made up the *ControlScore* with the three replicates that made up the *ExpScore*. Because weak fitness scores that are considered significant based on the *P* value may not be biologically significant, we also employed a threshold cut-off of 0.5. These scores were used to identify significantly depleted genes in the WT+flg22 exudate condition. An Rscript was written to generate an excel file of the final fitness scores > 0.5 with the ID, location, description, score per *ControlScore* replicate and score per *fitness* replicate, as well as count significant genes in each condition and identify commonalities between different conditions (see Appendix for full script). The *t* test was computed in the excel file. This dataset was used to identify operons containing genes with depleted fitness scores. In order to identify genes with depleted fitness scores on the same operon, the Integrated Microbial Genomes & Microbiomes database [14] was utilized. For the purposes of this paper, only some operons and some *highly* significant genes (fitness score > 1.2) were analyzed in depth. Significant genes (fitness score > 0.5, P value <0.05) are highlighted in yellow for each table.

Results:

Growth Curves:

To establish the ability of the BarSeq libraries to grow in MS medium and plant exudate, growth curves on plant exudates and MS in various levels of glucose were performed (1%, 0.1%, and 0%).

With a goal of obtaining at least three replications, it was found that growth was optimal for WCS417 libraries in exudates with the addition of 0.1% glucose. WCS417 library growth is not supported in plant exudates or MS only. Early stationary phase occurred at ~7 hours. The growth curves in the other exudates looked very similar to the example growth curve of WCS417 in fls2 exudate (figure 4).

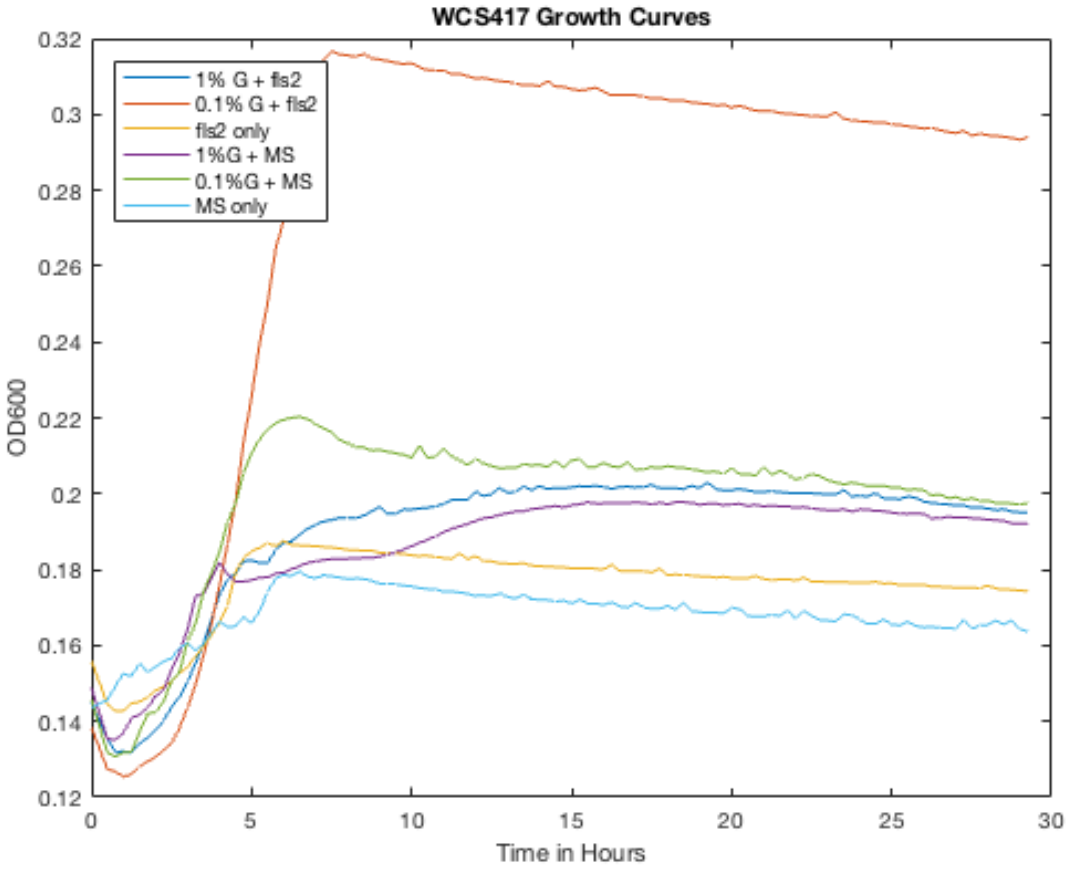

Figure 4: Growth curves for WCS417 in MS and fls2 exudate + 0, 0.1%, or 1% glucose

For the MF376 library, it was found that growth was optimal in both plant exudates + 0.1% glucose and MS + 0.1 % glucose. Early stationary phase occurred at (~12 hours). MF376 library growth was also not supported in plant exudates or MS only. Growth curves for the MF376 library in other exudates also looked very similar to the fls2 example (figure 5).

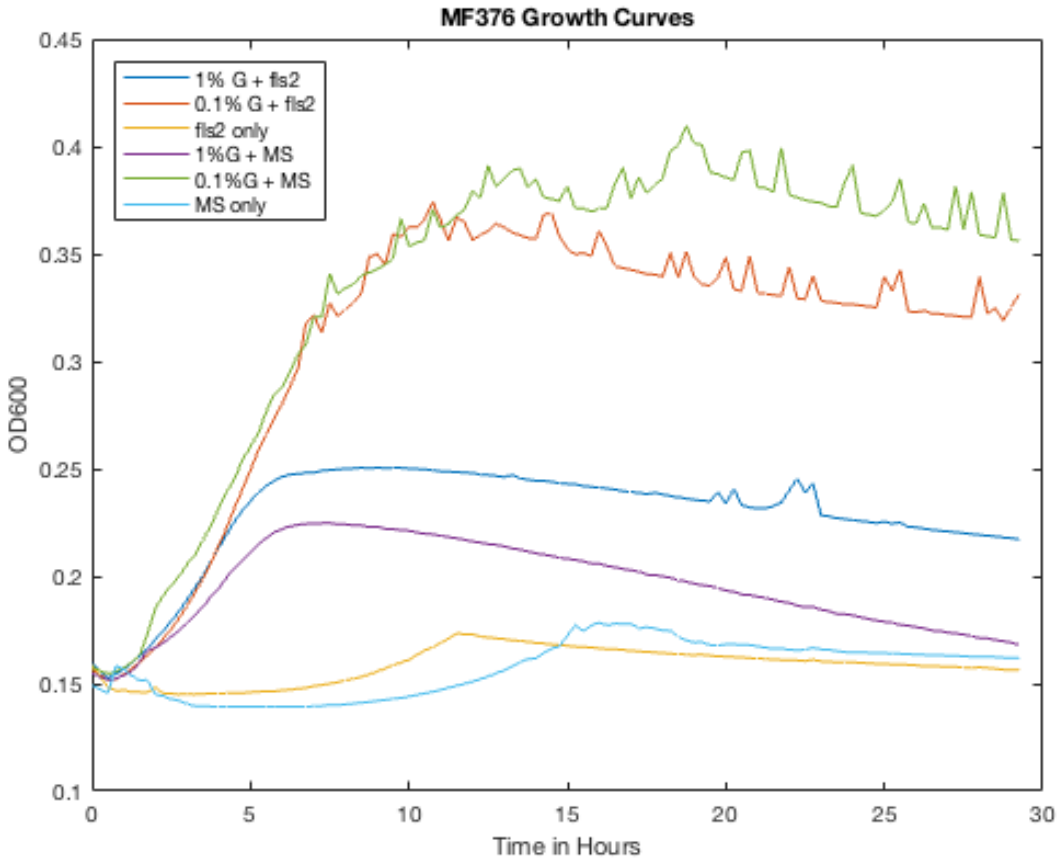

Figure 5: Growth curves for MF376 in MS and fls2 exudate + 0, 0.1%, or 1% glucose

TOC Analyzer:

Exudates collected after 24-hour exposure to flg22 were analyzed for non-purgeable organic carbon (NPOC). The results (Table 1) allowed for the normalization of NPOC levels across each exudate condition before libraries were grown.

Table 1: TOC Analyzer Results

Exudate Condition	Dilution	NPOC measured [ppm]	NPOC in sample [ppm]
WT + flg22	1/100	8.90	890
WT – flg22	1/100	9.83	983
fls2 + flg22	1/100	10.14	1014
fls2 – flg22	1/100	11.14	1114
Sid2 + flg22	1/100	7.39	739
Sid2 – flg22	1/100	10.51	1051
5x MS media	1/100	20.69	2069

Preliminary Bar-Seq Results:

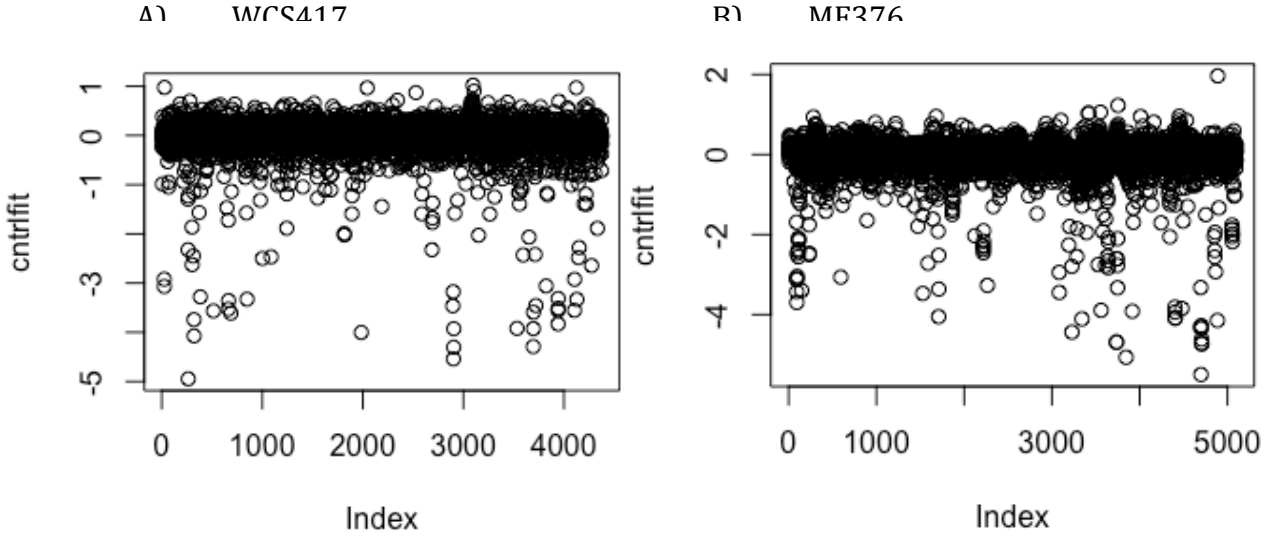

Figure 5: An illustration of gene fitness deficits in WT+flg22 A) Averaged WT+flg22 fitness scores for WCS417. B) Averaged WT+flg22 fitness scores for MF376

Preliminary Bar-Seq results displayed that there were indeed significant fitness defects in the Wild-Type+flg22 exudate conditions for MF376 and WCS417 compared to the respective input (IPT) conditions. Index refers to the gene, and cntrlfit refers to the fitness score after comparison of gene abundance in the averaged WT+flg22 replicates to the averaged IPT replicates. Each point on the graph is a gene and its associated fitness score.

After obtaining the initial illustration of the genetic differences between mutants grown in WT+flg22 exudate compared with 2xYT input, we shifted our attention to genes that were significantly different in one of the plant mutant exudate conditions compared to genes in WT+flg22. For example: there could be a gene with lower fitness in WT+flg22 condition compared to 2xYT that is related to metabolism, which likely would not be different in the fls2+flg22 condition. However, genes with lower fitness scores in

the WT+flg22 condition but not in the fls2+flg22 condition likely have to do directly with that gene's ability to survive against antimicrobials or other compounds released by a plant immune system in response to flg22 recognition. Significant genes in the following tables (Table 2 and Table 3) refer to genes with a fitness score difference > 0.5 after the fitness score in the experimental exudate was subtracted by the corresponding fitness score in wild-type exudate+flg22 and after removing genes with a p value > 0.05 (see Methods).

Table 2: Quantifying significant genes for each condition

Exudate Condition	# Significant Genes WCS417	# Significant Genes MF376
WT	89	74
fls2+flg22	84	72
fls2	74	93
sid2+flg22	60	87
sid2	83	65
MS	91	206

Table 3: Shared Significant Genes

Exudate Condition	# Shared WCS417	# Shared MF376
WT & fls2+flg22	41	27
WT & fls2	40	24
Fls2+flg22 & fls2	22	19
sid2+flg22 & sid2	27	22

The following tables below detail operons that contain significant genes, and a few highly significant individual genes. These tables do not constitute all of the significant genes or operons that were found. The operons and genes listed were chosen for their significant fitness defects and interesting applications in mitigating plant immunity. Differences that are significant (P-value < 0.05, *fitness* > 0.5) are highlighted in yellow for each exudate condition (referred to as Condition Score). The description of the genes in each table comes from the genes table generated by the genetic information datasets from the Integrated Microbial Genomes & Microbiomes database (see Methods) [14].

Results for WCS417:

Table 4: WCS417 Operon 1 Fitness Score Comparison

ID	Description	WT+ flg22 score	WT score	Fls2 + flg22 score	Fls2 score	Sid2+flg22 score	Sid2 score	MS score
PS417_04375	phospholipid/cholesterol/gamma-HCH transport system ATP-binding protein	-0.536	0.839	1.174	0.868	0.041	-0.084	0.425
PS417_04385	phospholipid/cholesterol/gamma-HCH transport system substrate-binding protein	-0.338	0.727	0.533	0.623	-0.354	-0.303	0.310
PS417_04390	phospholipid transport system substrate-binding protein	-0.523	0.588	0.763	0.805	0.053	0.140	0.550

This operon codes for the ABC transporter complex MlaFEDB, consisting of two ATP-binding proteins and two transmembrane proteins [17]. The complex is involved in the phospholipid transport pathway that maintains lipid asymmetry in the outer membrane via retrograde trafficking of phospholipids from the outer to inner membrane [17]. This complex has importance for survival against the plant immune response—the bacterial outer membrane in gram-negative bacteria forms a permeation barrier to protect the cell from harmful compounds, and the defining characteristic of this membrane is lipid asymmetry, or phospholipids in the inner leaflet and lipopolysaccharides in the outer leaflet [18]. The MlaFEDB complex selectively removes outer leaflet lipopolysaccharides to maintain the barrier function of the membrane [18].

Table 5: WCS417 Operon 2 Fitness Score Comparison

ID	Description	WT+ flg22 score	WT score	Fls2+ flg22 score	Fls2 score	Sid2+ flg22 score	Sid2 score	MS score
PS417_ 17290	Multidrug efflux pump subunit AcrB	-1.769	2.517	2.426	2.334	1.391	-0.109	2.204
PS417_ 17295	RND family efflux transporter, MFP subunit (AcrA)	-1.680	2.517	2.153	2.087	1.241	-0.006	1.920
PS417_ 17300	efflux transporter, outer membrane factor (OMF) lipoprotein, NodT family (TolC channel)	-1.371	1.855	1.921	1.811	1.124	0.056	1.537

This operon codes for the multidrug efflux pump AcrAB-TolC. The pump has been confirmed by many studies to confer resistance to a wide variety of compounds, including: antibiotics, dyes, detergents, and disinfectants [19]. AcrB uses proton motive force to pump unwanted compounds out of the cell from the inner membrane through the TolC channel [20]. AcrA is the periplasmic adapter protein and is believed to be involved in the maintenance and assembly of the complex, and transmits conformational changes in AcrB and TolC [20]. TolC is simply the outer membrane channel. The pump has also been shown to have a role in the colonization of various host organisms [20]. Some of the antimicrobial agents that are expelled by the pump consist of antimicrobial peptides, bile salts, and long chain fatty acids [20]. The inactivation of this pump by lack of even one structural component directly affects the colonization ability of the bacteria [20].

Table 6: WCS417 Operon 3 Fitness Score Comparison

ID	Description	WT + flg22 score	WT score	Fls2+ flg22 score	Fls2 score	Sid2+flg22 score	Sid2 score	MS score
PS417_11820	Signal transduction histidine kinase	-1.993	0.888	0.730	0.089	-0.642	0.0337	2.177
PS417_11825	two-component system, OmpR family, response regulator	-2.023	0.984	1.185	1.327	0.001	-0.356	2.3757

This operon codes for a two-component system that represents a stimulus-response mechanism. A stimulus is sensed by histidine kinase, and transmitted to a

response regulator. The regulator binds to DNA and mediates a cellular response [21]. The types of stimuli that are sensed by this system consist of: osmotic changes, temperature, small molecules, and antimicrobials [21].

Table 7: WCS417 Other Significant Genes Fitness Score Comparison

ID	Description	WT+ flg22 score	WT score	Fls2+ flg22 score	Fls2 score	Sid2+flg22 score	Sid2 score	MS score
PS417_06530	Outer membrane protein OmpA	-1.318	0.888	1.003	0.089	-0.642	0.034	2.177
PS417_01955	penicillin-binding protein 1A	-2.449	0.634	0.846	0.212	0.836	1.382	2.343

OmpA is a major protein in the bacterial cell membrane that is important for outer membrane stability. It has been confirmed that a deletion makes a mutant more sensitive to an acidic environment, high osmolarity, and pooled human serum. OmpA proteins also have important pathogenic roles including intracellular survival, bacterial adhesion, and evasion of host defenses. The data on evasion of host defenses are primarily in the realm of the central nervous system, respiratory system, and urogenital diseases [23].

Penicillin Binding Proteins are widely known for their role in biosynthesis of the cell wall, as well as maintaining its shape and integrity. More recently, however, they have been linked to virulence in a variety of bacterial species [25].

Results for MF376:

Table 8: MF376 Operon 1 Significant Fitness Scores

ID	Description	WT+ flg22 score	WT score	Fls2+ flg22 score	Fls2 score	Sid2+flg22 score	Sid2 score	MS score
H281DRAFT_00315	Glycosyltransferase	-2.503	0.837	1.205	0.270	0.722	0.782	-0.342
H281DRAFT_00316	ADP-heptose:LPS heptosyltransferase	-1.755	0.418	0.840	0.512	0.409	0.568	-0.649
H281DRAFT_00317	Lipid A core - O-antigen ligase and related enzymes	-2.467	0.889	1.244	1.088	1.037	1.419	-1.312

These genes code for the production of lipopolysaccharides (LPS) that make up the cell membrane. These consist of lipid A, non-repeating core oligosaccharides, and O-antigens. The glycosyl transferase protein is involved with the biosynthesis of the outer core region of lipopolysaccharides [26]. The heptosyltransferase is involved with the synthesis of the inner core region of the LPS, and its absence has been proven to result in a truncated LPS as well as decreased virulence for pathogenic gram-negative bacteria [27]. Lipid A core O-antigen ligase is required for the final litigation step of the O-antigen onto the lipid A core block to complete the assembly of LPS [28].

Table 9: MF376 Operon 2 Significant Fitness Scores

ID	Description	WT+ flg22 score	WT score	Fls2+ flg22 score	Fls2 score	Sid2+f lg22 score	Sid2 score	MS score
H281D RAFT_ 06072	3-isopropylmalate dehydrogenase (EC 1.1.1.85)	-4.707	0.259	0.072	0.779	1.188	0.199	0.785
H281D RAFT_ 06073	3-isopropylmalate dehydratase, small subunit (EC 4.2.1.33)	-4.615	1.435	0.784	0.824	1.462	0.348	1.759

This operon codes for a sub-pathway that synthesizes L-Leucine from 3-methyl-2-oxobutanoate. This sub-pathway is also part of amino acid synthesis in bacteria [29]. It has been shown that in some bacteria, the adaptation to amino acid starvation also includes the induction of virulence gene expression, possibly allowing the bacteria to evade host defenses [30].

Table 10: MF376 Operon 3 Significant Fitness Scores

ID	Description	WT+ flg22 score	WT score	Fls2+f lg22 score	Fls2 score	Sid2+ flg22 score	Sid2 score	MS score
H281D RAFT_0 4719	tol-pal system beta propeller repeat protein TolB	-2.654	-0.225	-0.077	-0.410	1.007	0.784	0.589
H281D RAFT_0 4721	Cell division and transport-associated protein TolR (TC 2.C.1.2.1)	-2.075	0.305	0.176	0.503	0.589	-0.328	0.295

Operon 3 codes partially for the Tol-Pal mutliprotein complex. The complex is necessary for maintaining the outer-membrane in gram-negative bacteria. Although it has not been heavily studied, it has been confirmed that Tol-pal genes play a role in bacterial pathogenesis [31]. Mutations in this gene lead to membrane blebbing, release of periplasmic contents, and hypersensitivity towards antibiotics [32].

Table 11: MF376 Operon 4 Significant Fitness Scores

ID	Description	WT+ flg22 score	WT score	Fls2+flg22 score	Fls2 score	Sid2+flg22 score	Sid2 score	MS score
H281DRAFT_01985	Protoheme IX farnesyltransferase	-0.722	0.328	0.242	0.468	0.368	0.151	1.309
H281DRAFT_01986	Required for cytochrome oxidase assembly	-0.572	0.049	0.149	0.163	-0.300	-0.099	1.195
H281DRAFT_01987	SCO1/SenC/PrrC, involved in biogenesis of respiratory and photosynthetic systems	-0.488	-0.127	0.165	0.255	0.302	-0.001	1.262
H281DRAFT_01988	Uncharacterized conserved protein	-0.140	-0.161	0.091	-0.005	0.030	0.092	1.057
H281DRAFT_01989	Protein unknown function (DUF2909)	-0.252	0.158	0.442	0.178	0.315	0.133	0.263
H281DRAFT_01990	Heme/copper-type cytochrome/quinol oxidase, subunit 3	-0.536	0.057	-0.021	-0.037	0.062	-0.098	1.09

Operon 4 codes for Heme-O biosynthesis. Heme is necessary for proper functioning of hemoproteins, which are involved in energy generation and detoxification of host immune effectors [33]. Bacteria must either synthesize their own heme or acquire it from the host. The relationship between heme synthesis and bacterial pathogenesis is largely understudied, and most of the organisms studied are related to colonization of a human host [33].

Table 12: MF376 Operon 5 Significant Fitness Scores

ID	Description	WT+ flg22 score	WT score	Fls2+flg22 score	Fls2 score	Sid2+flg22 score	Sid2 score	MS score
H281D RAFT_05641	ABC-type transport system involved in resistance to organic solvents, ATPase component	-1.047	0.294	-0.265	0.226	-0.019	0.295	0.771
H281D RAFT_05642	conserved hypothetical integral membrane protein	-0.540	0.020	0.173	-0.346	0.190	-0.054	0.749
H281D RAFT_05643	ABC-type transport system involved in resistance to organic solvents, periplasmic component	-0.740	-0.042	-0.183	-0.068	0.016	-0.376	0.787
H281D RAFT_05644	Surface lipoprotein	0.136	0.089	0.014	-0.074	-0.032	0.063	0.049
H281D RAFT_0	ABC-type transport	-1.053	0.052	0.316	0.294	0.157	-0.018	0.988

5645	system, aux component							

Operon 5 codes for ATP binding cassetes or ABC transporters. These form one of the largest known protein families, and are widespread in bacteria [34]. Transporters couple ATP hydrolysis to active transport and translocation of substrates, including but not limited to: ions, sugars, lipids, proteins, drugs, and peptides [34]. The three major components are the integral membrane proteins, peripheral proteins that hydrolyze ATP, and a periplasmic substrate-binding protein [34]. These types of transporters are important virulence factors because they play a role in the secretion of toxins and antimicrobial agents [34].

Discussion:

As a result of conducting a Bar-Seq experiment on two plant-growth promoting bacteria that also happen to be good colonizers of the plant root and rhizosphere, a variety of genes were identified to increase the fitness of WCS417 or MF376 in the face of the plant immune response. Some were related to PTI, and some were not. Only positive scores from fitness score comparisons in experimental exudates to WT+flg22 exudates were analyzed, as it gives insight into genes that are required for survival against aspects of a fully functioning immune response, particularly with regard to plant recognition of flg22, and plant ability to synthesize salicylic acid (a signal in PTI).

There were a lot of unexpected results with regard to the sid2 exudate condition, and it is possible that the exudate was contaminated in some way.

The important genes for WCS417 coded for the ABC transporter complex MlaFEDB involved in maintenance of lipid asymmetry [17], the multidrug efflux pump AcrAB-TolC involved in the removal of unwanted compounds from the cell [19], a two-component system representing a stimulus-response mechanism [21], a protein important for outer membrane stability [23], and penicillin binding proteins [25]. The important genes for MF376 coded for the production of lipopolysaccharides (LPS) [27], a pathway that is part of amino acid synthesis [29], a complex necessary for maintaining the outer membrane [31], Heme-O synthesis [33], and a transport system responsible for pumping noxious compounds out of the cell [34].

The MlaFEDB complex in WCS417 was only found consistently significant in the WT compared to WT+flg22 condition, fls+flg22 compared to WT+flg22 condition, and fls2 compared to WT+flg22 condition (Table 4). The MS condition was *almost*

significant. Besides sid2 compared to WT+flg22, the commonality between all of these experimental exudates is that their plants either lack recognition of the PTI elicitor flg22, or flg22 was not added in conjunction with the plants. Given the results, the addition of flg22 into the WT+flg22 likely released chemicals related to PTI, and bacterial mutants lacking a proper barrier to protect themselves from noxious compounds would logically perish, compared to a system where immune chemicals are not released. It is peculiar that the same result did not occur in the sid2+flg22 experimental condition-- this would indicate that the operon is not salicylic-acid dependent.

The genes involved in the production of LPS for MF376 were found to be significant in every single experimental condition except for MS (Table 8). Its absence has been confirmed by other studies to result in a truncated LPS, as well as decreased virulence for pathogenic gram-negative bacteria [27]. Since the membrane is important in gram-negative bacteria to protect the cell against noxious compounds, it would make sense that these mutants in the WT+flg22 condition would have decreased fitness compared to conditions without flg22 recognition or salicylic acid production. It is a bit strange that mutants in the MS condition did not have a survival advantage compared to those in WT+flg22. There were no plant compounds whatsoever in the MS condition, and each of the other exudates started with ½ MS. The MS condition, however, had 5X MS instead of ½ MS, so perhaps it has something to do with the high concentrations of salts or other nutrients.

In MF376, the Tol-Pal multiprotein complex was found to be significant. Like the MlaFEDB complex in WCS417, This complex is necessary for maintaining the outer-membrane in gram-negative bacteria [31]. Interestingly, it was primarily found to be

significant only in the sid2 exudate conditions, MS condition, and slightly in the fls2 condition (Table 10). The fitness loss is sid2 dependent, but fls2 independent. However, it is interesting that there is no fitness benefit in the other exudate conditions, since in these conditions there is no flg22 recognition by the plant, and therefore presumably no PTI response. These results could indicate that perhaps the plant produces harmful chemicals without flg22 induction of PTI, and these chemicals require salicylic acid.

The AcrAB-TolC multidrug efflux pump complex was found to be important for survival in the WT+flg22 conditions when compared to every other experimental condition besides sid2 (Table 5). The necessity for the efflux pump in the WT+flg22 condition strongly suggests the fact that noxious compounds were indeed released by the plants upon flg22 recognition. They also confirm that flg22 recognition is important for the induction of PTI, and the fitness loss is salicylic acid-independent. It is peculiar that the sid2 experimental condition did not bring about the same results, especially considering that there was no flg22 response *and* there was a mutation in the sid2 gene.

Operon 5 in MF376 codes for ATP binding cassettes, which transport a variety of compounds out of the cell via ATP hydrolysis, and play a role in the secretion of toxins and antimicrobial agents. The function of these proteins is very similar to the AcrAB-TolC complex described above. However, it is clear that there was a fitness deficit in the WT+flg22 mutants (Table 12) and there was only a survival advantage in the MS experimental exudate. This would indicate that regardless of a PTI response, the plants likely produced something that ATP binding cassettes normally remove, regardless of any fls2 and sid2 mutations, and perhaps these chemicals are generally harmful. This

phenomena is worth further experimentation, given the differences between the two bacteria and their efflux pump functionality.

The fact that the importance of lipid asymmetry and multidrug efflux pumps were deemed significant in many conditions by our experiment is a good sanity check, considering that both systems have been heavily studied in a bacteria-plant context. The rest of the findings are not as well studied in a plant-bacterial colonization context.

A two-component system representing a stimulus-response mechanism with histidine kinase, discovered in WCS417, was found to be important for survival against the plant immune response in all experimental conditions except for both sid2 conditions (Table 6). The types of stimuli that are recognized by histidine kinase are osmotic changes, temperature, small molecules, and antimicrobials [21]. Two-component systems are abundant in most bacteria, and have been confirmed to be important in plant rhizosphere colonization, but the details are not heavily studied [35]. Another study found that *Salmonella typhimurium* had decreased survival rate in mouse macrophages when this system was mutated, confirming the importance for the system in a host immune context [31]. It is interesting that a mutation in the sid2 gene in the presence of flg22 had no effect on the bacterial genetic fitness scores for this system— similar to the LPS production in MF376, this would indicate that some form of an immune response is still occurring in the sid2+flg22 condition, and lack of the two-component signaling system would make it difficult to survive against it.

One of the single genes that we chose to mention in this paper is a gene that codes for OmpA in WCS417. OmpA is important for the maintenance of the bacterial cell membrane and has been found to be important for pathogenic roles such as evasion of

host defenses, but the majority of studies on this protein have to do with a human host, particularly in the central nervous system, respiratory system, and urogenital disease [23]. It was found to be significant in the WT, fls2+flg22, and MS experimental conditions (Table 7). Since there have been a lot of studies on the role of OmpA in mitigating the immune response in a human host, it would be interesting to explore parallels with mitigation against the plant immune response. The other single gene that we decided to mention in this paper codes for a penicillin binding protein in WCS417. This protein in particular is important for biosynthesis of the cell wall and has recently been linked to virulence in a wide variety of bacterial species [25]. It would also be interesting to understand why this gene is important in WCS417 but not in MF376, since they are both considered effective colonizers.

In MF376, the WT+flg22 condition had a deficit in the genes that code for the synthesis of L-Leucine. This sub-pathway is also part of the overall pathway involved in amino acid synthesis. This phenomenon has not been studied highly in the context of plant-bacterial colonization-- there exists one study that confirmed that in some bacteria, the adaptation to amino acid starvation also induces virulent gene expression, but MF376 was not one of the bacteria involved in this study [30]. It was found to be significant in each experimental condition besides sid2 (Table 8). Since little information is known about the role of this operon in enhanced colonization ability, this operon is worth further study.

In MF376, there were interesting findings on the importance of Heme-O biosynthesis for survival in WT+flg22 exudate, but the only condition with a fitness advantage was the MS condition (Table 10). Hemoproteins are involved in energy

generation and detoxification of host immune effectors, but the relationship between heme synthesis and bacterial pathogenesis is largely understudied [33]. Since there was only an advantage in the MS condition, this would indicate that the plants produced chemicals that harmed the mutants regardless of a PTI response related to flg22 recognition or salicylic acid production. The exact details of how this relates to the synthesis of hemoproteins is worth further study.

Although there were a multitude of genes and operons that were discovered using a Bar-Seq approach, there still exist genes and operons that likely contribute to the survival of an individual bacteria, but were masked due to the limitations of Bar-Seq. Because each mutant was grown in a population of thousands of other mutants, it is entirely possible that a mutant deficient in producing a single chemical would be able to obtain it from the other chemical-producing bacteria, and fitness deficits would not be detected. Essential genes would also be masked as mutants would likely die in the IPT conditions and a proper fitness score would not be determined.

The results of this paper confirmed known bacterial mechanisms for survival against the plant immune response, and identified new mechanisms with the potential for further study. It is clear that there are a lot of genes that code for complexes or pathways that contribute to bacterial survival in the face of the plant immune response, but the specifics of what exactly assists with the survival is unknown. It would be interesting to delve deeper into using plant mutants that cannot make specific defense compounds or hormones, and seeing if it is possible to match genes that are specific to these compounds. It would also be interesting to identify all of the chemicals in each exudate condition. For the scope of this experiment, we simply confirmed with the TOC levels

that the plants were indeed producing chemicals. The identities of the chemicals might assist in understanding why there were only fitness advantages in some of the experimental exudate conditions as opposed to others.

Bar-Seq could also potentially be utilized to understand the genetic mechanisms behind bacterial community dynamics. For this experiment, we utilized single isolates, but it would be worth combining a Bar-Seq library with other bacteria to see how this changes the results. Potential pairwise strain interactions or small community interactions could be explored in this context. All forms of further research from the results of this experiment hold great promise for the identification of the exact mechanisms behind plant colonization, so that these mechanisms can be utilized to eventually enhance beneficial microbial colonization in an agronomic setting.

References

1. Reganold, John P., Papendick, R. I., Parr, J. F., "Sustainable Agriculture." *Scientific American,* vol. 262, no. 6, 1990, pp.112-120.

2. Singh, R. L., Mondal, S., "Microbial Biotechnology and Sustainable Agriculture." *Woodhead Publishing, an Imprint of Elsevier*, 2018, pp. 185-205.

3. Aktar, M.W., Sengupta, D., Chowdhury, A., "Impact of pesticides use in agriculture: their benefits and hazards" *Interdisciplinary Toxicology,* (2009), pp. 1-12.

4. Cole, B. J., Feltcher, M. E., Waters, R. J., Wetmore, K. M., Mucyn, T. S., Ryan, E. M., Wang, G., Ul-Hasan, S., McDonald, M., Yoshikuni, Y., Malmstrom, R. R., Deutschbauer, A. M., "Genome-wide identification of bacterial plant colonization genes." *PLoS Biol, (2017).*

5. Wetmore, K.M., Price, M. N., Waters, R. J, Lamson, J. S., He, J., Hoover, C. A., Blow, M. J., Bristow, J., Butland, G., Arkin, A. P., Deutschbauer, A. M., "Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons." *MBio*, (2017).

6. Trivedi, P., Schenk, P. M., Wallenstein, M. D., Singh, B. K., "Tiny microbes, big yields: enhancing food crop production with biological solutions." *Microb. Biotechnol.*, 10, 999–1003.

7. Finkel, O. M., Castrillo, G., Paredes, S. H., Gonzales, I. S., Dangl, J. L., "Understanding and exploiting plant beneficial microbes." *Science Direct.* (2017).

8. Govan, J. R. W., Hughes, J. E., Vandamme, P., "Burkholderia cepacia: medical, taxonomic, and ecological issues." *J Med Bicrobiology.* (1996).

9. Jones, J. D., Dangl, J. L., "The Plant Immune System." *Nature.* (2006).

10. Millet, Y. A., Danna, C. H., Clay, N. K., Songnuan, W., Simon, M. D., Werck-Reichart, D., Ausubel, F. M., "Innate Immune Responses Activated in *Arabidopsis* Roots by Microbe-Associated Molecular Patterns." *The Plant Cell.* (2010).

11. Gaffney, T., Friedrich, L., Vernooij, B., Negrotto, D., Nye, G., Uknes, S., Ward, E., Kessmann, H., Ryals, J., "Requirement of Salycylic Acid for the Induction of Systemic Acquired Resistance." *Science.* (1993).

12. Delaney, T. P., Uknes, S., Vernooij, B., Friedrich, L., Weymann, K., Negrotto, D., Gaffney, T., Gut-Rella, M., Kessmann, H., Ward, E., Ryals, J., "A Central Role of Salicylic Acid in Plant Disease Resistance" *Science.* (1994).

13. “About Arabidopsis.” *TAIR - About Arabidopsis*, www.arabidopsis.org/portals/education/aboutarabidopsis.jsp.

14. *DOE Joint Genome Institute - IMG/ER*, img.jgi.doe.gov/cgi-b in/mer/main.cgi?section=FindGenes.

15. Ahuja, I., Kissen, R., Bones, Atle., “Phytoalexins in defense against pathogens.” *Science Direct*. (2012).

16. Razavi, S. M., “Plant Coumarins as Allelopathic Agents.” *Science Alert*. (2011).

17. “Intermembrane Phospholipid Transport System ATP-Binding Protein MlaF.” *European Bioinformatics InstituteProtein Information ResourceSIB Swiss Institute of Bioinformatics*, 7 Nov. 2018, www.uniprot.org/uniprot/P45031.

18. Abellon-Ruiz, J., Kaptan, S, S., Basle, A., Claudi, B., Bumann, D., Kleinekathofer, U., van den Berg, B., “Structural basis for maintenance of bacterial outer membrane lipid asymmetry.” *Nature*. (2017).

19. Hobbs, E. C., Xuefeng, Y., Paul, B. J., Astarita, J. L., Storz, G., “Conserved small protein associates with the multidrug efflux pump AcrB and differentially affects antibiotic resistance.” *PNAS.* (2012).

20. Perez, A., Poza, M., Fernandez, A., Fernandez, M., Mallo, S., Merino, M., Rumbo-Feal, S., Cabral, M, P., Bou, G., “Involvement of the AcrAB-TolC Efflux Pump in the Resistance, Fitness, and Virulence of *Enterobacter cloacae*.” *American Society for Microbiology*. (2012).

21. Bhate, M. P., Molnar, K. S., Goulian, M., DeGrado, W. F., “Signal Transduction in Histidine Kinases: Insights from New Structures.” *HHS Public Access.* (2015).

22. Wang, Y., “The function of OmpA in *Escherichia coli*.” *Biochemical and Biophysical Research Communications*. (2002).

23. Confer, A. W., Ayalew, S., “The OmpA family of proteins: roles in bacterial pathogenesis and immunity.” *Veterinary Microbiology.* (2013).

24. Plesa, M., Hernalsteens, J., Vandenbussche, G., Ruysschaert, J., Cornelius, P., “The SlyB outer membrane lipoprotein of *Burkholderia mutlivorans* contributes to membrane integrity” *Research in Microbiology. (*2006).

25. Hamilton, A., Popham, D. L., Carl, D. J., Lauth, x., Nizet, V., Jones, A. L., “Penicillin-Binding Protein 1a Promotes Resistance of Group B Streptococcus to Antimicrobial Peptides.” *Infection and Immunity. (*2006).

26. Leipold, M. D., Vinogradov, E., Whitfield, C., "Glycosyltransferases involved in the biosynthesis of the outer core region of Escherichia coli lipopolysaccharides exhibit broader substrate specificities than is predicted from lipopolysaccharide structures." *The Journal of Biological Chemistry.* (2007).

27. Grizot, S., Salem, M., Vongsouthi, V., Durand, L., Moreau, F., Dohi, H., Vincent, S., Escaich, S., Ducruix, A., "Structure of the Escherichia coli heptosyltransferasae WaaC: binary complexes with ADP and ADP-2-deoxy-2-fluoro heptose."

28. Han, W., Wu, B., Li, L., Zhao, G., Woodward, R., Pettit, Nicholas, Cai, L., Thon, V., Wang, P. G., "Defining Function of Lipopolysaccharide O-antigen Ligase WaaL Using Chemoenzymatically Synthesized Substrates." *Journal of Biological Chemistry.* (2011).

29. European Bioinformatics InstituteProtein Information ResourceSIB Swiss Institute of Bioinformatics. "3-Isopropylmalate Dehydrogenase." *European Bioinformatics InstituteProtein Information ResourceSIB Swiss Institute of Bioinformatics*, 13 Feb. 2019, www.uniprot.org/uniprot/P30125.

30. Kaiser, J. C., Heinrichs, D. E., "Branching Out: Alterations in Bacterial Physiology and Virulence Due to Branched-Chain Amino Acid Deprivation.", *mBio.* (2018)

31. Miller, S. I., Kukral, A. M., Mekalanos, J. J., "A two-component regulatory system (phoP phoQ) controls Salmonella typhimurium virulence." *PNAS*. (1989).

32. Wojdyla, J. A., Cutts, E., Kaminska, R., Papadakos, G., Hopper, J. T. S., Stansfield, P. J., Staunton, D., Robinson, C. V., Kleanthous, C., "Structure and Function of the *Escherichia coli* Tol-Pal Stator Protein TolR." *Journal of Biological Chemistry.* (2015).

33. Choby, Jacob. E., Skaar, E. P., "Heme Synthesis and Acquisition in Bacterial Pathogens." *HHS Public Access.* (2016).

34. *KEGG PATHWAY: smz02010*, www.genome.jp/dbget-bin/www_bget?smz02010.

35. Borland, S., Prigent-Combaret, C., Wisniewski-Dye, F., "Bacterial hybrid histidine kinases in plant-bacteria interactions." *Microbiology.* (2016).

36. Price, M. N., Wetmore, K. M., Waters, R. J., Callaghan, Ray, J., Kuehl, J. V., et al. "Deep Annotation of Protein Function across Diverse Bacteria from Mutant Phenotypes". *bioRxiv.* (2016).

37. "AMPure XP for PCR Purification." *Beckman Coulter Life Sciences*, www.beckman.com/reagents/genomic/cleanup-and-size-selection/pcr.

Appendix:

PCR Quality:

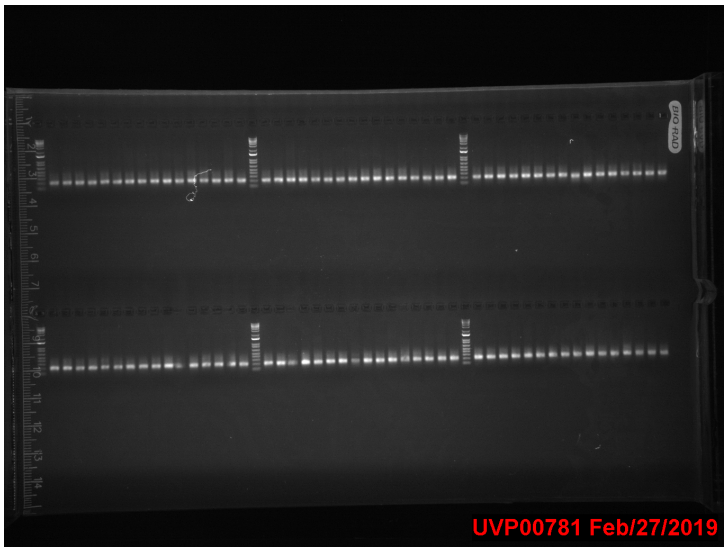

Figure 3: High Quality Gel after PCR

Computational Pipeline Instructions:

Commands:

```
perl berkeleylab-feba-391e1be315ad/bin/MultiCodes.pl -out ~name~ -index ~index~
        <filename> -n25
perl berkeleylab-feba-391e1be315ad/bin/combineBarSeq.pl out WCS417_ML3.pool IPT-
        A-1 codes AND ALL OTHER CODES
perl berkeleylab-feba-391e1be315ad/bin/SetUpOrg.pl -name WCS417 -gff
        2585427642.gff -fna 2585427642.fna
perl berkeleylab-feba-391e1be315ad/bin/BarSeqR.pl -org WCS417 -pool
        g/WCS417/out.poolcount -outdir html/WCS417
Rscript /home/apklein/Documents/Barseq/berkeleylab-feba-
        391e1be315ad/bin/RunFEBA.R WCS417 html/WCS417
        /home/apklein/Documents/Barseq/berkeleylab-feba-391e1be315ad >
        html/WCS417/log
```

Instructions:

1. Download fastq files
2. Clone repository into desired folder
3. Gunzip fastq files, unzip repository, take info down from metadata to run first script
4. Invoke Multicodes.pl script for each fastq file in desired set
5. Put pool definition file inside directory (WCS417_ML3.pool)

6. Do combineBarSeq.pl script with codes
7. Put gff and fna file in directory
8. Create a directory g/BacteriaName within current directory
9. Run SetUpOrg.pl script
10. Get the FEBA_Barseq.tsv file, edit as necessary, and move it into g/WCS417
11. Make a new directory html/WCS417
12. Move out.poolcount to g/BacteriaName and rename to setA.poolcount
13. Run the BarSeqR.pl script
14. Run the FEBA.R script

Fitness Analysis Rscript:

```
# Example with MF376 MS data

# fitness score calculation

#read data

fit_logratios <- read.delim("~/Downloads/MF376_MS_fit_logratios.txt")

#define columns

locusID = fit_logratios[,c(1)]
sysName = fit_logratios[,c(2)]
desc = fit_logratios[,c(3)]

IPT1=fit_logratios[,c(4)]
IPT2 = fit_logratios[,c(5)]
IPT3 = fit_logratios[,c(6)]

cntrl1 = fit_logratios[,c(7)]
cntrl2 = fit_logratios[,c(8)]
cntrl3 = fit_logratios[,c(9)]

exp1 = fit_logratios[,c(10)]
exp2 = fit_logratios[,c(11)]
exp3 = fit_logratios[,c(12)]

a <- data.frame(IPT1, IPT2, IPT3)
IPTfit <- rowMeans(a, na.rm=TRUE) #average each replicate

b <- data.frame(cntrl1, cntrl2, cntrl3)
cntrlfit <- rowMeans(b,na.rm=TRUE)

c <- data.frame(exp1, exp2, exp3)
```

```
exfit = rowMeans(c,na.rm=TRUE)

expfit1 = exp1-cntrl1
expfit2 = exp2-cntrl2
expfit3 = exp3-cntrl3

d <- data.frame(expfit1, expfit2, expfit3)
ExpFit = rowMeans(d,na.rm=TRUE)

#Create dataset for analysis

fitness = data.frame(locusID, sysName, desc, IPT1, IPT2, IPT3, IPTfit, cntrl1, cntrl2,
          cntrl3, cntrlfit, exp1, exp2, exp3, exfit,expfit1, expfit2,expfit3,ExpFit)

#Generate Initial WT+flg22 Plots

plot(cntrlfit)
abline(reg1)

#create dataset including WT+flg22 replicates and experimental replicates

fit = fitness[,c(19)]
fit1 = fitness[,c(16)]
fit2 = fitness[,c(17)]
fit3 = fitness[,c(18)]
ID = fitnessWT[,c(1)]
sysName = fitnessWT[,c(2)]
desc = fitnessWT[,c(3)]

data = data.frame(ID,sysName,desc,cntrl1, cntrl2, cntrl3, fit1, fit2, fit3, fit)
head(data)

#Set Threshold

a = data[(data[,c(10)]>=0.5),]
head(a)

#write to excel file

library("xlsx")
write.xlsx(a, file = "AllMF376MS_fitness.xlsx", sheetName = "a", col.names = TRUE,
          row.names = TRUE, append = FALSE)

#Count significant genes and find commonalities between different datasets
#Each variable (A, B, C, D, E, G) is arbitrary
```

```
A <- read.xlsx("1MF376sid2flg22_fitness.xlsx", sheetName = "a", colnames<- "TRUE")
B <- read.xlsx("1MF376sid2_fitness.xlsx", sheetName = "a", colnames<- "TRUE")
C <- read.xlsx("1MF376WT_fitness.xlsx", sheetName = "a", colnames<- "TRUE")
D <- read.xlsx("1MF376fls2flg22_fitness.xlsx", sheetName = "a", colnames<- "TRUE")
E <- read.xlsx("1MF376fls2_fitness.xlsx", sheetName = "a", colnames<- "TRUE")
G <- read.xlsx("1MF376MS_fitness.xlsx", sheetName = "a", colnames<- "TRUE")

a <- C[,c(1)]
b <- D[,c(1)]
c <- A[,c(1)]
d <- B[,c(1)]
e <- E[,c(1)]
g <- G[,c(1)]


a = intersect(#any two of the above)
vector.list<- list(#any combo of a,b,c,d,e)
b = Reduce("intersect",vector.list)
```