ENHANCING TRANSCRIPTION IN *ESCHERICHIA COLI* BW25113 AND *PSEUDOMONAS PUTIDA* KT2440 USING BACTERIOPHAGE LAMBDA ANTI-TERMINATOR PROTEIN Q

A Dissertation

by

JIBRAN AMANULLAH KHAN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Katy Kao
Co-Chair of Committee,	Steve Lockless
Committee Members,	Deborah A. Siegele
	Arul Jayaraman
Head of Department,	Thomas McKnight

May 2021

Major Subject: Biology

ABSTRACT

Functional characterization of metagenomic DNA often involves expressing heterologous DNA into genetically tractable microorganisms such as *Escherichia coli*. Expression of heterologous genes can suffer from limitations due to lack of recognition of foreign promoters or presence of intrinsic terminators on foreign DNA downstream of both the vector-based promoter and the transcription start site. Anti-terminator proteins are a possible solution to overcome this limitation. When bacteriophage lambda infects *E. coli*, it relies on the transcription machinery of the host to transcribe and express the phage DNA. Lambda genome encodes two anti-terminator proteins, namely N and Q (λQ), which regulate the expression of late early and late genes of the phage lambda, respectively. *E. coli's* RNA polymerase recognizes the P_R' promoter on the lambda genome and forms a complex with bacteriophage lambda anti-terminator Q, to overcome the terminator t_R'. Here we show the use of protein Q to efficiently transcribe a gene cluster containing intrinsic terminators from *Lactobacillus plantarum* in *Escherichia coli*. In addition, we expand the use of anti-terminator protein Q by showing it can function in *Pseudomonas putida* KT2440. The results show higher expression of a fluorescent reporter located ~12.5 kbp downstream from the promoter, when the transcription is driven by P_R' promoter in presence of protein Q.

DEDICATION

I dedicate this dissertation to my family who has been a constant support and backbone that took me through my graduate school. I would also like to dedicate this work to my friends back in India and the one's I was fortunate to make in the USA.

ACKNOWLEDGEMENTS

I would like to thank my committee chair and co-chair, Dr. Katy Kao and Dr. Steve Lockless and my committee members, Dr. Deborah Siegele and Dr. Arul Jayaraman, for their guidance and support throughout the course of this research. I especially thank Dr. Katy Kao for accepting me as her student at a time when I was searching for a challenging and industrial oriented opportunity in midst of my graduate school. Her constant feedback and support helped me to stay on track and achieve my goals in a timely manner. I thank Dr. Steve Lockless for an overall experience at my time in his lab. I enjoyed being a part of his lab where I learned a lot of new things about research and time outside of it. Most importantly, I made connections in his lab which will last forever with me. I thank Dr. Deborah Siegele for always being there when I had questions or two and always went to great extent to improve my understanding and clarify concerns. I thank Dr. Arul Jayaraman, for his support throughout these years, especially during the pandemic when he made it possible for me to continue my dissertation work in the department and provided all the necessary paperwork for it. A special thanks to Dr. Xiaorong Lin who served on my committee previously and Dr. Qing Sun to fill in as my committee during my prelims.

Thanks also go to the department faculty and staff for making my time at Texas A&M University a great experience. Special thanks to Dr. Arne Lekven and Dr. Rene Garcia for serving as my graduate advisors, Dr. Thomas McKnight the department head and Jennifer Bradford for coordinating and keeping me on top of all the formalities in

the department. I would like to thank Dr. Christine Merlin and Dr. Ying Zhang for teaching me to use the RT-PCR machine in the Department of Biology.

Thanks also go to my friends which I was fortunate to make in my journey through graduate school. I would like to thank the Lockless lab, especially Dr. Shian Liu, Dr. Sarah Beagle and Katrina Hofstetter. They took me as their little brother and for not just being my awesome lab mates but for accepting me for who I am and making my life so much easier by just being around. I would like to thank members in the Kao lab, especially commend Dr. Avinash Godara, for surviving in my presence as his only lab mate for over two years. I would also thank my internship supervisor, Dr. Adam Guss and his lab for showing me a new path of research and making my experience at ORNL so much fun.

Finally, thanks to my mother and father for their constant support and patience and to my loving brothers and sister for their continuous encouragement.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Katy Kao (chair), Dr. Steve Lockless (co-chair), Dr. Deborah A. Siegele of the Department of Biology and Dr. Arul Jayaraman of the Department of Chemical Engineering.

The work in chapter 3 was in collaboration and under supervision of Dr. Adam Guss (ORNL).

All work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a Texas A&M university Presidents excellence T3 grant.

This work was also made possible in part by DOE's center for bioenergy innovation (CBI). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the College of Science at Texas A&M University.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

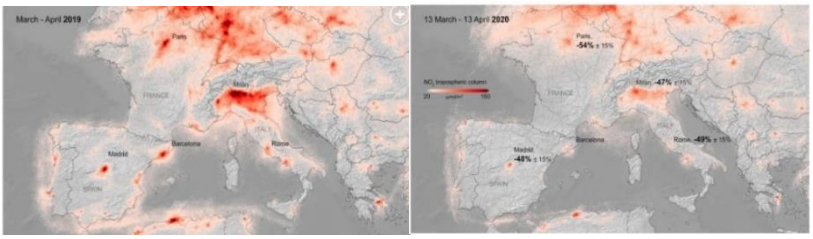
INTRODUCTION

The pandemic humankind is facing

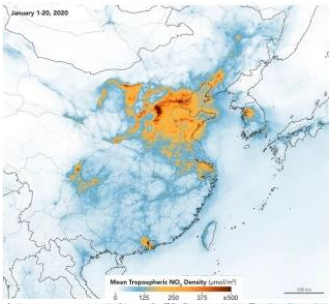
The Earth is facing an unprecedented time as humans enter a new decade in the 21st century. The coronavirus disease 2019 (Covid-19) pandemic has hit every possible inhabitable land on the planet. In its unbiased attack, it has crippled all countries including developed, developing and under-developed. The fear of the highly contagious infection has forced most countries to resort to nation or state-wide lockdowns or shelter-at home orders. Because of these dire times, impact of various human activities on the environment has garnered limelight. Air quality is one area where such impact is distinctly visible.

Various agencies have reported a drop in nitrogen dioxide (NO_2) levels in various parts of the world, an effect owing to lockdowns (Figure 1). Nitrogen dioxide (NO_2) is primarily emitted from burning fossil fuels [5] for transportation and electricity generation and is an indicator for air pollution (https://airquality.gsfc.nasa.gov/)

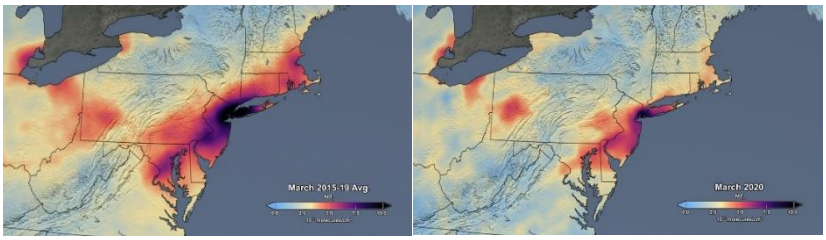
a) Europe



c) China



b) U.S.A



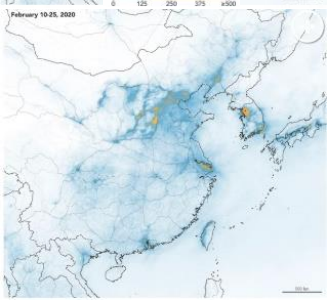


Figure 1 Satellite images taken pre and during pandemic.
The satellite images indicate the levels of NO_2 from different parts of the planet; a) Europe; b) China and c) U.S.A. As indicted in the images, NO_2 levels detected were low at the end of the first quarter of 2020 compared to the first quarter of 2019, an effect possibly due to lock downs/limited activities in various regions. Image courtesy: The European space agency (a) and NASA (b and c).

NO_2 belongs to the group of nitrogen oxide (NO_x) gases, which falls under the EPA's pollutant category (https://www.cdc.gov/air/pollutants.htm). NO_2 is usually recycled by plants using the nitrogen cycle. However, fossil fuel combustion releases excess NO_2 into the air, which the plants cannot keep up with.

Like NO_2, other greenhouse gases are emitted by combustion of fossil fuel derived products. For example, excessive carbon released by combustion of fossil fuels elevates the existing CO_2 in the environment. Given the negative impact on the environment, it is harmful to burn fossil fuel. The consequences of recent lockdowns and

reduction in activities leading to reduced greenhouse gas emissions, clearly indicate an urgency to divert our attention to environmentally friendly alternatives such as biofuels. Recent improvements in producing biofuels involves using unused farm waste as substrates (feedstocks). Primarily, lignocellulosic biomass when hydrolyzed provides cellulose which is broken down and used as a source for production of biofuels [6]. Studies predict a promising 19-48% lower greenhouse gas emissions when compared to gasoline combustion [7,8]. Also, the carbon released by burning of biofuels can be recycled and reused by plants during photosynthesis resulting in no net increase of carbon in the air. Various models developed by researchers, have predicted benefits of using biobased fuel, such as bioethanol, biodiesel, etc., over gasoline extracted from fossil fuels [7,9]. Lignin, a phenol polymer, is a component from the lignocellulose hydrolysis which is under-utilized compared to cellulose [10]. Lignin contributes to one third of the lignocellulosic biomass when separated from cellulose and hemicellulose component and is primarily used to generate energy [10]. However, the separation process yields more lignin than is needed for generating energy. Hence, an alternate strategy is required to use lignin in manufacturing processes to boost the bio-based economy [11]. Not many industrial microbes can naturally degrade lignin due to lack of enzymes that can convert the phenol polymer to consumable intermediates. However, recent efforts have been made to utilize these unused components of lignin from the biomass to extract energy [12] . Due to these limitations, production of biofuels on a larger scale also faces challenges [9]. The process in biofuel industry requires refinement to completely replace petroleum-based gasoline production.

The introduction of biotechnology in industrial processes aims to reduce toxic emissions to generate the final yield of the product, but to also considerably reduce the cost of the overall process [13,14].

Biotechnology industries are rigorously applying resources towards making the manufacturing process more robust and environmentally friendly [15]. One of the areas where the bio-based industries are focusing their attention is the efficient extraction of biofuels or chemicals by applying synthetic biology techniques to genetically engineer microbes. Back in 2010, ExxonMobil contributed ~$0.5 billion to Synthetic genomics, Inc. The investment was made to engineer algae to produce biofuels. Algae do provide a possible alternative as a chassis for biofuel production, given that they can survive in salt-water and produce fuels from atmospheric CO_2 and sunlight [16]. However, the production is not feasible yet, given the large cost associated with establishing algal farms and with the extraction process. On the same line, James Liao proposed engineering a photosynthetic cyanobacteria to make engine-compatible fuels, such as butanol, to avoid the refinement of motor engine [16]. However, the pitfalls associated with using cyanobacteria are similar to those for using algae for producing biofuels. Cyanobacteria are grown in ponds and need photons from sunlight. The penetration of photons is barely 10 cm into the ponds where the microbes will be grown. To be successfully used as an alternative, it will require millions of acres of space to cultivate microbes [16]. Due to issues with increased cost in establishing farming area with algae and photosynthetic microbes, an alternative is to look for non-photosynthetic bacteria for production of biofuels.

Other bio-based industrial applications of using bacteria for production

Use of bacteria in industrial applications is not a novel concept. *Rhodococcus* and *Pseudomonas* species have widely been used for bioremediation of oil spills [17]. The manufacturing of surfactants from petrochemicals emit carbon dioxide levels equivalent to the amount of carbon dioxide released by the combustion of 3.6 billion gallons of gasoline annually [18]. To address this issue, Modular genetics Inc., engineered microbes that convert soybean hulls into surfactants for use in personal care products [18]. Bacteria have also been used to produce antibiotics. Poet-DSM, as a company, is a good example of using synthetic biology to enhance the production of the synthetic antibiotic, Cephalexin. They also came up with a mixture of enzymes that are capable of breaking down the recalcitrant lignocellulosic biomass from agricultural residues to simple five carbon and six carbon sugars [18]. Other beneficial application, is the use of the Gram-negative bacteria *Lactococcus lactis* to secrete interleukin 10, which is used to treat Crohn's disease [19,20]. With the growing need to achieve efficiency, biotechnology is making strides in developing biocatalysts with superior performance or tolerance for application-specific conditions [21]. Due to the demand of enzymes in industries, there has been an increased interest in identifying new enzyme variants which enhances kinetic parameters and overall process of manufacturing.

Exploring the environment as a source of novel traits and strains

Micro-organisms are ubiquitous [22,23]. Prokaryotes represent the largest proportion of living organisms (estimated that there are 10^{30} microbial cells on earth [24,25] with an estimated 10^{3}–10^{5} microbial species in 1 g of soil [26]), and less than 1% can be cultured using existing methodologies [22,27-29]. Despite such a low percentage, the likelihood of identifying novel traits from uncultured environmental organisms is largest. If we assume that a single genome encodes 4000 proteins (as is the case with the bacterium *Escherichia coli*), then 4×10^{8} potential proteins might be expected in just 1 g of soil. Assume that 40% of these proteins display catalytic activity [30], we might expect to find 1.6×10^{8} biocatalysts, which highlights the vast inventory of biological functions that nature has to offer [21]. There are so many beneficial genes/clusters and strains that have been identified to date. For example, proteorhodopsin-containing bacteria are proposed to have an impact on the carbon and energy balance in the ocean [31-33]. Likewise, microbial exploration revealed a novel Na^{+}/H^{+} antiporter [34], strains that express lipases and amylases [35], as well as genes that confer tolerance to ligno-cellulose derived inhibitors such as alcohols, aldehydes and organic acids [36]. Nearly half of commercially available pharmaceuticals products are from bacteria, primarily from those that can be cultured and studied.

What is the underlying issue in exploring the uncultivable strains in labs? Due to the inability to cultivate pure culture of bacterial species from environment, they cannot be characterized further to identify beneficial pathways. It may be difficult to culture bacteria from nature for many reasons, such as lack of necessary nutrients or surfaces,

interdependencies between members of the community, accumulation of excess toxic waste products from native metabolic pathways, sub-optimal physical parameters such as pressure, temperature, atmospheric gas composition and variable doubling time in pure culture vs in ecological biography (131).

Carl Woese and George Fox pioneered the use of rRNA genes as markers for phylogenetic analysis [37]. Building on that work, Pace and colleagues laid the foundation for a new branch of microbial ecology [38,39], by using 5S and 16S ribosomal RNA (rRNA) as evolutionary markers in early 1980's [40-43]. Earlier, there were limitations using this approach, since not many genome sequences were available to predict the organism using a template and the cost associated with the sequencing. However, with a drop in cost to sequence a genome, there has been an increased interest in identifying novel strains by sequencing more and more genomes to further characterize novel species. The need to culture non-trivial organisms can be overcome by directly isolating genomic DNA from natural sources and cloning it in standard laboratory organisms [44]. Since the 1980's, extensive work has been done to identify genes of uncultivable strains from soils, water reservoirs, forests, and microbiomes of animals. In addition, uninhabitable places, such as acid mine drainage areas, Yellowstone springs, and the Sargasso sea ecosystem, were also explored for novel strains [44]. In 1998, realizing the benefit of using 16S rRNA sequencing to discover uncultivable microbes, Handelsman and her lab coined the term 'metagenomics' [45]. Metagenomics means the study of a genome acquired from environmental sources. Metagenomics avoids the necessity of

culturing a microbe in laboratory conditions. In the last two decades, metagenomics has enabled access to a biological and molecular biodiversity in natural environment [21].

Types of metagenomics

Metagenomics is broadly classified into 2 types: traditional or sequence-based metagenomics and functional metagenomics [21]. In the traditional approach, a novel strain is computationally predicted to belong to a genus based on conserved domains from the 16S *rRNA* database of a phylogeny. Similarly, function of a novel gene is discovered by aligning conserved domains in the database. Since the traditional approach is based on mining existing DNA database, identification of novel strains or genes can be restricted. A promising application of phylogenetic anchor-guided sequencing is to collect and sequence many genomic fragments from one taxon. The alternative to a phylogenetic marker-driven approach is to sequence random clones. The distribution and redundancy of functions in a community, linkage of traits, genomic organization, and horizontal gene transfer can all be inferred from traditional sequence-based analysis [44]. However, this approach also has drawbacks. The use of phylogenetic markers as the initial identifiers to indicate taxonomic affiliation is limited by the small number of available markers that provide reliable placement in the tree of life.

In contrast, functional metagenomics is about discovering novel strains or genes/pathways based on the phenotype of interest. Functional metagenomics is where DNA is acquired from environmental sources and cloned into a laboratory strain. After cloning it into a suitable host, the DNA library is expressed in a heterologous host to

identify a phenotype of interest. *Escherichia coli* (*E. coli*) is the primary choice as a host for functional metagenomic studies [46], as a lot of genetic tools to modify the genome already exist. In addition, the complete genome of *E. coli* is annotated which helps in altering the genome to fit the type of functional screening of heterologous DNA.

Beneficial outcomes of functional metagenomics

Initial work in functional metagenomic screened for halo formations on agar, indicating the presence or absence of enzymes such as lipase and amylase [35,47]. Esterases are quite beneficial in the food and detergent industries and other bioremediation processes. Novel esterases, functional at 7^{o}C, were identified using a metagenomic library acquired from Antarctica desert soil [48]. Similarly in 2005, Rhee et al. were able to identify a novel esterase from a Solfataric hot springs, which was stable at high temperature. Tetrazolium chloride dye was used as an indicator by Henne et al. [49] to identify 4-hydroxybutryate utilizing clones. Even though useful, such screening can only be applied if the objective is to breakdown a known metabolite in the growth media. It was rarely fruitful to identify other biological activities such as novel pathways that lead to an anabolic activity [24]. Some modified selection strategies involved knocking out a gene in the host strains and then performing a complementation assay by using a metagenomic library to rescue the deficiency. Applying this strategy, labs have discovered novel Na^+ (Li^+ /H^+)/H^+ antiporters [34], a novel DNA polymerase [50] and biotin [51] producing pathways from the genomic library. Another widely used strategy is to expose the metagenomic-library-carrying host to inhibitory concentrations of a particular

toxin or metabolite, such as antibiotics [52-58]. Such strategies were used to identify novel aminoglycoside resistant clones, which encoded a new cluster of 6'-N-acetyltransferases [52]. A combined PCR-based approach followed by a functional assay aided in identifying antibiotics such as terragine [59]; turbomycin A and turbomycin B [60]; erythromycin and rifamycin were first cloned from soil with a PCR based approach [61].

Cellulase, an enzyme required for breakdown of cellulose, is found in many microbes. Cellulase is used in many industries, including textile, food, and paper industries and has lately garnered a lot of attention in biobased butanol industries [62]. Various environmental samples, such as soil [63,64], hindgut of higher termites [65], rabbit cecum [66], compost [67], fresh cast from earthworm, cow and calf rumen [68-70], have been investigated to find novel cellulase encoding genes [21].

With the recent boom in the field of synthetic biology, scientists around the globe are trying to progress metagenomics by using genetic circuits. Recent interest has helped the development of various tools such as SIGEX [71], METREX [72], PIGEX [73], where the overall goal is to produce an output, mostly by reporting fluorescence once the circuit has received an input in the form of a desired substrate, metabolite or product, respectively. Using SIGEX, clones that were regulated by naphthalene, later identified as a novel nag2 operon, and benzoate were identified from a pool of 152,000 clones. Additionally, METREX was applied to identify N-acyl homoserine lactone (AHL) compounds from the gypsy moth gut, forest soil, and activated sludge from a coke plant [24].

Pitfalls of functional metagenomics

The field of metagenomics has brought a lot of promise by enabling exploration of various environmental sources. However, metagenomics is still a developing field and it is estimated that only 0.000001% or less enzymes encoded in a gene library can be detected by conventional functional screens [74]. Major limitations of using functional metagenomics lies in the need to express heterologous DNA efficiently.

Success of functional metagenomic depends on many factors such as, propagation of metagenomic DNA in a heterologous host, efficient transcription of the DNA by recognizing heterologous regulatory elements or vector-fused promoters, successful translation of mRNA and if needed post translational modification of proteins and suitable screening strategies for target genes. [13,24]. A major setback comes with the difficulty in using a single or a very limited pool of hosts. Currently, *E. coli* has been applied as a chassis to express DNA from diverse organisms [35,75-79]. However, it is not always an ideal choice as a host. *E. coli* is limited in its function and structure to support expression of heterologous DNA from diverse group. *E. coli* is estimated to express only 40% of heterologous proteins from a randomly cloned metagenomic library [80]. Part of the reason could be the vectors that are used to express the metagenomic DNA library inserts. For instance, in a transcriptional fusion, a strong promoter pre-exists on the vector, however the ribosome binding sites (RBS) for initiation of translation is expected to be provided by the inserted DNA. On contrary, transcription could itself fail to provide a functional mRNA if the DNA dependent RNA polymerase (RNAP) hits a terminator which exits prior to the RBS and coding sequence of a potentially interesting

gene [80]. Additionally, using non-optimal cloning vectors for making metagenomic libraries could result in lower success rate of identifying a beneficial phenotype. [21]. The work by Warren et al. [81] highlights this by expressing DNA from various bacterial species in *E. coli*. Based on transcriptional profile in the study, *E. coli* expressed genes of closely related *Haemophilus influenzae* better compared to a distant related *Pseudomonas aeruginosa*, emphasizing limitation of expressing DNA acquired from very distant species. Some promoters from phylogenetically similar species are more likely to be recognized by the host sigma factors compared to distantly related [82]. To overcome this limitation, various studies have also been conducted where sigma factors from different species are expressed in *E. coli* [83]. Few studies show expression of tRNA's in *E. coli* from diverse species to enhance expression at protein level [84-86] . These studies have been published and shown to improve transcription profile to certain extent. Thus, *E. coli* as a host can be genetically manipulated to enhance screening during functional metagenomic studies.

Scope of this dissertation

A drawback in transcribing foreign DNA is associated with recognition of foreign promoters by host transcription machinery. Therefore, while constructing a DNA library, usually a strong *E. coli* promoter is used to initiate the transcription process. Various promoters such as lac, tac and T7 are usually found in cloning vectors [87]. T7 promoter has been tested to be one of the strong promoters known. The promoter can only be recognized by T7 RNA polymerase (RNAP), which is orthogonally expressed in

E. coli and the T7 promoter is not recognized by the native *E. coli* transcription machinery [88]. To use T7 promoters, strains have been engineered to express T7 polymerase. Thus, one may use a cloning strain that already expresses the specific RNA polymerase or if the background strain of *E. coli* does not fit the purpose, then express the polymerase on a vector or integrate the gene encoding the polymerase on the genome. In addition, what is often over-looked is that one of the important phases in transcription is termination. If we consider, a phenotype of interest which depends on expression of multiple genes which are not in an operon but in a cluster within a locus separated by terminators. The expression of genes involved will not be an issue if all the promoters driving transcription are recognized by the foreign host. However, if the promoter is not efficiently recognized by the foreign host, and there is a recognizable promoter at the beginning of the DNA cluster, then the terminators could hinder identification of downstream genes for the phenotype of interest. Based on most studies performed on *E. coli*, termination occurs primarily in two ways: Rho dependent and Rho-independent termination. Rho dependent termination relies on protein Rho. Predicting termination based on Rho protein is not as trivial as in the case of Rho-independent terminators [89], due to lack of canonical structure such as in the case of Rho-independent terminators. Rho-independent or intrinsic termination occurs by secondary structures developed by the nascent mRNA transcript. The canonical structure comprises of a G-C rich palindrome sequence, which forms a stem-hair pin loop, followed by a tail of U's. The secondary structure destabilizes the RNA-DNA hybrid formed at the transcription bubble and prevents the read through by RNA polymerase resulting in

transcription termination [90]. Anirban et al. [91] came up with a genome scanner for terminators (GeSTer) to predict intrinsic terminators across genomes of 313 species [92-95]. After analyzing ~1 million genes from various species, they concluded that atleast 35.6% of the genes carry an intrinsic terminator downstream of them. Similarly, there are many tools developed by biologist to predict terminators, such as TransTermHP [95], Arnold [96], which have been tested to predict terminators with ~90% accuracy. This suggests that many microbes from various taxa of bacteria, employ the intrinsic terminator strategy for terminating transcription. Intrinsic terminators can pose an issue in identifying potential genes or pathways which require transcription of genes in different clusters separated by terminators.

A possible way of overcoming terminators is using an accessory anti-terminator protein which help RNAP transcribe past intrinsic terminators. The focus of this dissertation is to demonstrate the use of an anti-terminator protein Q from bacteriophage lambda [97] in overcoming terminators while transcribing heterologous DNA in a foreign host.

Bacteriophage lambda anti-terminator protein Q

The bacteriophage lambda (λ) harbors two transcription regulators that have been shown to play a role as anti-terminators, namely protein N (λN) and Q (λQ)[98]. The first evidence of λQ acting as a positive regulator to transcribe the late genes of the lambda genome was documented in early 70's [99]. However, the gene sequence of the λQ [100] and its role as an anti-terminator using in-vitro studies was shown in early 80's [101].

Subsequent studies show the role of the Q binding element (QBE) within the Q utilizing site (*qut*) for the attachment of λQ to the RNA polymerase - sigma 70 (σ70) holoenzyme [102-106]. The *qut* site consists of an additional -10-like pause-inducing sequence which is essential for pausing the RNAP and allowing time for the λQ to bind to the holoenzyme (Figure 2a) [105,107] . Also, the qut site of lambda contains a -35 like sequence which is required for the recruitment of λQ to RNAP [106,108].

Recruitment of the λQ to the RNAP is triggered during formation of the early elongation complex by the RNAP with σ70 and non-template DNA strand. The region 2 of the σ70 subunit then slides along the non-template strand, encounters the -10 like promoter sequence which pauses the holoenzyme, exposes part of the QBE which allows room for the attachment and binding of the λQ to the RNAP (Figure 2b). Pausing of the holoenzyme, at the -10 like sequence, is a critical stage for attachment of the λQ to the RNAP. Studies by Ko et al. show that mutation in region 2 of the σ70, leading to reduced pause at the -10 sequence, results in failure of RNAP to bind to the λQ, resulting in sensitivity to the terminator sequence [109]. Between the initiation and pause of the early elongation complex, the transcription machinery synthesizes 15-17 nucleotides of nascent RNA. Synthesis of nascent RNA promotes RNA-mediated destabilization where, interaction between region 4 of the σ70 and the β-subunit of the core RNA polymerase is dissolved (Figure 2c) [110]. Due to this triggered destabilization, λQ finds an opportunity to compete with region 4 of the σ70 to bind to the RNAP core complex. Finally, λQ binds to the β-component of the core RNA polymerase [4], and tags along during the elongation phase of transcription (Figure 2d). The key feature of this complex

is that the λQ-bound RNAP complex evidently reduces pausing at the -10 like sequence and is also capable of overcoming downstream terminators compared to a λQ-free RNAP complex [106].

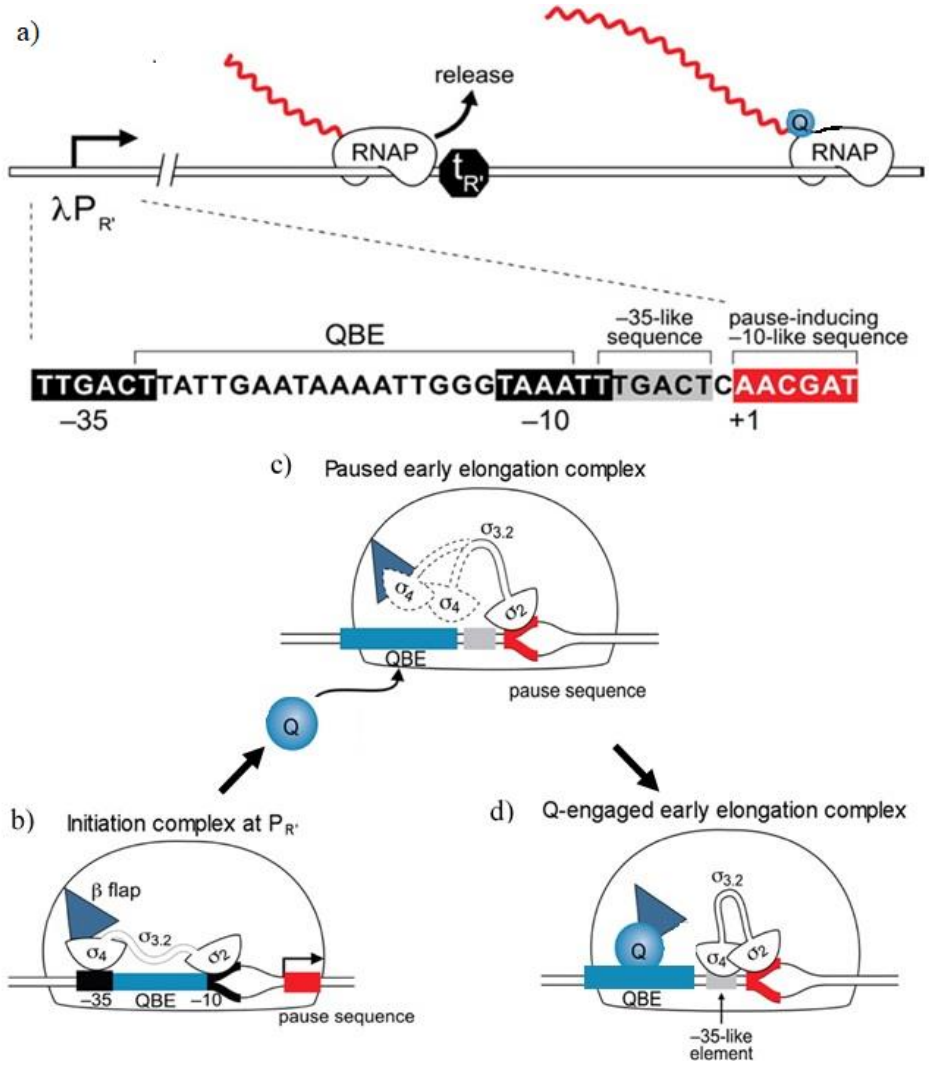


Figure 2 Mechanism of λQ binding to RNA polymerase to overcome the terminator t_R'. Image adapted from Deighan et al, 2008[4].
a) DNA fragment carrying the P_R' promoter required to form the RNAP and λQ complex. The complex helps in overcoming the terminator t_R' which terminates transcription in absence of λQ; b) regions 2 and 4 of sigma 70 (σ70) bind to the -10 and -35 elements of RNAP respectively; c) destabilization of region 4 of σ70 induces recruitment of λQ at the Q binding element (QBE); d) λQ bound to the β-flap region of RNAP.

Potential use of bacteriophage lambda protein Q on a metagenomic level

Studies were conducted using various reporters to report function of λQ in *E. coli* [4,102,104-106,111] . Ghosh et al., have shown that RNA polymerase complexed with λQ can overcome *rrnBT1-T2* terminators [112]. Other studies have reported a combination of either t_R, alone [102,104] and/or *E. coli*'s native terminator (*rrnBT1-T2*) to overcome and express a reporter gene [92,112]. In addition, work by Yang and Roberts, show use of Q^{82} from bacteriophage 82 to overcome a rho-dependent terminator, a feature widely studied using anti-terminator protein N [113].

Transcription termination using intrinsic terminators is conserved across bacteria [91]. A study conducted by Gabor et al., suggests that the average distance between a start codon of a coding sequence and preceding intrinsic terminator is ~15 kbps in bacteria. This piece of information is particularly important when a metagenomic library is expressed in a foreign host during functional screening. Perhaps a large cluster of metagenomic DNA acquired from soil, which has a good ratio of bacteria, may be difficult to transcribe completely and limit identification of beneficial or novel clusters due to the terminators interrupting the transcription process [80]. For instance, vectors are used to express the metagenomic library in a host. These vectors carry host-specific promoters upstream of the insertion site, so that success of transcription is not solely based on recognition of foreign promoters. However, if the inserted DNA carries an intrinsic terminator upstream of the transcription start site, it will hinder the efficiency of RNAP to transcribe genes downstream of the terminator. My work proposes the use of

bacteriophage lambda's anti-terminator protein Q to overcome such limitations that arise due to the presence of intrinsic terminators.

Specific aims

Metagenomics is the study of genetic material obtained from community of microbes found in the environment. A major boost to the metagenomics field started when techniques were developed that overcame the need of culturing and isolating individual members of the community [114-116]. These advances aided in studying communities which are not easy to culture under standard microbiological conditions. The ever improving high throughput DNA sequencing strategies have changed the dynamics of the metagenomics field [33,44]. Function based metagenomics is a method of constructing a genomic library of DNA fragments and expressing their functional proteins in foreign hosts to study the gene's characteristics. This approach has led to the identification of novel antibiotic resistance genes, degradative enzymes, and more [44]. Although the approach has been quite useful, there are challenges that need to be resolved in order to utilize this approach to its potential. Not all genes extracted from the environmental strains are expressed in foreign hosts, limiting the identification of other potential genes or proteins. This limitation could be partly due to unrecognizable promoters on foreign DNA. The issue with foreign promoters can be overcome by driving the expression of the inserted foreign DNA using a host promoter. However, phenotypes which rely on expression of multiple genes not in an operon may not benefit from a host recognized promoter, since terminators in between the clusters of genes could lead to pre-mature

termination of transcription. The long-term goal is to develop an improved strategy to construct a metagenomic library to identify genetic determinants associated with biosynthesis of desired compounds or phenotypes of industrial relevance. The overall objective of this proposal is to develop a new strategy, primarily involving bacteriophage lambda anti-termination protein Q to overcome internal intrinsic terminators in heterologous DNA. The function of λQ will be tested in two different host backgrounds, *Escherichia coli* and *Pseudomonas putida* KT2440. To test its application, the developed strategy will be used to identify potential pathways in a library created from environmental isolates to enhance the metabolism of industrial strains. In order to achieve our goal, we propose two aims which are as follows:

AIM 1: To enhance the expression of heterologous DNA with intrinsic terminators in Escherichia coli *using bacteriophage lambda anti-terminator protein Q*

Functional screening of metagenomic library is often conducted in a laboratory standard host such as *E. coli*. However, *E. coli's* transcription machinery is shown to recognize foreign promoters from closely related species better than distantly related species [82]. In addition, transcription could potentially terminate pre-maturely if a foreign DNA carries internal terminators. Thus, in order to assist in the transcription process, we aim to use bacteriophage λQ. We aim to overcome the barriers due to internal transcription terminators in foreign DNA, which are a potential issue for identifying a cluster of genes separated by intrinsic terminators. To test this we construct a synthetic cluster from *Lactobacillus plantarum*, which carries native intrinsic terminators and a native *E. coli*

terminator *rrnBT1*. The expression of cluster is driven by either a lac promoter or P_R' promoter in presence or absence of λQ. We *hypothesize* that in presence of λQ, expression of genes downstream of the intrinsic terminators will be higher compared to expression of cluster driven by lac promoter.

AIM 2: To enhance the expression of heterologous DNA with intrinsic terminators in Pseudomonas putida *using bacteriophage lambda anti-terminator protein Q*

Functional metageomics uses *E. coli* or other hosts based on the phenotype screening under study. Intrinsic terminators are not just a feature associated with *E. coli*, but rather it is distributed across the bacteria domain [91,94]. Thus, use of an anti-terminator protein in others hosts beyond *E. coli* could also prove beneficial for metagenomic studies. As far as we know, bacteriophage lambda anti-terminator protein Q has only been tested in *E. coli* and never in any other backgrounds. In this study, we aim to test the function of anti-terminator λQ in *P. putida* KT2440. We *hypothesize* that the P_R' promoter will be recognized by *P. putida* KT2440's RNAP and that λQ will help in overcome intrinsic terminators that could hinder transcription of a heterologous DNA.

CHAPTER II

EXPANDING THE USE OF LAMBDA PROTEIN Q IN *ESCHERICHIA COLI* BW25113 FOR EXPRESSION OF HETEROLOGOUS DNA

Introduction

There is a desire to shift from petrochemical-based to bio-based approaches to manufacture fuels and chemicals [117,118]. Recent bio-based processes involve use of farm wastes and lignocellulosic biomass as substrates for production. However, existing limitations such as additional cost to breakdown these recalcitrant feedstocks and inability of the cell to utilize all the depolymerized substrate undermine the efficiency of bio-based processes. Pre-treatment of lignocellulosic biomass is a necessary step before adding the substrate to bioreactors, since industrially relevant microbes are not efficient at utilizing the feedstocks in their hetero-polymer form. Various physio-chemical pre-treatment steps generate inhibitory by-products that have a negative impact on cellular function due to lack of detoxification mechanism in the host, resulting in lower productivity [36]. These challenges give rise to a need for identification of novel strains or genes that can aid in efficient production of desired products. For example, increase in yield can be achieved by enhancing the tolerance of production hosts to toxic by-products by engineering pathways to convert them to consumable metabolic intermediates. Strains capable of breaking down these recalcitrant substrates have been found in nature [6,119,120]; the genes responsible for these desired phenotypes can

potentially be engineered in production hosts to improve their tolerance and eventually cut down on cost for production.

Functional metagenomics has been used to identify novel genes or metabolic pathways that could possibly be engineered in microbial hosts to produce industrially relevant products [34-36,121,122]. Typically, a model organism – such as *Escherichia coli* (*E. coli*), is used to express the metagenomic library. However, regardless of the host used, potential limitations such as inability of the host machinery in recognizing heterologous promoters [82] or the presence of intrinsic terminators on inserted DNA between vector-provided promoters and transcription start site can existing, resulting in operons or genes going unidentified [80].

Intrinsic and factor dependent terminators (Rho-dependent in *E. coli*) are widely distributed in bacterial genomes. Based on analysis of ~300 bacterial genomes, ~35% of genes are predicted to be terminated using intrinsic terminators [91]. The role of an anti-terminator protein is to help DNA dependent RNA polymerase (RNAP) to overcome terminators arising in gene clusters, hence aiding in expression of the downstream genes [90]. Anti-terminator proteins are known to be encoded in many bacteriophage genomes [4,113]; the bacteriophage lambda genome encodes two anti-terminators N and Q [90]. The anti-terminator λQ recognizes a Q binding element (QBE) segment on the P_R' promoter which facilitates attachment to a paused RNAP [4,103,106,110]. The RNAP bound to the λQ has previously been shown to overcome intrinsic terminators, such as *rrnBT1*-T2, during transcription (Figure 3) [4,106,112].

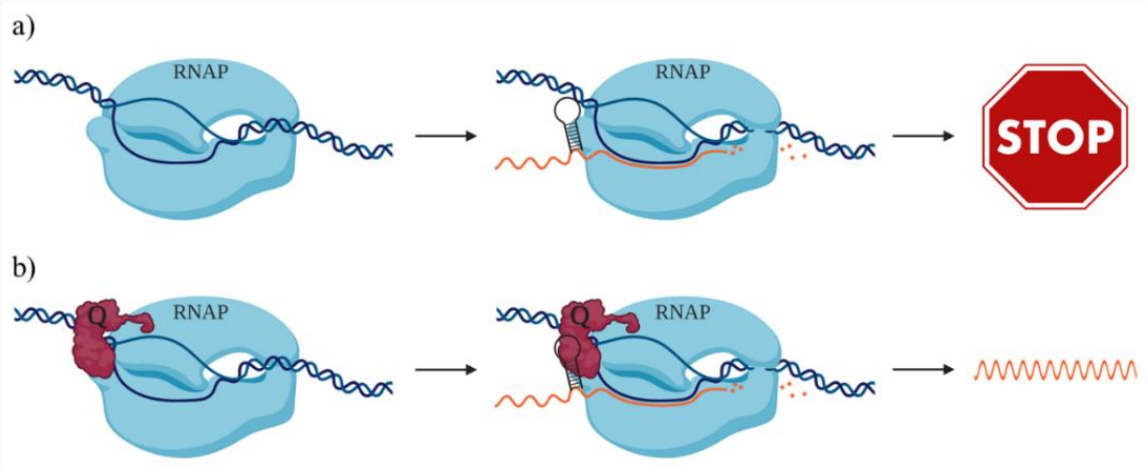


Figure 3 Illustration on how anti-terminator Q protein works.
a) The transcription of mRNA (orange) terminates when RNAP encounters a terminator sequence while the mRNA is synthesized. b) RNAP overcomes the terminator and results in complete synthesis of the mRNA, in presence of protein Q.

The goal of this study was to test the use of the bacteriophage lambda anti-terminator protein Q to overcome a widely used *E. coli* intrinsic terminator as well as a predicted terminator to transcribe heterologous DNA from the *L. plantarum* genome. The *hypothesis* was that in presence of the anti-terminator λQ, the expression of heterologous genes downstream of an intrinsic terminator would significantly increase when compared to its absence in the cell. The expression levels of mRNA, using qRT-PCR studies in *E. coli* clearly indicate higher expression of the gene downstream of the intrinsic terminators in presence of λQ.

Results

Expression of gene downstream of terminator (cps3I) does not increase when lac promoter (P_{lac}) is induced with IPTG in E. coli

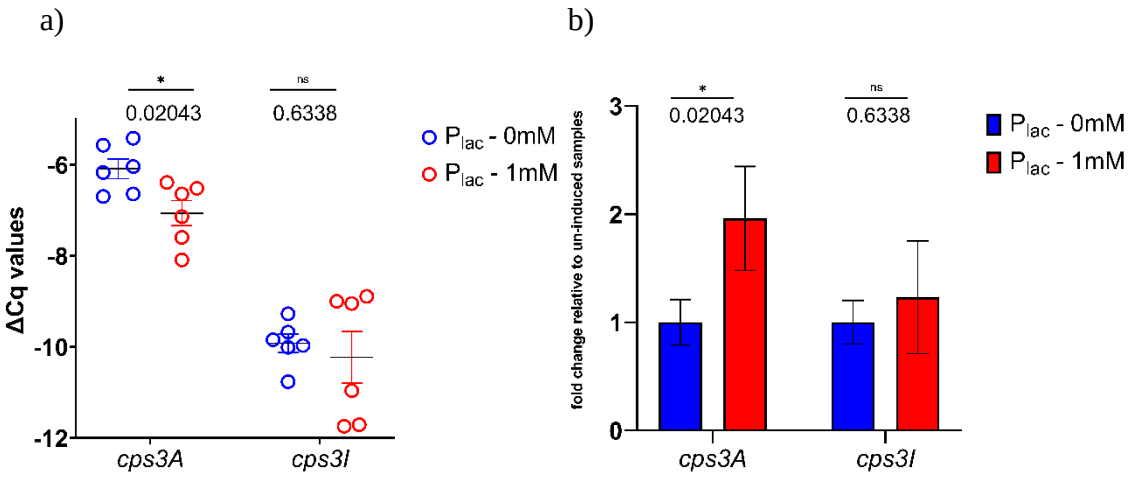


Figure 4 Expression of *cps3I* is not increased upon IPTG induction with the lac promoter. Blue and red circles/bars represent the expression levels detected in strain which was not-induced (P_{lac}-0mM) or induced (Pl_{ac}-1mM) with 1mM IPTG, respectively. **a** ΔCq values represented in a scatter plot for all the replicates. The circular symbols represent Cq values of *cps3A* or *cps3I* normalized to housekeeping gene *hcaT*. **b** Fold change expression using the $2^{-\Delta\Delta Ct}$ method of gene *cps3A* and *cps3I* normalized to reference gene *hcaT*, relative to the expression of the un-induced strain (P_{lac}-0mM). The Cq values collected during the qRT-PCR study were from duplicates of three biological replicates for both strains under study. The error bars represent standard error taking into account propagation of error. The p-values were calculated using a two-tailed student's t-test where the asterisk represent: NS = non-significant; *P<0.05.

To determine the effectiveness of the *E. coli* RNAP to overcome transcriptional terminator, a synthetic gene cluster was constructed from the cps3 gene cluster of *Lactobacillus plantarum*. The synthetic gene cluster contains two internal transcriptional terminators, *rrnBT1* and a predicted internal terminator (Figure 6). As control, the cps3 gene cluster was expressed under the lac promoter on plasmid pJKK2 and used to test the efficiency of *E. coli* RNAP to overcome intrinsic terminators. The expression levels of the first and last genes (*cps3A* and *cps3I*) from the modified cps3 cluster in the presence or absence of IPTG induction were measured using qRT-PCR (Figure 4). Results showed an increase in expression of *cps3A* but not of the downstream *cps3I* in the presence of 1mM IPTG induction, suggesting that induction of the lac promoter was not sufficient to overcome transcriptional termination by the *rrnBT1* terminator.

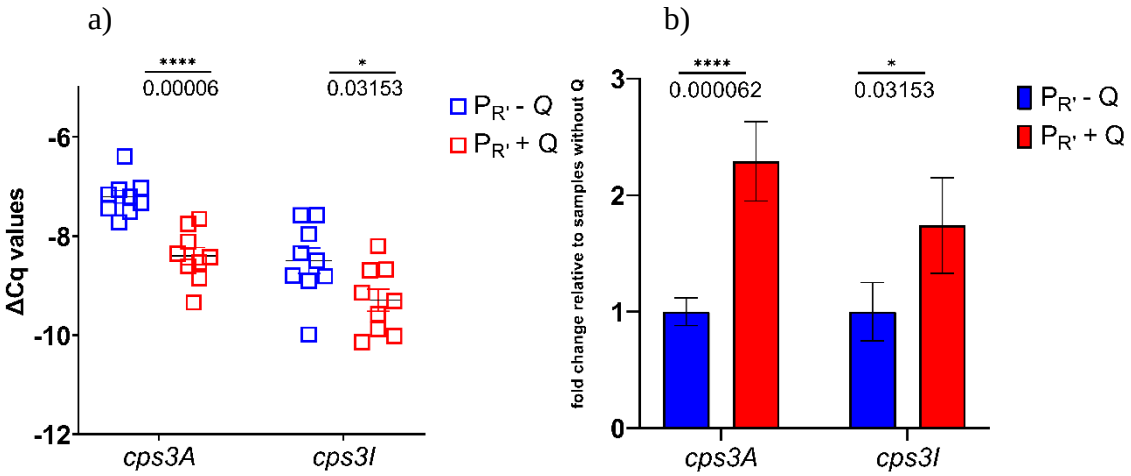


Figure 5 Expression of *cps3A* and *cps3I* increases significantly in presence of Q protein. Blue and red circles/bars represent the expression levels detected in strain which was un-induced (PR-Q) or induced (PR+Q) with 1mM arabinose, respectively. **a** ΔCq values represented in a scatter plot for all the replicates. The circular symbols represent Cq values of *cps3A* or *cps3I* normalized to housekeeping gene *hcaT*. **b** Fold change expression using the $2^{-\Delta\Delta Ct}$ method of gene *cps3A* and *cps3I* normalized to reference gene *hcaT*, relative to the expression of the un-induced strain (P_R'-Q). The Cq values collected during the qRT-PCR study were from triplicates of three biological replicates for both strains under study. The error bars represent standard error taking into account propagation of error. The p-values were calculated using a two-tailed student's t-test where the asterisk represent: *P<0.05 and ****P<0.0001.

To determine the effectiveness of the *E. coli* RNAP complex to overcome transcriptional terminators in presence of λQ, the P_R' promoter was used to express the synthetic *cps3* gene cluster on plasmid pJKK3. The gene encoding λQ was integrated on *E. coli* genome under the control of the arabinose promoter (P_{BAD}). A strain containing pJKK3 without the λQ was used as negative control. The expression of the last gene after the internal terminators, *cps3I,* was statistically significantly higher ($p<0.05$) in the presence of λQ compared to the control strain that did not express λQ (Figure 5). This suggests that the λQ helps to stabilize the RNAP-DNA complex to help overcome the intrinsic terminators during transcription. We also observed the expression of the first gene, *cps3A*, increased in the presence of λQ compared to the negative control, suggesting that λQ expression enhances transcription from the P_R' promoter.

Discussion

Previously, levels of β-galactosidase activity were reported to demonstrate the efficiency of RNAP-Q complex to read-through terminators such as *rrnBT1* [4,104]. Hence, we tested the efficiency of λQ to overcome a widely used *E. coli* terminator, *rrnBT1* [112,123], in order to transcribe capsular polysaccharide cluster (cps3) from *Lactobacillus plantarum*. We use qRT-PCR to report expression levels of heterologous genes, downstream and upstream of a terminator, to show efficient transcription by RNAP when complexed with the anti-terminator λQ. In addition, we compare the expression levels of genes when a lac promoter (P_{lac}) is replaced with P_R' promoter.

As predicted, in the control experiment (Figure 4 and Table 1), there was an increase in expression level of gene *cps3A* when P_{lac} was used. However, the results suggest that the RNAP is not stable when it encounters the *rrnBT1* terminator as indicated by the non-significant difference observed in the expression levels of gene *cps3I,* which is downstream of the terminator, in samples which were induced by IPTG when compared to un-induced samples. In presence of λQ, expression levels of *cps3A* are significantly higher when compared to strain that do not express λQ (Figure 5 and Table 1). In addition, there is a significant difference observed in expression levels of *cps3I* from samples that express λQ when compared to samples that do not. This suggests that the RNAP overcomes the terminator when bound to λQ. Also, the variation seen between biological replicates of samples with λQ is lower than samples with P_{lac} that were induced with IPTG in the control experiment (Figure 5), suggesting stable RNAP-Q complex that efficiently reads through terminator *rrnBT1*.

Table 1 Statistical analysis of the qRT-PCR data.
The p-values were calculated using the log transformed normalized expression relative to un-induced samples.

Sample	Gene	ΔΔCq relative to un-induced samples	p-value
P_{lac} - 0mM	*cps3A*	1 (±0.21)	
	cps3I	1 (±0.20)	
P_{lac} - 1mM	*cps3A*	1.96 (±0.48)	2.04E-02
	cps3I	1.23 (±0.52)	6.34E-01
$P_{R'}$ - Q	*cps3A*	1 (±0.12)	
	cps3I	1 (±0.25)	
$P_{R'}$ + Q	*cps3A*	2.29 (±0.34)	6.16E-05
	cps3I	1.74 (±0.41)	3.15E-02

For this study, we used the native bacteriophage lambda Q to overcome a strong terminator *rrnBT1* in *E. coli*. However, intrinsic terminators in bacteria stop the transcription process with variable efficiencies owing to characteristics, such as stem loop length, poly-U trail, etc. [124,125]. Also, variant of λQ has been created which is shown to increase the anti-termination efficiency [4]. Hence, it will be interesting to test the effectiveness of λQ and its variants to overcome intrinsic terminators that vary in termination efficiency.

Methods

Strains and growth media

Refer to the Table 6 (Appendix A) for genotypes of the strains mentioned in this section. All plasmids are maintained in *E. coli* DH5alpha except the pLA2 vector series which are replicated in *E. coli* BW25141. All mRNA expression analysis, including genome modifications to integrate the λQ encoding gene, were performed in *E. coli* BW25113 strain.

Luria-Bertani (LB) broth/agar (10% tryptone; 5% yeast extract; 8% NaCl; 1.2% agar for plates) was used for culturing all the *E. coli* strains during the cloning procedure. *Lactobacillus* MRS broth (Himedia) was used to culture *Lactobacillus plantarum* for genomic DNA extraction. M9 minimal media was used for the growth, mRNA expression and fluorescence assay studies. The recipe used for 50 mL of M9 minimal media is as follows: 10 mL of 5X M9 salts (1 L of 5X M9 salts: 64 gms of $Na_2HPO_4.H_2O$, 15 gms of KH_2PO_4, 2.5 gms of NaCl, 2.5 gms of NH_4Cl); 5µl of 1 M $CaCl_2$; 100 µl of 1 M $MgSO_4$; 50 µl of 1000X metal ion mix; 400 µl of 40% glucose; 5 µl of thiamine (10mg/ml); and H_2O. For induction purpose, a final concentration of filter sterilized 1 mM IPTG or arabinose was added 1 or 1.5 hours post inoculation for mRNA expression studies in *E. coli* to induce the lac or P_R' promoter, respectively. To maintain plasmids in LB or M9 minimal media, appropriate antibiotics were added (kanamycin (50µg/ml) or ampicillin (100µg/ml)). *E. coli* cultures were maintained at 37ºC.

DNA and RNA preparation

The heterologous DNA cluster *cps3* for qRT-PCR studies was amplified from the genomic DNA of *Lactobacillus plantarum*. *L. plantarum* was grown for 24 hours in *Lactobacillus* MRS broth and the genome was extracted using the Quick gDNA miniprep kit (Zymo research) following manufacturer protocol.

All the plasmids were purified from strains cultured in LB broth incubated at 37^{o}C or 30^{o}C for temperature sensitive plasmid overnight for ~16 hours. Next day, the extraction was performed as per the instructions in the Zyppy Plasmid Miniprep kit (Zymo Research).

For RNA preparation, the respective strains were streaked from frozen stocks on LB plates supplemented with ampicillin and incubated at 37^{o}C. Next day, a culture (3 biological replicates) was started by picking an isolated colony and transferring to M9 minimal media with ampicillin. The culture was incubated in a shaking incubator at 37^{o}C at a speed of 250 r.p.m. for ~16-18 hours. The density of the culture was measured using Genesys 10S UV-Vis spectrophotometer (Thermo scientific) at 600nm. A fresh culture was started by normalizing to an O.D.$_{600nm}$ of 0.1 in a total of 3 ml broth. The cultures were grown at 37^{o}C at a speed of 250 r.p.m. The P_{lac}-1mM and PR+Q samples were induced with 1mM IPTG or arabinose after 1 and 1.5 hours (O.D.$_{600nm}$ ~0.2) respectively. The cells were harvested at mid-exponential phase. The RNA from each sample was extracted by following the instructions in the Quick-RNA miniprep kit (Zymo research).

Construction of plasmids

The integration of λQ encoding gene on the *E. coli* genome was performed using pLA2 vector following the CRIM method developed by Haldimann and Wanner [126]. Classical cloning techniques were used to construct the pLA2+Q vector. Phusion high-fidelity DNA polymerase (ThermoFisher Scientific) was used for amplification purposes. Refer Table 9 (Appendix A) for all primer sequences mentioned in this section. In brief, primers P1 and P2 were used to amplify Q gene using the synthesized oligo from Integrated DNA technologies (Table 8, Appendix A) as template. Double digestion of pLA2 and the amplicon was performed using restriction enzymes NdeI and KpnI. T4 DNA ligase was used to ligate the digested vector and amplicon and the mixture was transformed in chemical competent *E. coli* BW25141 strains (Competent cells were prepared using Mix and Go *E. coli* transformation kit from Zymo research). There were two types of strains constructed to test the effect of λQ - a) PR-Q: strain transformed with pLA2 vector that does not carry the gene encoding λQ; b) PR+Q: Strain transformed with pLA2+Q vector carrying the gene encoding λQ on the genome. pBAD-RFP(EC) was used to introduce the heterologous DNA (*cps3* cluster) amplified from the genomic DNA (gDNA) of *Lactobacillus plantarum*. Refer Table 9 (Appendix A) for all primer sequences mentioned in this section. All the amplifications were performed using the long amp *taq* 2X master mix (New England Biolabs). The pBAD-RFP(EC) vector was first modified by removing *araC, araBAD* and GFP genes and replacing with two restriction enzyme sites XhoI and XbaI using overlap PCR with primers P3 and P4. Next, the *cps3ABglf2* segment of the cps3 cluster was amplified

using primers P5 and P6 from gDNA of *L. plantarum* and inserted in the newly constructed vector which was digested with XhoI and XbaI along with the amplicon. The newly constructed vector also carries the P_R' promoter which was incorporated on the primer P5. The pBAD-*cps3ABglf2* vector was modified further such that two new restriction enzyme sites, NheI and NcoI, were introduced between the terminators *rrnBT1* and *rrnBT2*. This was done with the help of overlap PCR using primers P7 and P8. Lastly, the *cps3FHI* segment of the cps3 cluster was amplified using primers P9 and P10. The amplicon and the vector were digested with NheI and NcoI and the fragment was introduced using ligation. The final vector contains the P_R' promoter and the *cps3ABglf2* segment followed by a terminator *rrnBT1* which carries the *cps3FHI* fragments downstream of it (Figure 6). The final construct was used as a template and using overlap PCR, primers P11 and P12 were used to introduce the P_{lac} promoter in place of P_R' promoter (pJKK2). Finally, two vectors were created; one vector carries the P_{lac} promoter upstream of the modified cps3 (pJKK2) cluster and the second vector carries the P_R' promoter upstream of the modified cps3 cluster (pJKK3). The vector with P_R' promoter upstream of cps3 cluster was introduced in *E. coli* BW25113 [127] strain which was either integrated with gene encoding for λQ on the genome (P_R'+Q) or not (P_R'-Q). The expression of λQ was regulated by using an arabinose promoter and 1mM arabinose was used to induce production of λQ in these studies.

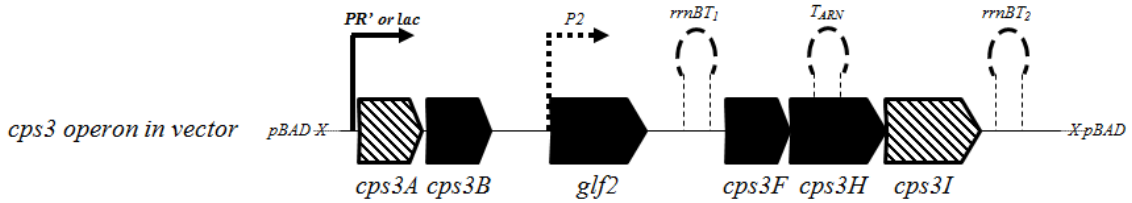


Figure 6 Synthetic cps3 gene cluster.
The pBAD vector carries either a P_R' or lac promoter upstream of the cps3 cluster. In addition, the *cps3ABglf2* segment and *cps3FHI* segment were constructed between the terminator *rrnBT1*. P2 and T_{ARN} are the predicted native promoter and terminator respectively.

qRT-PCR analysis

The qRT-PCR analysis was done in the Quant studio 6 flex system from Applied Biosystems by following a one-step qPCR protocol. Primers P13 to P20 were used for detecting expression of genes. The qscript One-step SYBR Green qRT-PCR, low ROX kit (Quantabio) was used for this purpose. In brief, the final concentration of each components were assembled (10µl in total) as follows: 5µl of master mix, 250nM of forward and reverse primers (0.5 µl of 5µM stock), 1.8 µl of nuclease-free water, 40ng of RNA template (2ul of 20ng/µl stock) and 0.2 µl of reverse transcriptase. The reaction protocol steps were performed as follows: cDNA synthesis at 50C for 10 mins; *taq* activation at 95^{o}C for 5 mins; PCR cycle (45 cycles) at 95^{o}C for 10 sec, 60^{o}C for 20 sec and 72^{o}C for 60 sec; and lastly a melt curve at 65^{o}C for 60 sec and 97^{o}C for 1 sec. Fluorescence was monitored during each extension step of PCR cycle, and a melting curve analysis was performed after each run to confirm the amplification of specific transcript.

qRT-PCR calculation and statistical analysis

The qRT-PCR data was analyzed using the comparative Cq method ($2^{-\Delta\Delta Cq}$ method) [128]. The raw data and calculations are in the supplementary section (Appendix A). The results reflect the values of target gene (*cps3A* and *cps3I)* of induced strains (P_{lac}-1mM or P_R'+Q) relative to un-induced strains (P_{lac}-0mM or P_R'-Q) (Figure 4 and 5). The analysis was done by normalizing the Cq values of *cps3A* and *cps3I* with the housekeeping gene *hcaT*, resulting in ΔCq. The average of the ΔCq for all replicates was log-transformed to yield the ΔCq expression. ΔΔCq values were calculated by comparing the ΔCq expression values of the uninduced samples (P_{lac}-0mM or P_R'-Q) with the induced samples (P_{lac}-1mM or P_R'+Q) for each target gene. The error bars represent the standard error taking into account propagation of error.

CHAPTER III

EXPANDING THE USE OF LAMBDA PROTEIN Q IN *PSEUDOMONAS PUTIDA* KT2440

Introduction

Escherichia coli is commonly used as a host for many functional metagenomic studies due to its simple structure and wide array of available genetic tools[21]. However, *E. coli,* like other standard laboratory microbes, does not contain all regulatory elements and features to successfully express genes from same or different phylum of microbes [82,91]. One approach to overcome this issue is to expand the inventory of alternative hosts to be used for metagenomics studies. Alternate hosts can be used for functional metagenomic studies when a strain lacks certain traits or pathways that is crucial to identify a phenotype of interest. For example, if the metagenomic library is to be screened for clusters that enable utilization of aromatic compounds (such as ferulic acid). An ideal choice for a host would be a strain that naturally degrades the aromatic compound under study. Soil bacterium are known to encode necessary genes to utilize aromatic compounds to intermediates that can be harnessed for energy. By deleting its native genes, the host could be used to express the metagenomic library and select for the desired phenotype. This type of functional screening has previously been done to identify transporters and other novel pathways [21,129,130]. Regardless of benefits which an alternate host can offer, it could still face similar limitations of pre-mature termination due to intrinsic terminators between the promoters and transcription start site since

intrinsic terminators are widely distributed across bacteria [91,94]. Thus, successfully testing functionality of anti-terminator λQ in a host other than *E. coli* could prove beneficial for functional metagenomics studies. In this work, we chose a strain that has characteristics as simple as *E. coli*, that is routinely used due to ease of culturing in laboratory and that has tools developed to perform genetic modifications. We expand the use of the anti-terminator λQ from bacteriophage lambda in a soil bacterium *Pseudomonas putida*.

Pseudomonas putida (*P. putida*) is a gram-negative obligate aerobe which belongs to the bacteria phylum of ɣ-Proteobacteria. *P. putida* has potentially been used to clean-up oil spills [131,132]. *P. putida* has been successfully engineered to express heterologous DNA as well. Nielsen et al. used the butanol biosynthesis pathway from *Clostridium acetobutylicum* to successfully produce butanol in *P. putida* [133]. *P. putida* is suitable for heterologous expression of genes from GC-rich bacterial clades, like Actinobacteria, that are especially rich in secondary metabolite biosynthesis gene clusters [134]. In addition, the bacterium is known to degrade diverse molecules such as naphthalene [135], toluene [136], plastic [137] and aromatic compounds such as ferulate and vanillin [138]. Commonly used *E. coli* promoters, such as the lac family of inducible promoters, have been previously tested in *P. putida* KT2440 with variable success [133,139-141]. Thus, promoters such as P_R', which are readily recognized by *E. coli* RNAP can potentially work in *P. putida* and enable the formation of *P. putida* RNAP and λQ complex. The laboratory growth requirements for *P. putida* are very similar to *E. coli*, and many genetic tools are developed for modification and expression of desired genes

[142]. All these factors potentially make *P. putida* an alternative as a host for metagenomic screening.

In this study, we show the use of an anti-terminator λQ from the bacteriophage lambda, to overcome premature transcriptional termination due to intrinsic terminators in a heterologous DNA in two different host backgrounds, *E. coli* and *P. putida* KT2440. The expression levels of mRNA, using qRT-PCR studies in *E. coli* clearly indicate higher expression of the gene downstream of the intrinsic terminators in presence of λQ. In addition, fluorescent assay performed in *P. putida* KT2440, clearly indicate ~5 fold increase in mKate2 gene in strain expressing λQ when compared to a widely used lac promoter (P_{lac}).

Results

Expression of protein Q in Pseudomonas putida *KT2440 did not result in observable growth deficit*

To expand the use of bacteriophage lambda protein Q to express heterologous DNA containing transcriptional terminators beyond *E. coli*, we introduced the gene encoding λQ on the genome of *Pseudomonas putida* KT2440. A heterologous DNA cluster (~12.5 kbps), from halophiles acquired from the environment [143], was integrated on the genome using serine phage integrase. In addition, a reporter mKate2 gene lacking its own promoter, was integrated downstream of the heterologous DNA cluster (Figure 11). The expression of heterologous DNA cluster was driven by either lac or P_R' promoter in presence or absence of λQ. The resulting strains were first tested for any growth deficiency due to the integration and expression of heterologous DNA in the constructed *P. putida* KT2240 strains (Figure 7). Results showed no significant growth burden due to the integration of heterologous DNA in all the strains with or without expression of λQ.

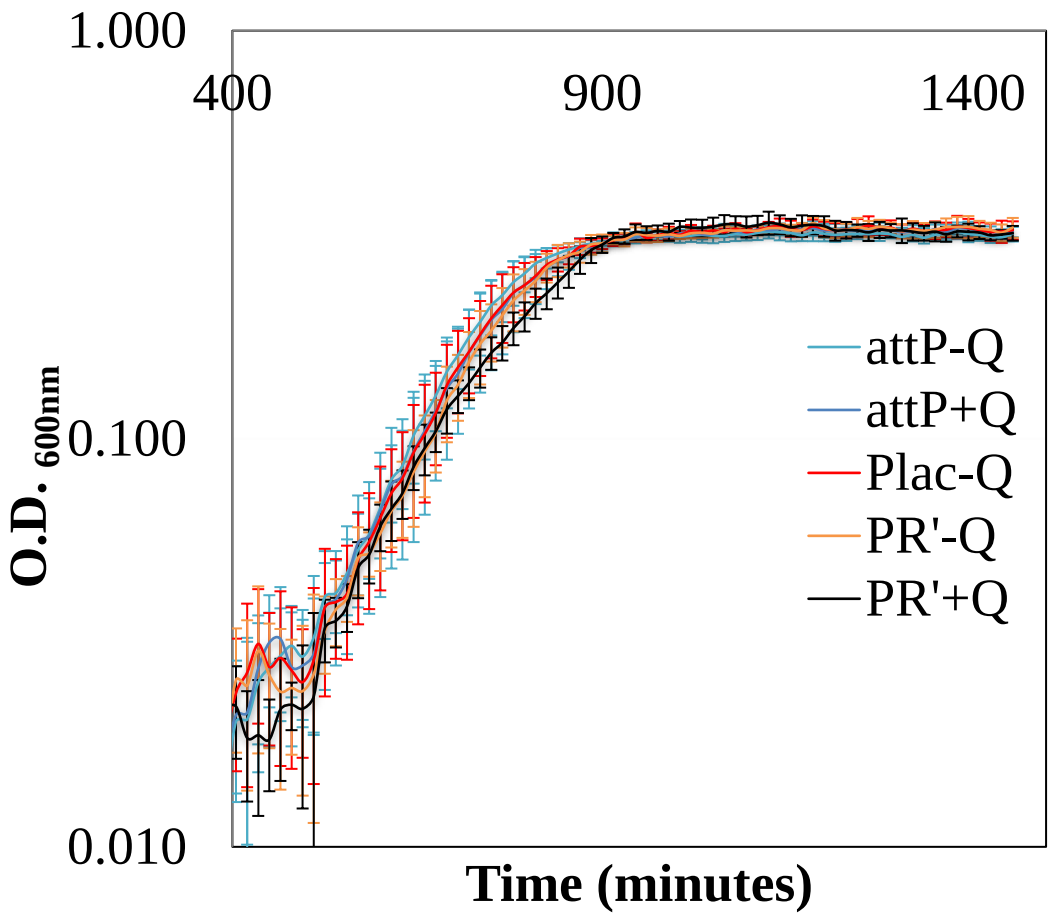


Figure 7 No significant growth difference was observed between strains under study. The strains under study were cultured in M9 minimal media + Kanamycin for ~25 hours in TECAN plate reader. Every 15 minutes, absorbance was measured at 600nm. All strains show similar growth. The growth curve represents an average of triplicates of 3 biological replicates. The error bars represent standard deviation.

The P_R' promoter and anti-terminator protein Q are functional in P. putida *KT2440*

To test the functionality of λQ to overcome internal intrinsic terminators in *P. putida* KT2440, we measured the fluorescence levels of reporter mKate2 in strains carrying the heterologous DNA cluster under the P_R' promoter in presence or absence of λQ. The fluorescence from mKate2 protein was analyzed using 588nm/633nm excitation/emission filters. Interestingly, we observed a significant increase in fluorescence in P_R' + Q strains when compared to P_R' – Q (Figure 8 and 9). No fluorescence was observed in the P_R' - Q strain suggesting that the internal intrinsic terminators in the heterologous DNA interrupts transcription. This obstruction of RNAP complex could be a failure of the RNAP to reach the mKate2 gene resulting in no fluorescence (Figure 9). These results are encouraging, since they firstly display recognition of P_R' promoter by the RNAP in *P. putida* KT2440, a host other than *E. coli*. Secondly, the results indicate that the transcription machinery can overcome intrinsic terminators in heterologous DNA, in presence of anti-terminator λQ, which is indicated by high levels of mKate2 fluorescence only when λQ is expressed in the cell.

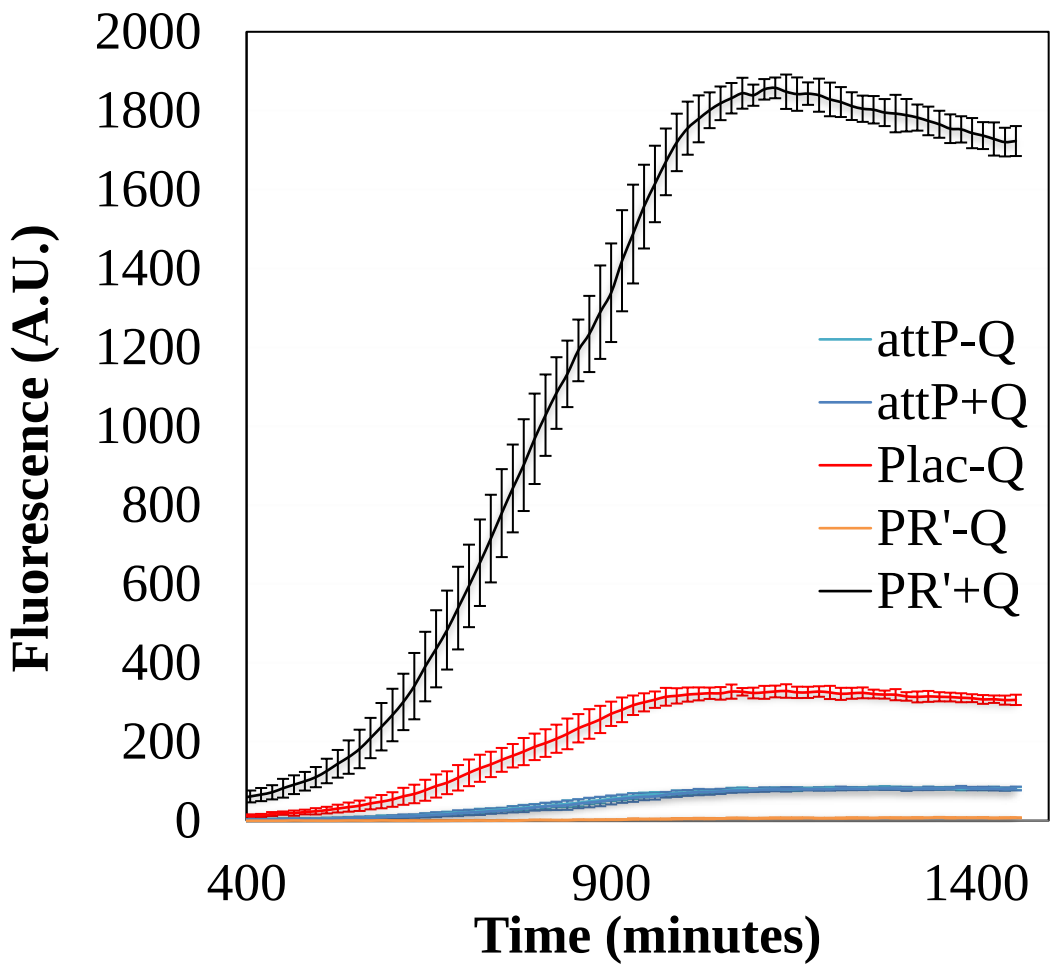


Figure 8 Fluorescence of mKate2 is higher in presence of Q protein.
The strains under study were cultured in M9 minimal media + Kanamycin for ~25 hours in TECAN plate reader. Every 15 minutes, fluorescence was measured using 588nm/633nm excitation/emission filters. The negative control (attP-Q and attP+Q) show minimal fluorescence. The strains with P_R' upstream of heterologous DNA in absence of Q (green) show no fluorescence. The strains with P_R' upstream of heterologous DNA in presence of Q (black) show higher fluorescence compared to DNA carrying P_{lac} (red). The fluorescence curve represents an average of triplicates of 3 biological replicates. The error bars represent standard deviation.

Expression of mKate2 is significantly higher when the expression of heterologous DNA is driven by P_R' promoter in presence of the protein Q compared to lac promoter

The *E. coli* lac promoter is frequently used to express genes in *P. putida* KT2440. We wanted to promote use of P_R' promoter studies which could be limited by the presence of intrinsic terminators in the DNA. Thus, we tested the fluorescence levels of mKate2 when integrated downstream of the heterologous DNA cluster whose expression is driven by either a lac or P_R' promoter. There was a significantly higher fluorescence of mKate2 observed when the expression of heterologous DNA cluster was driven by P_R' promoter in presence of λQ (black bars) compared to P_{lac} (red bars). The observation suggests that in the P_R' + Q strain, the λQ helps to stabilize the RNAP-DNA complex to overcome the intrinsic terminators which are predicted to be on the heterologous DNA cluster. This stability of RNAP-DNA complex might be lacking in the P_{lac} – Q strain translating into lower levels of fluorescence of mKate2.

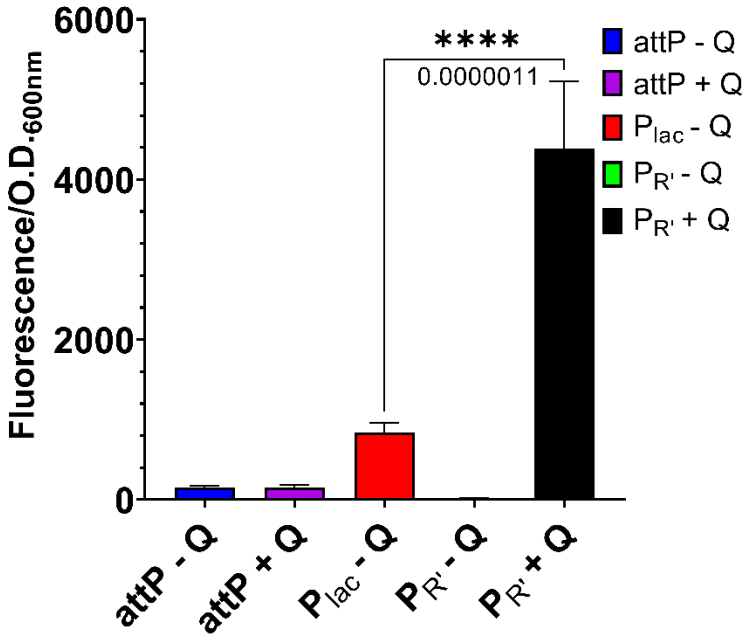


Figure 9 Fluorescence of mKate2 protein is exponentially increased when the cluster carries P_R' promoter in presence of Q protein.
Dark blue: negative control strains that carry attP fragment alone without Q. Light blue: negative control strains that carry attP fragment with Q. Red: P_{lac} strain. Green: P_R' in absence of Q protein. Black: P_R' in the presence of Q protein. The bars indicate average of fluorescence normalized to $O.D._{600nm}$ at mid-log phase (t=600 mins) from triplicates of three biological replicates for each strains under study. The data plotted represents mean ± standard deviation (SD). The p-values were calculated using a two-tailed student's t-test where the asterisk represent: ****P<0.0001.

Discussion

The role of anti-terminator λQ has widely been studied in *E. coli*. Due to these studies, various key sequences and mechanism of λQ acting as an anti-terminator have been identified. However, its use has not been explored beyond *E. coli*. In this study, we test the use of P_R' promoter and λQ in a gram-negative ɣ-Proteobacterium *Pseudomonas putida* KT2440.

In this work, a cluster from environmental isolate was amplified and used to test the use of λQ. The cluster mimicked a heterologous metagenomic library with intrinsic terminators that could potentially cause issues and terminate transcription at various locations. Based on the growth curve (Figure 7), it is apparent that expression of λQ in *P. putida* KT2440 is not toxic, since growth is not significantly different between strains that express λQ compared to the strains that do not. The results of this work clearly indicate, that the mKate2 reporter protein is highly expressed when the heterologous DNA carries P_R' promoter in a strain that expresses λQ. This result is significant for two reasons. Firstly, there is no expression of mKate2 protein observed in strains lacking λQ (Figure 8 and 9). This suggests that the intrinsic terminators occurring in heterologous DNA are not overcome unless λQ is docked on the RNAP despite the recognition of P_R' promoter. Secondly, the expression levels of mKate 2 when the heterologous DNA cluster carries the P_R' promoter are ~5 fold higher compared to levels when the cluster carries P_{lac}. This suggests that when RNAP initiates transcription by recognizing a widely used P_{lac} promoter, it does not efficiently overcome the terminators. Overall, the results clearly indicate that the P_R' promoter and λQ can be used for their anti-terminator

role in *P. putida* KT2440 and prove more useful than using a P_{lac} promoter for identification of genes/pathways in functional metagenomics studies.

It was important to show the use of anti-terminator λQ in a different host other than *E. coli* to efficiently transcribe heterologous DNA. This helps in promoting the use of λQ in studies which require transcription of long clusters of DNA, such as in the field of metagenomics.

Methods

Strains and growth media

Refer to the Table 6 (Appendix A) for genotypes of the strains mentioned in this section. All plasmids are maintained in *E. coli* DH5alpha.

Luria-Bertani (LB) broth/agar (10% tryptone; 5% yeast extract; 8% NaCl; 1.2% agar for plates) was used for culturing all the *E. coli* and *Pseudomonas putida* KT2440 strains during the cloning procedure. M9 minimal media was used for the growth and fluorescence assay studies. The recipe used for 50 mL of M9 minimal media is as follows: 10 mL of 5X M9 salts (1 L of 5X M9 salts: 64 gms of $Na_2HPO_4.H_2O$, 15 gms of KH_2PO_4, 2.5 gms of NaCl, 2.5 gms of NH_4Cl); 5µl of 1 M $CaCl_2$; 100 µl of 1 M $MgSO_4$; 50 µl of 1000X metal ion mix; 400 µl of 40% glucose; 5 µl of thiamine (10mg/ml); and H_2O. To maintain plasmids in LB or M9 minimal media, appropriate antibiotics were added (kanamycin (50µg/ml) or ampicillin (100µg/ml)). *E. coli* and *P. putida* KT2440 cultures were maintained at 37ºC and 30ºC, respectively.

For extraction of genomic DNA from environmental isolates, strains were grown in TGP (composition per liter (pH-7): 17 gm of tryptone; 3 gm of Soy peptone; 5 gm of NaCl; 2.5 gm of K_2HPO_4; post autoclaving, added the following after filter sterilizing: 4 mL of glycerol; 4 gm of sodium pyruvate) broth in a shaking water bath at 55^{o}C for 48 hours at 70 r.p.m.

DNA and RNA preparation

The heterologous DNA cluster for studies in *P. putida* KT2440 was amplified from genomic DNA extracted from environmental isolate (G23C002). For this purpose, G23C002 strain was grown for 48 hours in TGP broth and the genome was extracted as per the protocols using the Quick gDNA miniprep kit (Zymo research).

All the plasmids were purified from strains cultured in LB broth incubated at 37^{o}C overnight for ~16 hours. Next day, the extraction was performed as per the instructions in the Zyppy Plasmid Miniprep kit (Zymo Research).

Construction of plasmids

Integration of λQ encoding gene in *P. putida* KT2440 was performed using suicide vector pK18mobsacB. The strain *P. putida* KT2440 carried the Bxb1-attB attachment sequence in the PP_RS24740 locus. In this study, the strain was modified by integrating the λQ encoding gene (Table 8, Appendix A) by replacing it with non-essential gene cluster from PP_RS17495 to PP_RS17485. The replacement was carried out using suicide vector pK18mobsacB (ATCC 87097) [144]. The final pJKK4 vector was

constructed using a classical cloning technique in the following sequence. The λQ encoding gene (pJAK4) (Table 8, Appendix A) and a total of ~1100 bps of homologous sequence comprising portions of PP_RS17485 and PP_RS17495 (pJAK2) were synthesized and ordered in vector form from Twist Biosciences. First, using restriction enzymes EcoRI and HindIII, the pJAK2 and pK18mobsacB vectors were digested. Using T4 DNA ligase, the *P. putida* homologous region acquired from pJAK2 fragment was inserted in pK18mobsacB resulting in pJAK7. Next, using inverse PCR with the help of primers P21 and P22, restriction sites BamHI and XhoI was introduced in pJAK7 resulting in pJAK8. Using P23 and P24, Q encoding gene was amplified from pJAK4. The amplicon and pJAK8 were digested with BamHI and XhoI and the fragment was inserted using T4 DNA ligase resulting in pJAK9. The final vector was used to integrate λQ encoding gene on the genome using the strategy laid out by Johnson and Beckham (Figure 10) [145]. This resulted in two strains, one where the λQ encoding gene is present (denoted with a suffix +Q in Figures) and absent (denoted with a suffix -Q in Figures) on the genome.

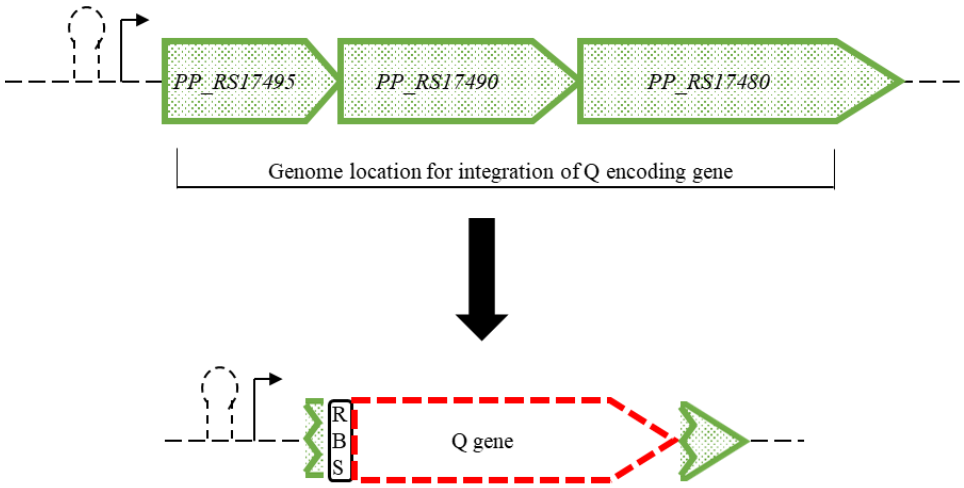


Figure 10 Illustration showing insertion of Q protein encoding gene on *P. putida* KT2440 genome.
The genes PP_RS17495-PP_RS17490-PP_RS17480 on P. putida KT2440 genome are replaced by a protein Q encoding gene with an RBS upstream of it using suicide vector pK18mobsacB.

Integration of heterologous DNA cluster in *P. putida* KT2440 genome was performed using phage integrase BxB1. Serine phage integrases have widely been used as a genetic tool to unidirectionally integrate genes on *P. putida* genome between unique attachment sites. The integration requires a pair of specific sequences, about 40-70 nucleotides in length, namely attB (attachment pads on bacteria) and attP (attachment pads on phage or DNA to be integrated). These sites are recognized by a unique serine phage integrase, which binds to these sequences as a tetramer at the site of exchange and conducts the cross over between the sites [146-148]. Phage integrase systems are very robust and time efficient when compared to other modes of introducing genes on the genome, including widely used homologous recombination techniques and strategies using suicide vectors [141,149]. Thus, due to the ease and sophistication of the phage integrase system, they were used as a technique in this study to introduce DNA on the genome.

Mycobacteriophage BxB1 large serine integrase [150] was used to integrate heterologous DNA in *P. putida* KT2440. BxB1 integrase has been used previously to rapidly introduce DNA in *P. putida* KT2440 [141].

For the purpose of testing transcription of heterologous DNA through a series of intrinsic terminators, DNA was amplified from environmental isolate (G23C002) belonging to genus *Aeribacillus* [151]. The hypothetical cluster is ~12.5 kbps in length and encodes atleast three intrinsic terminators predicted by the ARNold finding terminator tool [96]. Also, a fluorescent reporter protein was used as a proxy to report changes in expression of the cluster due to different promoters and presence or absence of λQ. The mKate fluorescent proteins [152] have been successfully used previously in *P. putida* KT2440 [153]. In this work a brighter variant of mKate called mKate2 was used [154].

The following protocol has been developed by Jay Huenemann and Dr. George Peabody V in Dr. Adam Guss's lab at Oak Ridge National Lab (ORNL) [*Manuscript in progress*]. In order to integrate the heterologous DNA on the genome, the cluster was ligated on the 3'end to a fragment (attP fragment) of genes encoding for mKate2 [154] fluorescent protein (without a promoter) and kanamycin resistance gene (Kan^r) along with the BxB1-attP attachment sequence which is essential for integration between the attachment sites on the genome and the introduced DNA fragment.

The attP fragment was ordered on a vector (pJAK3) from Twist biosciences. The fragment was amplified using unphosphorylated primer P25 and P26. The heterologous DNA cluster was amplified using genomic DNA template extracted from environmental isolate G23C002 [151] using either P27 and P28 carrying P_{lac} at the 5'end or P29 and P28

carrying P_R' at the 5'end. The heterologous DNA amplicons (~12.5 kbps) were treated with fast DNA end repair kit (Thermo fisher scientific) to phosphorylate the 5' ends. The attP fragment and heterologous DNA were ligated using T4 DNA ligase (New England Biolabs) at equimolar concentration (0.06 pmol) in a total of 20 µl volume. The ligation mixture was incubated at room temperature for 2 hours. Competent *P. putida* KT2440 cells were prepared by modifying the protocol by Johnson and Beckham [145]. In brief, a colony from newly streaked *P. putida* KT2440 plate was transferred to 5 mls of LB broth. The tubes were incubated at 250 r.p.m. at 30^{o}C aerobically for ~18 hours. Next day, the culture was distributed in 1 ml aliquots in 1.5 ml sterile eppi tubes and centrifuged at 4000xg for 5 minutes at RT. The supernatant was discarded, and pellet were resuspended in 1 ml of 10% sterile glycerol. The washed culture was centrifuged as before, and the wash step was repeated twice. The supernatant from final centrifugation was discarded and the pellet was resuspended in ~80 µl of 10% glycerol. These electro-competent cells were added with 5µl of ligation mixtures of attP-heterologous DNA along with ~100 ng of pGW31 vector and transferred to 2mm electroporation cuvettes. The electroporation was conducted using ECM 399 exponential decay wave electroporation system (BTX) at 2100 mV. After electroporation, the mixture was added with 1 ml of LB broth and recovered at 30C for 2 hours. After incubation, the culture was transferred to sterile eppi tubes, centrifuged at 4000xg for 5 minutes and the pellet was resuspended in 100µl of sterile milliQ water and spread on LB + kanamycin plates. The plates were incubated at 30^{o}C overnight and next day observed for transformants and confirmed for integration using PCR. The strains of *P. putida* KT2440 with or

without λQ encoding gene were integrated with the attP fragment alone as control strains, referred as attP+Q and attP-Q respectively. *P. putida* KT2440 strains that were integrated with heterologous DNA cluster, carrying the P_{lac} promoter, along with attP fragment are designated as 'P_{lac}-Q'. *P. putida* strains carrying the λQ gene that were integrated with heterologous DNA cluster, carrying the P_R' promoter, along with attP fragment are designated as 'P_R'+Q' and without the λQ encoding gene were designated as 'P_R'-Q' (Figure 11)

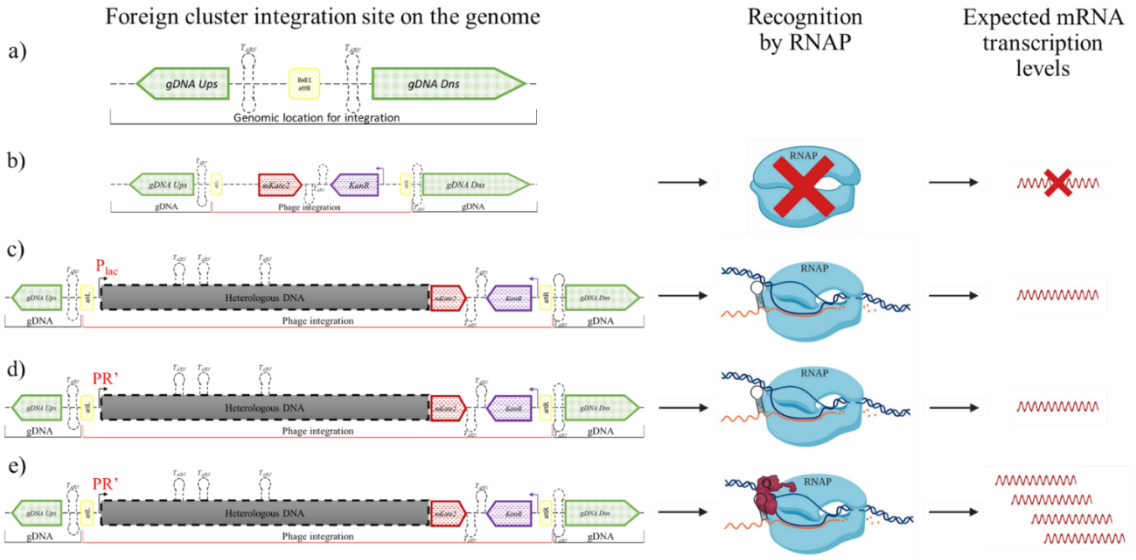


Figure 11 *P. putida* constructs used in the study.
a) Genome location showing presence of BxB1-attB site (yellow); b) Insertion of the mKate2 (red) - KanR (purple) - Bxb1-attP (yellow) fragment (attP fragment) using phage integrase system. mKate2 does not carry a promoter upstream; c) The heterologous DNA (black) with Plac inserted along with the attP fragment on the genome; d) The heterologous DNA (black) with P_R' inserted along with the attP fragment on the genome; e) The heterologous DNA (black) with P_R' inserted along with the attP fragment on the genome.

Fluorescent plate reader assays

All the strains were streaked for isolation on LB + Kan plates. Three biological replicates of each were picked and transferred to M9 + kanamycin broth. The cultures were incubated at 30ºC and 250 rpm for ~18 hours. After incubation, the optical density (O.D.) was measured at 600 nm using Genesys 10S UV-Vis spectrophotometer (Thermo scientific). The O.D. was normalized to ~0.1 in a 1 ml of M9 + kanamycin broth. For the assay, ~20 µl of the normalized culture was added to 180 µl of M9 + kanamycin broth in a 96-well plate (Falcon, flat bottom with low evaporation lid). The 96-well plates were inserted in the TECAN infinite M Nano^{+} plate reader to measure O.D. and fluorescence every 15 minutes (100 cycles) for a total of 25 hours at 30ºC. Each cycle comprised of three orbital shaking steps every 5 minutes for 15 seconds at ~250 r.p.m. After the third shake, O.D. was measured at 600nm and then fluorescence of mKate2 was measured by using 588nm/633nm excitation/emission filters. In Figure 9, the fluorescence values were normalized to the $O.D._{600nm}$ of each culture at 600 minutes (refer to Table 11 and 12 for calculations). The error bars indicted in the graphs represent error of propagation [155].

CHAPTER IV

IDENTIFICATION OF BENEFICIAL CLUSTERS USING BACTERIOPHAGE LAMBDA PROTEIN Q AND FUTURE DIRECTIONS

Background

The highly regulated transcription process is typically terminated by intrinsic or factor dependent terminators [91]. The broad focus of this dissertation is to promote use of anti-terminator λQ to accelerate identification of beneficial phenotype from a metagenomic library. In functional metagenomics, DNA acquired from various sources is usually inserted on a vector and introduced in a host for expression and functional screening based on phenotype of interest [21]. Transcription could fail, if intrinsic terminators are between the promoter and transcription start site of gene prepared from a metagenomic library [80]. Hence, to overcome this limitation, we would like to test the use of bacteriophage lambda protein Q to potentially enhance screening of a metagenomic library for a phenotype.

The Kao lab has access to a few environmental strains (Table 2), primarily halophiles, acquired from the lab of Dr. Heather H. Wilkinson in the Department of Plant pathology and microbiology (Texas A&M University).

The following work was initiated to test the use of λQ on genomic DNA from isolates acquired from various niches in USA.

Table 2 Characteristics of environmental isolates acquired from various locations in USA.

Sr. No.	Strain ID	Soil characteristics			State	Genus
		T (°C)	pH	Na^+ (mg kg^{-1})		
1	**E07C003**	19	6.8	13136	OK	Aeribacillus
2	**E08C020**	19	7.4	15534	OK	Anoxybacillus
3	**G13D008**	11	7.3	18741	NM	Aeribacillus
4	**G19C023**	12	7.4	5946	NM	Aeribacillus
5	**G23C002**	14	7.4	10017	NM	Aeribacillus
6	**G23C019**	14	7.4	10017	NM	Aeribacillus
7	**H20D004**	10	7.1	18262	CA	Geobacillus
8	**J18D015**	30	7.4	221	TX	Aeribacillus
9	**K49C015**	47	ND	ND	UT	Aeribacillus
10	**N09C011**	20	6.88	2476	GA	Geobacillus
11	**F09D026**	Characteristics not identified				
12	**E08M013**					

Data extracted from [151]

Results and discussions

Screening for beneficial phenotype using environmental isolates

Halophiles are the future stars for biobased industries [156]. These strains have been explored for clusters and enzymes that could benefit production of environment friendly biosurfactants, biodegradable plastics and biofuels [156,157]. Lignocellulose is the most abundant source of renewable material that could be further utilized as a substrate for manufacturing of such products. Lignin and cellulose are a heterocomplex present in the cell walls of woody plants. Halophiles have widely been found to carry enzymes like laccase, cellulase that can degrade lignin and cellulose respectively [158]. Various halophiles have been tested to breakdown lignin into more simple intermediates to manufacture products such as biofuels [158]. Ferulic acid, another by-product of lignocellulose biomass pretreatment process, helps in cross-linking of lignin with hemi-cellulose [159,160]. Halophiles are known to degrade ferulic acid to much simpler forms of metabolic intermediates [156,158]. *E. coli*, for instance, cannot naturally breakdown ferulic acid due to lack of enzymes (Figure 12) [1]. However, *Pseudomonas* species are well known to breakdown various aromatic compounds including coumaric acid and ferulic acid [12,161-163]. Thus, a metagenomic library could be expressed in an *E. coli* host or a *Pseudomonas* mutant strain lacking its native pathway to breakdown ferulic acid as a carbon source (Figure 12). Breakdown of ferulic acid involves multiple enzymes in *Pseudomonas sps*. Ferulic acid is converted to protocatechuate acid (PCA) by a series of four enzymes (Figure 14) [1,162]. The PCA, then enters β-ketoadipate pathway and with the help of enzymes encoded in different clusters, breaks down PCA into TCA cycle

intermediates (Figure 15) [164]. Thus, to completely utilize ferulic acid by an organism, all the clusters encoding enzymes required in the pathway must be expressed. Some bacteria contain these clusters in an operon and other contain clusters of these genes far apart from each other [165,166]. For example, *Pseudomonas aeruginosa* contains all the genes, required to convert PCA to TCA intermediates, on several clusters which are ~85kb apart (based on the genome analysis performed using PseudoCyc [167]). Terminators between these clusters will potentially limit identification of these various genes involved in the complete pathway. Even if a long cluster of DNA from such strains is inserted on a vector, the expectation would be that either the foreign promoters in the clusters are recognized by the host RNAP, or the terminators are overcome in order to drive expression using the vector-based promoters. To test the later, the anti-terminator λQ could be used to screen such phenotypes, which will promote expression of multiple clusters by de-sensitizing presence of terminators at the end of operons to successfully transcribe the required genes when they are far apart.

The strains from Dr. Wilkinson's lab were initially screened for growth phenotypes of interest. They were tested for growth in various broths such as Luria Bertani (LB), M9 minimal media (with glucose or glycerol as carbon source) (Table 3) and TGP [168] (Table 4). In addition, the environmental strains were screened for tolerance to n-butanol (Table 3) and various antibiotics (Table 5) in the media.

Table 3 Testing tolerance of environmental isolates to n-butanol using BioscreenC.

Sr. No.	**Strain ID**	Media	LB			M9 + glucose			M9 + glycerol		
		Butan-ol (v/v)	0%	1%	2%	0%	1%	2%	0%	1%	2%
1	**E07C003**		++++	++	-	-	-	-	-	-	-
2	**E08C020**		-	-	-	-	-	-	-	-	-
3	**G13D008**		+	-	-	-	-	-	-	-	
4	**G19C023**		+	+	-	-	-	-	-	-	-
5	**G23C002**		-	-	-	-	-	-	-	-	-
6	**G23C019**		++	++	+	++++	+++	-	+	-	-
7	**H20D004**		+++	++	+++	++++	+++	-	+	++	-
8	**J18D015**		+++	+++	-	+++	+++	++	++	-	++
9	**K49C015**		-	-	-	-	-	-	-	-	-
10	**N09C011**		-	-	-	-	-	-	-	-	-
11	**F09D026**		++	-	-	-	-	-	-	-	-
12	**E08M013**		++	++	-	-	-	-	-	-	-

(biological replicate of 1 with 3 technical replicates; every '+' represents $O.D._{600nm} = 0.2$)

Table 4 Measuring growth of environmental isolates in TGP broth using BioscreenC.

Sr. No.	Strain ID	$O.D._{600nm}$ in TGP media
1	**E07C003**	-
2	**E08C020**	++++
3	**G13D008**	++++
4	**G19C023**	+
5	**G23C002**	++
6	**G23C019**	++
7	**H20D004**	++
8	**J18D015**	+
9	**K49C015**	-
10	**N09C011**	-
11	**F09D026**	+
12	**E08M013**	++

(biological replicate of 1 with 3 technical replicates; every '+' represents $O.D._{600nm} = 0.2$)

Table 5 Testing tolerance of environmental isolates to various antibiotics in broth.

Sr. No.	Strain ID	antibiotic resistance (post 48 hours)				
		LB	Chloram-phenicol	Erythro-mycin	Kana-mycin	Tetra-cycline
1	**E07C003**	**+**	-	-	**+**	-
2	**E08C020**	-	-	-	-	-
3	**G13D008**	**+**	-	**+**	-	-
4	**G19C023**	+	-	-	-	-
5	**G23C002**	-	-	-	-	-
6	**G23C019**	**+**	-	**++**	-	-
7	**H20D004**	**++**	-	**++**	-	-
8	**J18D015**	-	-	**++**	-	-
9	**K49C015**	-	-	-	-	-
10	**N09C011**	-	-	-	-	-
11	**F09D026**	-	-	-	-	-
12	**E08M013**	**++**	-	-	-	-

(biological replicate of 1 with 3 technical replicates; every '+' represents $O.D._{600nm} = 0.2$)

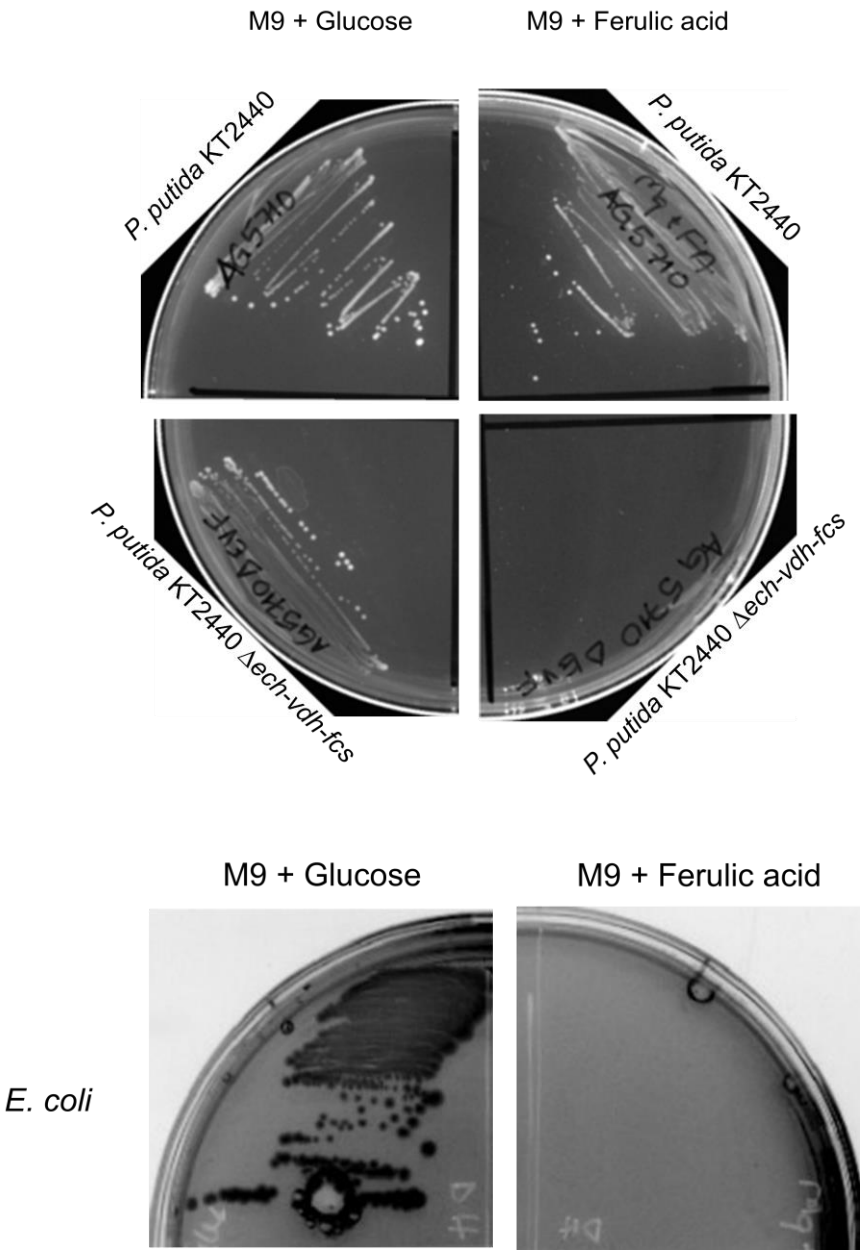


Figure 12 Testing growth of *E. coli* BW25113 and *P. putida* KT2440 on ferulic acid as sole carbon source.

Using plate-based assay, environmental isolates from the collection were streaked on M9 minimal media plates with ferulic acid as carbon source (Figure 13). Based on the analysis, six of the environmental strains (E07C003, F09D026, E08C020, N09C011, G23C002 and G19C023) were found to grow on ferulic acid as carbon source.

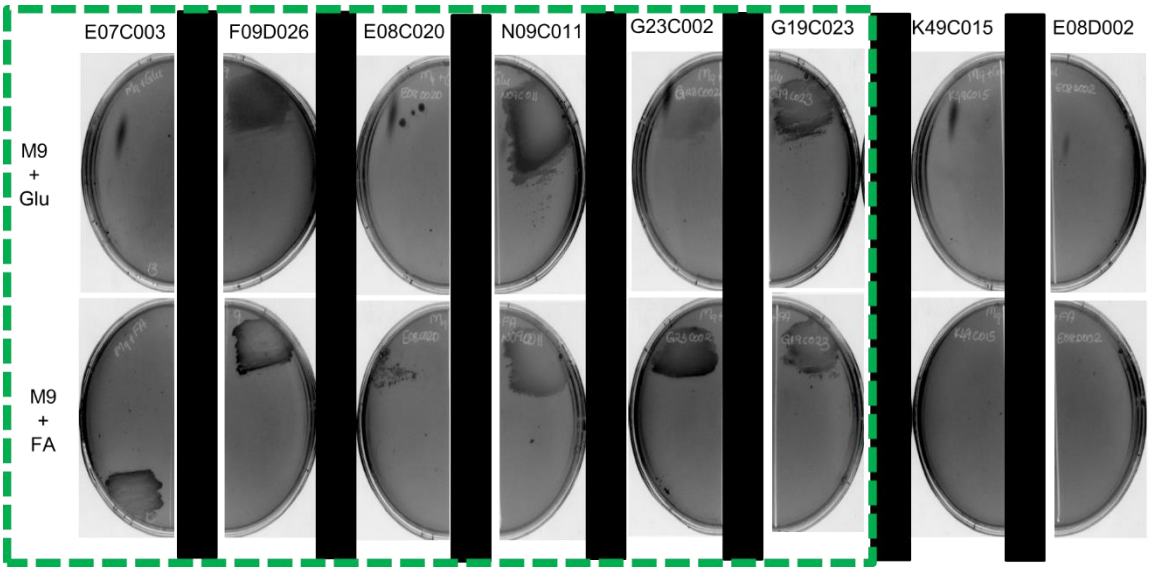


Figure 13 Plate images confirming growth of environmental isolates in M9 media with ferulic acid as carbon source.

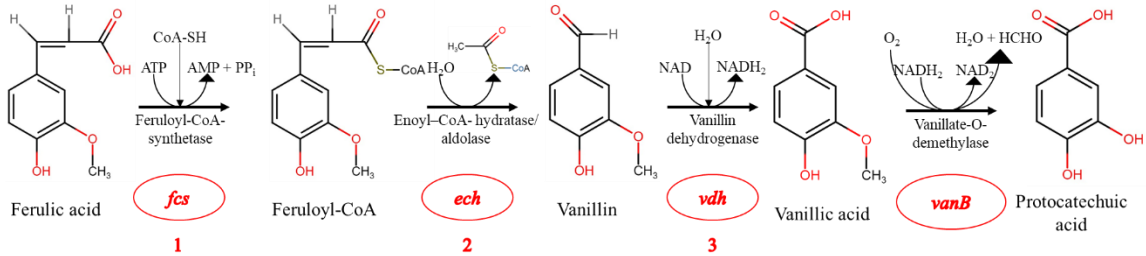


Figure 14 *Pseudomonas fluorescens* is known to break down ferulic acid with the help of various enzymes.
Structures created with the help of https://chem-space.com/search and adapted from [1]

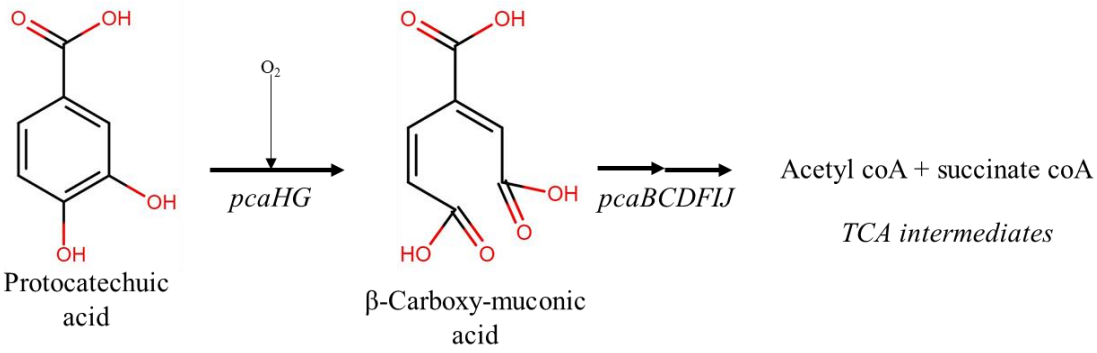


Figure 15 Conversion of protocatechuate to TCA intermediates by *Pseudomonas* sps. Structures created with the help of https://chem-space.com/search and adapted from [2]

Identification of potential ferulic acid breakdown pathway

Ferulic acid breakdown has previously been explored in many bacteria. The pathway to utilize the aromatic compound has been studied in detail in *Pseudomonas* sps. [12,161-163]. Based on the studies, ferulic acid (FA) is broken down into protocatechuate acid (PCA) with the help of four enzymes: feruloyl-coA-synthetase (*fcs*); enoyl-coA-hydratase (*ech*); vanillin dehydrogenase (*vdh*) and vanillate-o-demethylase (*vanB*). Using the amino acid sequence of these enzymes, BLAST search was performed to identify similar homologs in the environmental isolates that were successfully tested to grow on ferulic acid as sole carbon source. Analysis of 3 of these environmental isolates (E07C003, G23C002 and N09C011) are shown in Figure 16.

Based on the analysis, some of the environmental strains encode homologs of enzymes predicted to be involved in the ferulic acid breakdown pathway. Some environmental strains have more than one homolog of the similar genes placed in different clusters on the genome. For example, isolate G23C002, contains all the predicted homologs of *fcs*, *ech*, *vdh* and *vanB* in a region spanning ~24 kbps. However, these genes are not in an operon and could potentially have various terminators predicted in between them. Presence of terminators could potentially limit identification of such clusters or multiple genes involved in a pathway if the foreign promoter is not recognized in the host. Thus, if the genomic DNA from such environmental isolates is used to screen for growth on ferulic acid, then use of anti-terminator λQ could help in overcoming the terminators barrier.

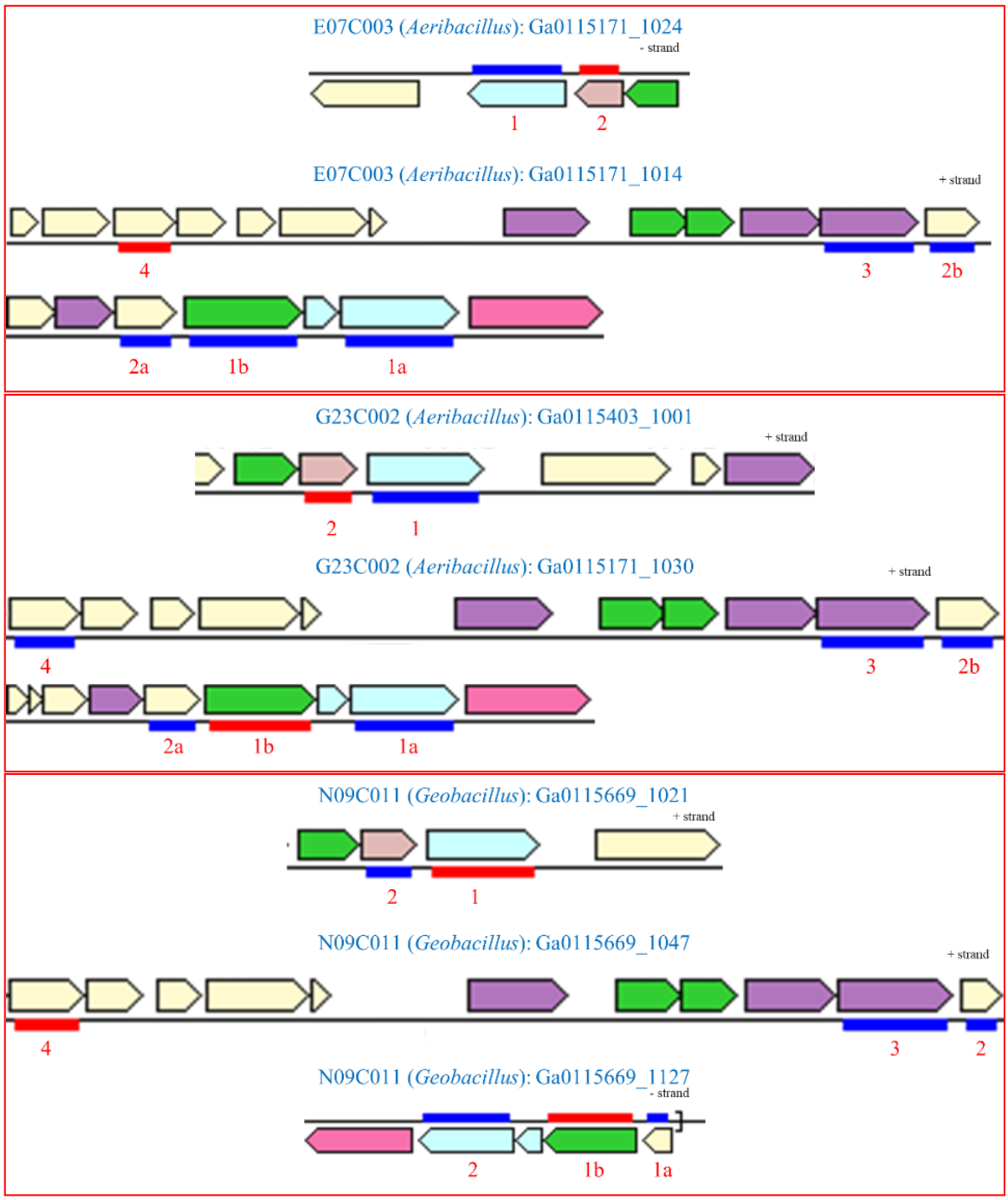


Figure 16 JGI Prediction of various homologs of genes involved in ferulic acid to vanillin breakdown in environmental isolates.
The homolog predictions were based on the amino acid sequence of enzymes involved in ferulic acid breakdown pathway in *Pseudomonas* sps. The numbers represent homologs as follows: *fcs* like – 1(a and b); *ech* like – 2(a and b); *vdh* like – 3; *vanB* like – 4

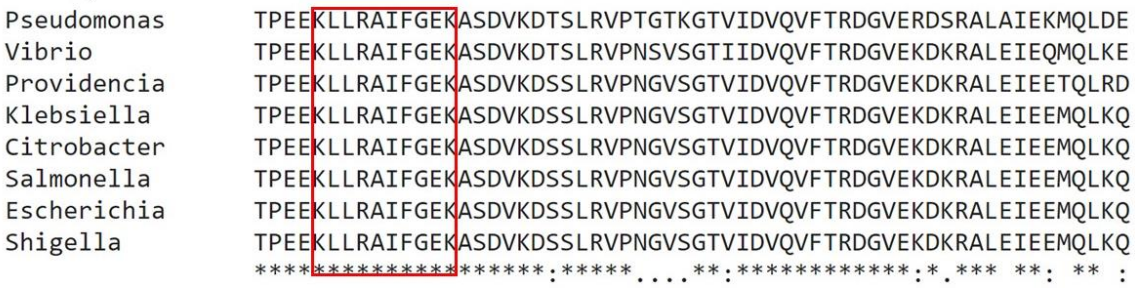
Future directions

A few environmental isolates were successfully screened to grow on media plates with ferulic acid as sole carbon source. Computational analysis predicted presence of homologs of enzymes involved in the breakdown of ferulic acid in *Pseudomonas* sps. This paves way to explore use of bacteriophage lambda protein Q on a genomic library constructed from these environmental isolates to screen for growth on ferulic acid in presence and absence of λQ. The hypothesis is that, if the complete set of genes (*fcs, ech, vdh, vanB*) is required for breakdown of ferulic acid, predicted to be in separate but neighbor clusters on the genome of environmental isolates, then the presence of λQ will assist in expression of the genes eventually helping in identification of the genes required for the pathway. This set of functional screening could be performed in *E. coli* and *P. putida* KT2440. In case of *E. coli*, it natively lacks genes to breakdown protocatechuate acid (PCA) into TCA intermediates. Hence, *E. coli* would have to be genetically engineered to convert PCA to TCA intermediates, as shown by Sonya et al. [169], and then used as a host to screen for ferulic acid breakdown. On contrary, a mutant *P. putida* KT2440 strain (lacking native *fcs, ech, vdh, vanB* genes) could be used as a host to screen for the ferulic acid breakdown phenotype. *P. putida* KT2440 makes a better choice as a host for screening of ferulic acid breakdown pathway since it naturally carries the downstream pathway of converting PCA to TCA intermediates.

In addition, Protein Q has been shown to attach to the β-flap portion of the RNAP by replacing the σ70 region 4 at the flap-tip helix portion (amino acid residues 900-909) of the β-subunit [4,110]. Computational analysis using *E. coli* RpoB (encoding the

β-subunit of RNAP) amino acid sequence suggest that various other species of bacteria carry homologs of the β-subunit. Most of the identified genus, like *E. coli*, belong to the phylum proteobacteria. Upon scanning the amino acid sequences of a subset of these bacteria, it was found that flap-tip helix portions of the β-subunit was highly conserved (see Figure 17). Most of the identified genus, like *E. coli*, belong to the phylum proteobacteria. In addition, various studies have either tested recognition of promoters on genes from Shigella [170] and Citrobacter[171] by *E. coli's* transcription machinery or conservation of promoter sequences between *E. coli* and *Klebsiella* sp.[172]. This suggests that there is a possibility of the $P_{R'}$ promoter to be recognized by the transcription machinery of other such background strains. In this work, we showed the use of anti-terminator protein Q in *Pseudomonas putida* KT2440, a different host other than *E. coli*, to efficiently transcribe heterologous DNA. This will help in expanding the use of protein Q in possibly other strains and facilitate studies which require transcription of long clusters of DNA, such as in the field of metagenomics.

a)



b)

Genus name	NCBI sequence ID	Amino acid similarity (%)
Shigella	WP_052992158.1	99.93
Salmonella	WP_023226537.1	99.03
Citrobacter	WP_024130652.1	98.66
Klebsiella	WP_025713260.1	98.14
Providencia	WP_014656956.1	93.52
Vibrio	WP_038140685.1	85.10
Pseudomonas	WP_122043688.1	72.94

Figure 17 *E. coli rpoB* homologs in other eubacteria.
a) An amino acid sequence alignment of the *rpoB* homologs from various genus highlighting the conserved flap-tip portion (red box) of the β-subunit where protein Q is predicted to bind to the RNAP. b) represent the amino acid similarity between *E. coli rpoB* sequence and the homologs in other genus.

Methods

Strains and growth media

Refer to Table 6 (Appendix A) for environmental strains mentioned in this section.

Growth assays were performed in different media. Luria-Bertani (LB) broth (10% tryptone; 5% yeast extract; 8% NaCl; 1.2% agar for plates) was used for growth assay to test tolerance in presence of butanol (0-2% v/v). M9 minimal media broth was used to test growth in presence of glucose or glycerol as carbon source. The recipe used for 50 mL of M9 minimal media is as follows: 10 mL of 5X M9 salts (1 L of 5X M9 salts: 64 gms of $Na_2HPO_4.H_2O$, 15 gms of KH_2PO_4, 2.5 gms of NaCl, 2.5 gms of NH_4Cl); 5µl of 1 M $CaCl_2$; 100 µl of 1 M $MgSO_4$; 50 µl of 1000X metal ion mix; 400 µl of 40% glucose or glycerol (as carbon source); 5 µl of thiamine (10mg/ml); and H_2O. TGP broth (composition per liter (pH-7): 17 gm of tryptone; 3 gm of Soy peptone; 5 gm of NaCl; 2.5 gm of K_2HPO_4; post autoclaving, added the following after filter sterilizing: 4 mL of glycerol; 4 gm of sodium pyruvate) was used during growth assay and to culture the environmental strains by shaking in a water bath at 55^oC for 48 hours at 70 r.p.m. To test antibiotic susceptibility, antibiotics were added as follows: kanamycin (30µg/ml); chloramphenicol (30µg/ml); erythromycin (10µg/ml) and tetracycline (10µg/ml).

Growth assay

All growth assays were performed in BioscreenC at 50^{o}C for 36 hours. Overnight cultures were diluted in 1:100 ratio and final volume of 300µl was added to honeycomb microplates (BioscreenC MBR). The plates were centrifuged at medium speed once every hour. Absorbance was recorded using wide band filter (450-580nm).

Testing growth on ferulic acid

E. coli and *P. putida* KT2440 cultures were maintained at 37^{o}C and 30^{o}C, respectively. The environmental strains were maintained at 50^{o}C on plates which were vacuum sealed and incubated in a water bath set at 50^{o}C. The M9 minimal media were prepared to test growth on ferulic acid as sole carbon source, where the glucose from the standard composition was replaced by ferulic acid (final conc. 0.1%w/v).

Identification of E. coli rpoB *homologs*

The amino acid sequence (see Supplementary Fig. 3 for information on the sequences) of *E. coli rpoB* (*rpoB$_{EC}$*) was downloaded from NCBI (https://www.ncbi.nlm.nih.gov/). Using the protein BLAST tool, homologs of *rpoB$_{EC}$* were identified restricting the similarity range between 70-100%. The amino acid sequences for the genus were extracted and a multiple sequence alignment was performed using the clustal omega program (https://www.ebi.ac.uk/Tools/msa/clustalo/)[173].

CHAPTER V

CONCLUSIONS

Functional metagenomics is a beneficial tool to identify genes or clusters from microbes that cannot be easily cultured in laboratory conditions. Typically, genomic DNA acquired from environmental sources is cloned into a vector and then transformed into a host, such as *E. coli*, to screen for phenotype of interest. One of the limitations that arises in expression of heterologous DNA is due to non-compatible foreign promoters. Vector-based promoters, which are recognized by the host DNA-dependent RNA polymerase (RNAP), can minimize this issue. However, if the promoters on the heterologous DNA are not recognized by the host RNAP, multiple essential genes for a phenotype not in an operon and separated by terminators, will not be transcribed by the vector-based promoter upstream of the inserted DNA. The work showed in this dissertation focuses on the issue of overcoming terminators by promoting the use of bacteriophage lambda anti-terminator protein Q during functional metagenomic screening.

Bacteriophage lambda anti-terminator protein Q regulates the expression of late genes. The terminator t_R' located upstream of the cluster of late genes, encoding the proteins required for the lytic phase, inhibits the expression of these genes when not required [97]. However, when the phage needs to transition into the lytic phase, λQ is expressed and which in turn helps the host's RNAP to overcome the t_R' terminator and result in expression of late genes. In this study, we tested the use of bacteriophage

lambda protein Q to overcome intrinsic terminators to transcribe heterologous DNA in *E. coli* BW25113 and *P. putida* KT2440.

Previously, bacteriophage lambda protein Q has been shown to overcome native *E. coli* terminator *rrnBT1* [112]. In the first study, a cluster from *L. plantarum* was used as a source of heterologous DNA. The *cps3* cluster was split into two segments and cloned on a replicating vector backbone. The segments were separated by an *E. coli* terminator *rrnBT1*, in addition to a predicted terminator in the first segment. The qRT-PCR assay was used to test the difference in expression of gene *cps3A* (gene in the first segment) and *cps3I* (gene in the second segment downstream of the terminators) in presence and absence of λQ. Based on the qRT-PCR analysis, the expression of the gene *cps3I* (downstream of terminators) was significantly higher in presence of λQ compared to expression levels when the λQ was not present in the cells. The results suggest that the λQ is functioning in *E. coli* as expected and aids in transcription of heterologous DNA with more efficiency.

Functional metagenomic studies can be more successful when alternate hosts could be used based on the phenotype under study. In order to expand the use of λQ for functional metagenomic studies, we broaden its usage in another host besides *E. coli*. For the second study, we tested the efficiency of transcription in *P. putida* KT2440 in presence and absence of λQ. The transcription efficiency was reported by measuring the expression of a fluorescent protein mKate2, which was downstream of a heterologous cluster predicted to carry intrinsic terminators. The cluster was introduced with either a lac promoter (P_{lac}) or P_R' promoter. The expression of mKate2 gene in strains carrying

the heterologous DNA with P_R' promoter was tested in presence or absence of λQ. Based on the fluorescence assay, the mKate2 protein expression was significantly higher when the cluster carried a P_R' promoter and the transcription was assisted by λQ. This suggests two important findings. First, P_R' promoter is successfully recognized by *P. putida* KT2440 RNAP. Second, bacteriophage lambda protein Q helped RNAP to successfully overcome terminators in the heterologous DNA to express the mKate2. The later observation is more prominent when the expression of mKate2 was detected to be not as strong when the cluster carried lac promoter. These findings encourage and promote the use of P_R' promoter along with λQ in *P. putida* KT2440, especially when a metagenomic library is to be screened using functional metagenomics studies.

Intrinsic terminators are widely distributed across bacteria domain [91]. The terminators vary in efficiency and whether they act uni-directionally or bi-directionally to terminate transcription [95]. The first part of the study focused on overcoming an *E. coli* terminator *rrnBT1*. However, in order to encourage the use of λQ over a wide range of functional metagenomic studies, there is a need to test the efficiency of λQ to overcome intrinsic terminators with variable strength. Such study will further strengthen the role of λQ and promote its use when the DNA library is taken from random samples in order to limit the issues arising with the presence of intrinsic terminators.

In this study, we tested the use of native bacteriophage lambda protein Q. However, a variant of λQ (Q***), is shown to have higher expression of *lacZ* overcoming terminator t_R' when compared to the native λQ [4]. In the same study, λQ*** is not affected by absence of the Q binding element, which is essential for λQ to dock to

RNAP. Thus, it will be interesting to test the variant Q*** for expression of heterologous DNA.

Lastly, in this work, we demonstrated the use of bacteriophage lambda protein Q in *P. putida* KT2440. Studies on λQ have primarily been conducted in *E. coli*, the host organism for bacteriophage lambda. The successful expression of mKate2 in presence of λQ in *P. putida* KT2440, encourages widening the use of λQ in other hosts. *E. coli* based promoters are known to be recognized by transcription machinery of other bacteria [83,174,175]. Thus, there is a possibility that P_R' promoter from bacteriophage lambda could be recognized by foreign hosts and assist in recruiting λQ to host's RNAP. In addition, if P_R' promoter recognition by foreign host RNAP is an issue then, as shown by the studies by Deighan et al., λQ*** could be used with a LacUV5 promoter. Lambda Q*** is shown to be able to assist RNAP to overcome terminator t_R' when the transcript carries P_{lacUV5} [4]. This is a promising approach, since P_{lacUV5} is shown to be recognized by RNAP of other microbes as well [176-178]. Finally, λQ brings a lot of promise to be used in progressing the functional metagenomics field.

REFERENCES

1 Overhage, J., Priefert, H. & Steinbuchel, A. Biochemical and genetic analyses of ferulic acid catabolism in Pseudomonas sp. Strain HR199. *Appl Environ Microbiol* **65**, 4837-4847, doi:10.1128/AEM.65.11.4837-4847.1999 (1999).

2 Ribbons, D. W. & Evans, W. C. Oxidative metabolism of protocatechuic acid by certain soil pseudomonads: a new ring-fission mechanism. *Biochem J* **83**, 482-492, doi:10.1042/bj0830482 (1962).

3 Kuwada, N. J., Traxler, B. & Wiggins, P. A. Genome-scale quantitative characterization of bacterial protein localization dynamics throughout the cell cycle. *Mol Microbiol* **95**, 64-79, doi:10.1111/mmi.12841 (2015).

4 Deighan, P., Diez, C. M., Leibman, M., Hochschild, A. & Nickels, B. E. The bacteriophage lambda Q antiterminator protein contacts the beta-flap domain of RNA polymerase. *Proc Natl Acad Sci U S A* **105**, 15305-15310, doi:10.1073/pnas.0805757105 (2008).

5 Winkler, S. L. *et al.* Vehicle criteria pollutant (PM, NOx, CO, HCs) emissions: how low should we go? *npj Climate and Atmospheric Science* **1**, doi:10.1038/s41612-018-0037-5 (2018).

6 de Souza, W. R. in *Sustainable Degradation of Lignocellulosic Biomass - Techniques, Applications and Commercialization* Ch. Chapter 9, (IntechOpen, Croatia, 2013).

7 Wang, M., Han, J., Dunn, J. B., Cai, H. & Elgowainy, A. Well-to-wheels energy use and greenhouse gas emissions of ethanol from corn, sugarcane and cellulosic biomass for US use. *Environmental Research Letters* **7**, doi:10.1088/1748-9326/7/4/045905 (2012).

8 Lewandrowski, J. *et al.* The greenhouse gas benefits of corn ethanol – assessing recent evidence. *Biofuels* **11**, 361-375, doi:10.1080/17597269.2018.1546488 (2019).

9 Hill, J., Nelson, E., Tilman, D., Polasky, S. & Tiffany, D. Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proc Natl Acad Sci U S A* **103**, 11206-11210, doi:10.1073/pnas.0604600103 (2006).

10 Davis, K. *et al.* Recovery and Utilization of Lignin Monomers as Part of the Biorefinery Approach. *Energies* **9**, doi:10.3390/en9100808 (2016).

11 Ragauskas, A. J. *et al.* Lignin valorization: improving lignin processing in the biorefinery. *Science* **344**, 1246843, doi:10.1126/science.1246843 (2014).

12 Xu, Z., Lei, P., Zhai, R., Wen, Z. & Jin, M. Recent advances in lignin valorization with bacterial cultures: microorganisms, metabolic pathways, and bio-products. *Biotechnol Biofuels* **12**, 32, doi:10.1186/s13068-019-1376-0 (2019).

13 Alves, L. d. F., Silva-Rocha, R. & Guazzaroni, M.-E. in *Functional Metagenomics: Tools and Applications* Ch. Chapter 5, 75-94, (Springer International Publishing, Switzerland, 2017).

14 Herrera, S. Industrial biotechnology-a chance at redemption. *Nat Biotechnol* **22**, 671-675, doi:10.1038/nbt0604-671 (2004).

15 de Lorenzo, V. *et al.* The power of synthetic biology for bioproduction, remediation and pollution control: The UN's Sustainable Development Goals will inevitably require the application of molecular biology and biotechnology on a global scale. *EMBO Rep* **19**, doi:10.15252/embr.201745658 (2018).

16 Schmidt, C. W. Synthetic biology: environmental health implications of a new field. *Environ Health Perspect* **118**, A 118-123, doi:10.1289/ehp.118-a118 (2010).

17 Xu, X. *et al.* Petroleum Hydrocarbon-Degrading Bacteria for the Remediation of Oil Pollution Under Aerobic Conditions: A Perspective Analysis. *Front Microbiol* **9**, 2885, doi:10.3389/fmicb.2018.02885 (2018).

18 *Current Uses of Synthetic Biology for Renewable Chemicals, Pharmaceuticals, and Biofuels*, <https://archive.bio.org/articles/current-uses-synthetic-biology> (2013).

19 Moe-Behrens, G. H., Davis, R. & Haynes, K. A. Preparing synthetic biology for the world. *Front Microbiol* **4**, 5, doi:10.3389/fmicb.2013.00005 (2013).

20 Steidler, L. *et al.* Biological containment of genetically modified Lactococcus lactis for intestinal delivery of human interleukin 10. *Nat Biotechnol* **21**, 785-789, doi:10.1038/nbt840 (2003).

21 Guazzaroni, M. E., Silva-Rocha, R. & Ward, R. J. Synthetic biology approaches to improve biocatalyst identification in metagenomic library screening. *Microb Biotechnol* **8**, 52-64, doi:10.1111/1751-7915.12146 (2015).

22 Kellenberger, E. Exploring the unknown. The silent revolution of microbiology. *EMBO Rep* **2**, 5-7, doi:10.1093/embo-reports/kve014 (2001).

23 Brouillette, M. These microbial communities have learned to live at Earth's most extreme reaches. *Nature*, doi:10.1038/d41586-020-00697-y (2020).

24 van der Helm, E., Genee, H. J. & Sommer, M. O. A. The evolving interface between synthetic biology and functional metagenomics. *Nat Chem Biol* **14**, 752-759, doi:10.1038/s41589-018-0100-x (2018).

25 Turnbaugh, P. J. & Gordon, J. I. An invitation to the marriage of metagenomics and metabolomics. *Cell* **134**, 708-713, doi:10.1016/j.cell.2008.08.025 (2008).

26 Schloss, P. D. & Handelsman, J. Toward a census of bacteria in soil. *PLoS Comput Biol* **2**, e92, doi:10.1371/journal.pcbi.0020092 (2006).

27 Nalin, R., Ranjard, L., Nazaret, S. & Simonet, P. W. La biologie moléculaire en écologie microbienne du sol: application à l'analyse de la structure des communautés bacteriennes. *Bull. Soc. Fr. Microbiol.* **13**, 21-26 (1998).

28 Sleator, R. D., Shortall, C. & Hill, C. Metagenomics. *Lett Appl Microbiol* **47**, 361-366, doi:10.1111/j.1472-765X.2008.02444.x (2008).

29 Pedros-Alio, C. & Manrubia, S. The vast unknown microbial biosphere. *Proc Natl Acad Sci U S A* **113**, 6585-6587, doi:10.1073/pnas.1606105113 (2016).

30 Dinsdale, E. A. *et al.* Functional metagenomic profiling of nine biomes. *Nature* **452**, 629-632, doi:10.1038/nature06810 (2008).

31 Stewart, E. J. Growing unculturable bacteria. *J Bacteriol* **194**, 4151-4160, doi:10.1128/JB.00345-12 (2012).

32 Beja, O. *et al.* Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**, 1902-1906, doi:10.1126/science.289.5486.1902 (2000).

33 Tringe, S. G. & Rubin, E. M. Metagenomics: DNA sequencing of environmental samples. *Nat Rev Genet* **6**, 805-814, doi:10.1038/nrg1709 (2005).

34 Majernik, A., Gottschalk, G. & Daniel, R. Screening of environmental DNA libraries for the presence of genes conferring Na(+)(Li(+))/H(+) antiporter activity on Escherichia coli: characterization of the recovered genes and the corresponding gene products. *J Bacteriol* **183**, 6645-6653, doi:10.1128/JB.183.22.6645-6653.2001 (2001).

35 Rondon, M. R. *et al.* Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* **66**, 2541-2547, doi:10.1128/aem.66.6.2541-2547.2000 (2000).

36 Forsberg, K. J. *et al.* Identification of Genes Conferring Tolerance to Lignocellulose-Derived Inhibitors by Functional Selections in Soil Metagenomes. *Appl Environ Microbiol* **82**, 528-537, doi:10.1128/AEM.02838-15 (2016).

37 Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74**, 5088-5090, doi:10.1073/pnas.74.11.5088 (1977).

38 Lane, D. J. *et al.* Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci U S A* **82**, 6955-6959, doi:10.1073/pnas.82.20.6955 (1985).

39 Stahl, D. A., Lane, D. J., Olsen, G. J. & Pace, N. R. Characterization of a Yellowstone hot spring microbial community by 5S rRNA sequences. *Appl Environ Microbiol* **49**, 1379-1384, doi:10.1128/AEM.49.6.1379-1384.1985 (1985).

40 Olsen, G. J. *et al.* Sequence of the 16S rRNA gene from the thermoacidophilic archaebacterium Sulfolobus solfataricus and its evolutionary implications. *J Mol Evol* **22**, 301-307, doi:10.1007/BF02115685 (1985).

41 Pace, N. R., Stahl, D. A., Lane, D. J. & Olsen, G. J. in *Advances in Microbial Ecology Advances in Microbial Ecology* (ed K. C. Marshall) Ch. Chapter 1, 1-55 (Springer US, USA, 1986).

42 Pemberton, J. M., Vincent, K. M. & Penfold, R. J. Cloning and heterologous expression of the violacein biosynthesis gene cluster fromChromobacterium violaceum. *Current Microbiology* **22**, 355-358, doi:10.1007/bf02092154 (1991).

43 Pace, N. R. Opening the door onto the natural microbial world: molecular microbial ecology. *Harvey Lect* **91**, 59-78 (1995).

44 Handelsman, J. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* **68**, 669-685, doi:10.1128/MMBR.68.4.669-685.2004 (2004).

45 Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. & Goodman, R. M. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* **5**, R245-249, doi:10.1016/s1074-5521(98)90108-9 (1998).

46 Lewin, A., Lale, R. & Wentzel, A. in *Functional Metagenomics: Tools and Applications* (eds Trevor C. Charles, Mark R. Liles, & Angela Sessitsch) Ch. Chapter 2, 13-44 (Springer International Publishing, Switzerland, 2017).

47 Henne, A., Schmitz, R. A., Bomeke, M., Gottschalk, G. & Daniel, R. Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on Escherichia coli. *Appl Environ Microbiol* **66**, 3113-3116, doi:10.1128/aem.66.7.3113-3116.2000 (2000).

48 Heath, C., Hu, X. P., Cary, S. C. & Cowan, D. Identification of a novel alkaliphilic esterase active at low temperatures by screening a metagenomic library from antarctic desert soil. *Appl Environ Microbiol* **75**, 4657-4659, doi:10.1128/AEM.02597-08 (2009).

49 Henne, A., Daniel, R., Schmitz, R. A. & Gottschalk, G. Construction of environmental DNA libraries in Escherichia coli and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. *Appl Environ Microbiol* **65**, 3901-3907, doi:10.1128/AEM.65.9.3901-3907.1999 (1999).

50 Simon, C., Herath, J., Rockstroh, S. & Daniel, R. Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. *Appl Environ Microbiol* **75**, 2964-2968, doi:10.1128/AEM.02644-08 (2009).

51 Entcheva, P., Liebl, W., Johann, A., Hartsch, T. & Streit, W. R. Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. *Appl Environ Microbiol* **67**, 89-99, doi:10.1128/AEM.67.1.89-99.2001 (2001).

52 Riesenfeld, C. S., Goodman, R. M. & Handelsman, J. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environ Microbiol* **6**, 981-989, doi:10.1111/j.1462-2920.2004.00664.x (2004).

53 Diaz-Torres, M. L. *et al.* Novel tetracycline resistance determinant from the oral metagenome. *Antimicrob Agents Chemother* **47**, 1430-1432, doi:10.1128/aac.47.4.1430-1432.2003 (2003).

54 Henning, H. *et al.* Identification of novel benzoylformate decarboxylases by growth selection. *Appl Environ Microbiol* **72**, 7510-7517, doi:10.1128/AEM.01541-06 (2006).

55 Sommer, M. O. A., Dantas, G. & Church, G. M. Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science* **325**, 1128-1131, doi:10.1126/science.1176950 (2009).

56 Culligan, E. P., Sleator, R. D., Marchesi, J. R. & Hill, C. Functional environmental screening of a metagenomic library identifies stlA; a unique salt

tolerance locus from the human gut microbiome. *PLoS One* **8**, e82985, doi:10.1371/journal.pone.0082985 (2013).

57 Guazzaroni, M. E., Morgante, V., Mirete, S. & Gonzalez-Pastor, J. E. Novel acid resistance genes from the metagenome of the Tinto River, an extremely acidic environment. *Environ Microbiol* **15**, 1088-1102, doi:10.1111/1462-2920.12021 (2013).

58 Wichmann, F., Udikovic-Kolic, N., Andrew, S. & Handelsman, J. Diverse antibiotic resistance genes in dairy cow manure. *mBio* **5**, e01017, doi:10.1128/mBio.01017-13 (2014).

59 Wang, G. Y. *et al.* Novel natural products from soil DNA libraries in a streptomycete host. *Org Lett* **2**, 2401-2404, doi:10.1021/ol005860z (2000).

60 Gillespie, D. E. *et al.* Isolation of antibiotics turbomycin a and B from a metagenomic library of soil microbial DNA. *Appl Environ Microbiol* **68**, 4301-4306, doi:10.1128/aem.68.9.4301-4306.2002 (2002).

61 Seow, K. T. *et al.* A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: a means to access and use genes from uncultured microorganisms. *J Bacteriol* **179**, 7360-7368, doi:10.1128/jb.179.23.7360-7368.1997 (1997).

62 Bhat, M. K. Cellulases and related enzymes in biotechnology. *Biotechnol Adv* **18**, 355-383, doi:10.1016/s0734-9750(00)00041-0 (2000).

63 Kim, S. J. *et al.* Characterization of a gene encoding cellulase from uncultured soil bacteria. *FEMS Microbiol Lett* **282**, 44-51, doi:10.1111/j.1574-6968.2008.01097.x (2008).

64 Jiang, C. *et al.* Characterization of a novel beta-glucosidase-like activity from a soil metagenome. *J Microbiol* **47**, 542-548, doi:10.1007/s12275-009-0024-y (2009).

65 Warnecke, F. *et al.* Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**, 560-565, doi:10.1038/nature06269 (2007).

66 Feng, Y. *et al.* Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases. *Appl Microbiol Biotechnol* **75**, 319-328, doi:10.1007/s00253-006-0820-9 (2007).

67 Pang, H. *et al.* Identification of cellulase genes from the metagenomes of compost soils and functional characterization of one novel endoglucanase. *Curr Microbiol* **58**, 404-408, doi:10.1007/s00284-008-9346-y (2009).

68 Ferrer, M. *et al.* Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environ Microbiol* **7**, 1996-2010, doi:10.1111/j.1462-2920.2005.00920.x (2005).

69 Ferrer, M. *et al.* Functional metagenomics unveils a multifunctional glycosyl hydrolase from the family 43 catalysing the breakdown of plant polymers in the calf rumen. *PLoS One* **7**, e38134, doi:10.1371/journal.pone.0038134 (2012).

70 Wang, L., Hatem, A., Catalyurek, U. V., Morrison, M. & Yu, Z. Metagenomic insights into the carbohydrate-active enzymes carried by the microorganisms adhering to solid digesta in the rumen of cows. *PLoS One* **8**, e78507, doi:10.1371/journal.pone.0078507 (2013).

71 Uchiyama, T., Abe, T., Ikemura, T. & Watanabe, K. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nat Biotechnol* **23**, 88-93, doi:10.1038/nbt1048 (2005).

72 Williamson, L. L. *et al.* Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. *Appl Environ Microbiol* **71**, 6335-6344, doi:10.1128/AEM.71.10.6335-6344.2005 (2005).

73 Uchiyama, T. & Miyazaki, K. Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes. *Appl Environ Microbiol* **76**, 7029-7035, doi:10.1128/AEM.00464-10 (2010).

74 Fernandez-Arrojo, L., Guazzaroni, M. E., Lopez-Cortes, N., Beloqui, A. & Ferrer, M. Metagenomic era for biocatalyst identification. *Curr Opin Biotechnol* **21**, 725-733, doi:10.1016/j.copbio.2010.09.006 (2010).

75 Ding, M. & Yelton, D. B. Cloning and analysis of the leuB gene of Leptospira interrogans serovar pomona. *J Gen Microbiol* **139**, 1093-1103, doi:10.1099/00221287-139-5-1093 (1993).

76 Ferreyra, R. G., Soncini, F. C. & Viale, A. M. Cloning, characterization, and functional expression in Escherichia coli of chaperonin (groESL) genes from the

phototrophic sulfur bacterium Chromatium vinosum. *J Bacteriol* **175**, 1514-1523, doi:10.1128/jb.175.5.1514-1523.1993 (1993).

77 Chavez, S., Reyes, J. C., Chauvat, F., Florencio, F. J. & Candau, P. The NADP-glutamate dehydrogenase of the cyanobacterium Synechocystis 6803: cloning, transcriptional analysis and disruption of the gdhA gene. *Plant Mol Biol* **28**, 173-188, doi:10.1007/BF00042048 (1995).

78 Rondon, M. R., Goodman, R. M. & Handelsman, J. The Earth's bounty: assessing and accessing soil microbial diversity. *Trends Biotechnol* **17**, 403-409, doi:10.1016/s0167-7799(99)01352-9 (1999).

79 Rondon, M. R., Raffel, S. J., Goodman, R. M. & Handelsman, J. Toward functional genomics in bacteria: analysis of gene expression in Escherichia coli from a bacterial artificial chromosome library of Bacillus cereus. *Proc Natl Acad Sci U S A* **96**, 6451-6455, doi:10.1073/pnas.96.11.6451 (1999).

80 Gabor, E. M., Alkema, W. B. & Janssen, D. B. Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ Microbiol* **6**, 879-886, doi:10.1111/j.1462-2920.2004.00640.x (2004).

81 Alvarez-Alvarez, R. *et al.* Expression of the endogenous and heterologous clavulanic acid cluster in Streptomyces flavogriseus: why a silent cluster is sleeping. *Appl Microbiol Biotechnol* **97**, 9451-9463, doi:10.1007/s00253-013-5148-7 (2013).

82 Warren, R. L. *et al.* Transcription of foreign DNA in Escherichia coli. *Genome Res* **18**, 1798-1805, doi:10.1101/gr.080358.108 (2008).

83 Gaida, S. M. *et al.* Expression of heterologous sigma factors enables functional screening of metagenomic and heterologous genomic libraries. *Nat Commun* **6**, 7045, doi:10.1038/ncomms8045 (2015).

84 Sørensen, H. P., Sperling-Petersen, H. U. & Mortensen, K. K. Production of recombinant thermostable proteins expressed in Escherichia coli: completion of protein synthesis is the bottleneck. *Journal of Chromatography B* **786**, 207-214, doi:10.1016/s1570-0232(02)00689-x (2003).

85 Baca, A. M. & Hol, W. G. Overcoming codon bias: a method for high-level overexpression of Plasmodium and other AT-rich parasite genes in Escherichia coli. *Int J Parasitol* **30**, 113-118, doi:10.1016/s0020-7519(00)00019-9 (2000).

86 Kim, R. *et al.* Overexpression of archaeal proteins in Escherichia coli. *Biotechnology Letters* **20**, 207-210, doi:10.1023/a:1005305330517 (1998).

87 Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: advances and challenges. *Front Microbiol* **5**, 172, doi:10.3389/fmicb.2014.00172 (2014).

88 Shis, D. L. & Bennett, M. R. Synthetic biology: the many facets of T7 RNA polymerase. *Mol Syst Biol* **10**, 745, doi:10.15252/msb.20145492 (2014).

89 Di Salvo, M., Puccio, S., Peano, C., Lacour, S. & Alifano, P. RhoTermPredict: an algorithm for predicting Rho-dependent transcription terminators based on Escherichia coli, Bacillus subtilis and Salmonella enterica databases. *BMC Bioinformatics* **20**, 117, doi:10.1186/s12859-019-2704-x (2019).

90 Santangelo, T. J. & Artsimovitch, I. Termination and antitermination: RNA polymerase runs a stop sign. *Nat Rev Microbiol* **9**, 319-329, doi:10.1038/nrmicro2560 (2011).

91 Mitra, A., Angamuthu, K., Jayashree, H. V. & Nagaraja, V. Occurrence, divergence and evolution of intrinsic terminators across eubacteria. *Genomics* **94**, 110-116, doi:10.1016/j.ygeno.2009.04.004 (2009).

92 Carafa, Y. d. A., Brody, E. & Thermes, C. Prediction of rho-independent Escherichia coli transcription terminators. *Journal of Molecular Biology* **216**, 835-858, doi:10.1016/s0022-2836(99)80005-9 (1990).

93 Lesnik, E. A. *et al.* Prediction of rho-independent transcriptional terminators in Escherichia coli. *Nucleic Acids Res* **29**, 3583-3594, doi:10.1093/nar/29.17.3583 (2001).

94 Ermolaeva, M. D., Khalak, H. G., White, O., Smith, H. O. & Salzberg, S. L. Prediction of transcription terminators in bacterial genomes. *J Mol Biol* **301**, 27-33, doi:10.1006/jmbi.2000.3836 (2000).

95 Kingsford, C. L., Ayanbule, K. & Salzberg, S. L. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biology* **8**, R22, doi:10.1186/gb-2007-8-2-r22 (2007).

96 Naville, M., Ghuillot-Gaudeffroy, A., Marchais, A. & Gautheret, D. ARNold: a web tool for the prediction of Rho-independent transcription terminators. *RNA Biol* **8**, 11-13, doi:10.4161/rna.8.1.13346 (2011).

97 Casjens, S. R. & Hendrix, R. W. Bacteriophage lambda: Early pioneer and still relevant. *Virology* **479-480**, 310-330, doi:10.1016/j.virol.2015.02.010 (2015).

98 Nudler, E. & Gottesman, M. E. Transcription termination and anti-termination in E. coli. *Genes Cells* **7**, 755-768, doi:10.1046/j.1365-2443.2002.00563.x (2002).

99 Roberts, J. W. Transcription termination and late control in phage lambda. *Proc Natl Acad Sci U S A* **72**, 3300-3304, doi:10.1073/pnas.72.9.3300 (1975).

100 Daniels, D. L. & Blattner, F. R. Nucleotide sequence of the Q gene and the Q to S intergenic region of bacteriophage lambda. *Virology* **117**, 81-92, doi:10.1016/0042-6822(82)90509-8 (1982).

101 Grayhack, E. J. & Roberts, J. W. The phage lambda Q gene product: activity of a transcription antiterminator in vitro. *Cell* **30**, 637-648, doi:10.1016/0092-8674(82)90260-4 (1982).

102 Yang, X. J., Hart, C. M., Grayhack, E. J. & Roberts, J. W. Transcription antitermination by phage lambda gene Q protein requires a DNA segment spanning the RNA start site. *Genes Dev* **1**, 217-226, doi:10.1101/gad.1.3.217 (1987).

103 Guo, J. & Roberts, J. W. DNA binding regions of Q proteins of phages lambda and phi80. *J Bacteriol* **186**, 3599-3608, doi:10.1128/JB.186.11.3599-3608.2004 (2004).

104 Deighan, P. & Hochschild, A. The bacteriophage lambdaQ anti-terminator protein regulates late gene expression as a stable component of the transcription

elongation complex. *Mol Microbiol* **63**, 911-920, doi:10.1111/j.1365-2958.2006.05563.x (2007).

105 Ko, D. C., Marr, M. T., Guo, J. & Roberts, J. W. A surface of Escherichia coli sigma 70 required for promoter function and antitermination by phage lambda Q protein. *Genes Dev* **12**, 3276-3285, doi:10.1101/gad.12.20.3276 (1998).

106 Nickels, B. E., Roberts, C. W., Sun, H., Roberts, J. W. & Hochschild, A. The σ70 Subunit of RNA Polymerase Is Contacted by the λQ Antiterminator during Early Elongation. *Molecular Cell* **10**, 611-622, doi:10.1016/s1097-2765(02)00648-2 (2002).

107 Roberts, J. W. *et al.* Antitermination by bacteriophage lambda Q protein. *Cold Spring Harb Symp Quant Biol* **63**, 319-325, doi:10.1101/sqb.1998.63.319 (1998).

108 Shi, J. *et al.* Structural basis of Q-dependent transcription antitermination. *Nat Commun* **10**, 2925, doi:10.1038/s41467-019-10958-8 (2019).

109 Grayhack, E., Yang, X., Lau, L. & Roberts, J. Phage lambda gene Q antiterminator recognizes RNA polymerase near the promoter and accelerates it through a pause site. *Cell* **42**, 259-269, doi:10.1016/s0092-8674(85)80121-5 (1985).

110 Nickels, B. E., Roberts, C. W., Roberts, J. W. & Hochschild, A. RNA-mediated destabilization of the sigma(70) region 4/beta flap interaction facilitates engagement of RNA polymerase by the Q antiterminator. *Mol Cell* **24**, 457-468, doi:10.1016/j.molcel.2006.09.014 (2006).

111 Franklin, N. C. & Doelling, J. H. Overexpression of N antitermination proteins of bacteriophages lambda, 21, and P22: loss of N protein specificity. *J Bacteriol* **171**, 2513-2522, doi:10.1128/jb.171.5.2513-2522.1989 (1989).

112 Ghosh, B. *et al.* Specificity of antitermination mechanisms. Suppression of the terminator cluster T1-T2 of Escherichia coli ribosomal RNA operon, rrnB, by phage lambda antiterminators. *J Mol Biol* **222**, 59-66, doi:10.1016/0022-2836(91)90737-q (1991).

113 Yang, X. J. & Roberts, J. W. Gene Q antiterminator proteins of Escherichia coli phages 82 and lambda suppress pausing by RNA polymerase at a rho-dependent terminator and at other sites. *Proc Natl Acad Sci U S A* **86**, 5301-5305, doi:10.1073/pnas.86.14.5301 (1989).

114 National Research Council Committee on Metagenomics, C. & Functional, A. in *The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet The National Academies Collection: Reports funded by National Institutes of Health* (National Academies Press (US); National Academy of Sciences., 2007).

115 Thomas, T., Gilbert, J. & Meyer, F. Metagenomics - a guide from sampling to data analysis. *Microb Inform Exp* **2**, 3, doi:10.1186/2042-5783-2-3 (2012).

116 Gilbert, J. A. & Dupont, C. L. Microbial metagenomics: beyond the genome. *Ann Rev Mar Sci* **3**, 347-371, doi:10.1146/annurev-marine-120709-142811 (2011).

117 Koppolu, V. & Vasigala, V. K. Role of Escherichia coli in Biofuel Production. *Microbiol Insights* **9**, 29-35, doi:10.4137/MBI.S10878 (2016).

118 Saini, M., Chiang, C. J., Li, S. Y. & Chao, Y. P. Production of biobutanol from cellulose hydrolysate by the Escherichia coli co-culture system. *FEMS Microbiol Lett* **363**, doi:10.1093/femsle/fnw008 (2016).

119 Wilhelm, R. C., Singh, R., Eltis, L. D. & Mohn, W. W. Bacterial contributions to delignification and lignocellulose degradation in forest soils with metagenomic and quantitative stable isotope probing. *ISME J* **13**, 413-429, doi:10.1038/s41396-018-0279-6 (2019).

120 Lee, S., Kang, M., Bae, J. H., Sohn, J. H. & Sung, B. H. Bacterial Valorization of Lignin: Strains, Enzymes, Conversion Pathways, Biosensors, and Perspectives. *Front Bioeng Biotechnol* **7**, 209, doi:10.3389/fbioe.2019.00209 (2019).

121 Peng, Q. *et al.* Isolation of a novel alkaline-stable lipase from a metagenomic library and its specific application for milkfat flavor production. *Microb Cell Fact* **13**, 1, doi:10.1186/1475-2859-13-1 (2014).

122 Richardson, T. H. *et al.* A novel, high performance enzyme for starch liquefaction. Discovery and optimization of a low pH, thermostable alpha-amylase. *J Biol Chem* **277**, 26501-26507, doi:10.1074/jbc.M203183200 (2002).

123 Liang, S. *et al.* Activities of constitutive promoters in Escherichia coli. *J Mol Biol* **292**, 19-37, doi:10.1006/jmbi.1999.3056 (1999).

124 Cambray, G. *et al.* Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Res* **41**, 5139-5148, doi:10.1093/nar/gkt163 (2013).

125 Chen, Y. J. *et al.* Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat Methods* **10**, 659-664, doi:10.1038/nmeth.2515 (2013).

126 Haldimann, A. & Wanner, B. L. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J Bacteriol* **183**, 6384-6393, doi:10.1128/JB.183.21.6384-6393.2001 (2001).

127 Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640-6645, doi:10.1073/pnas.120163297 (2000).

128 Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* **3**, 1101-1108, doi:10.1038/nprot.2008.73 (2008).

129 Mirete, S. *et al.* Salt resistance genes revealed by functional metagenomics from brines and moderate-salinity rhizosphere within a hypersaline environment. *Front Microbiol* **6**, 1121, doi:10.3389/fmicb.2015.01121 (2015).

130 Elkins, C. A. & Savage, D. C. Identification of genes encoding conjugated bile salt hydrolase and transport in Lactobacillus johnsonii 100-100. *J Bacteriol* **180**, 4344-4349, doi:10.1128/JB.180.17.4344-4349.1998 (1998).

131 Kevles, D. J. Ananda Chakrabarty wins a patent: biotechnology, law, and society, 1972-1980. *Hist Stud Phys Biol Sci* **25**, 111-135 (1994).

132 Nikel, P. I., Martinez-Garcia, E. & de Lorenzo, V. Biotechnological domestication of pseudomonads using synthetic biology. *Nat Rev Microbiol* **12**, 368-379, doi:10.1038/nrmicro3253 (2014).

133 Nielsen, D. R. *et al.* Engineering alternative butanol production platforms in heterologous bacteria. *Metab Eng* **11**, 262-273, doi:10.1016/j.ymben.2009.05.003 (2009).

134 Loeschcke, A. & Thies, S. Pseudomonas putida-a versatile host for the production of natural products. *Appl Microbiol Biotechnol* **99**, 6197-6214, doi:10.1007/s00253-015-6745-4 (2015).

135 Fernandez, M., Niqui-Arroyo, J. L., Conde, S., Ramos, J. L. & Duque, E. Enhanced tolerance to naphthalene and enhanced rhizoremediation performance for Pseudomonas putida KT2440 via the NAH7 catabolic plasmid. *Appl Environ Microbiol* **78**, 5104-5110, doi:10.1128/AEM.00619-12 (2012).

136 Inoue, A. & Horikoshi, K. A Pseudomonas thrives in high concentrations of toluene. *Nature* **338**, 264-266, doi:10.1038/338264a0 (1989).

137 Ru, J., Huo, Y. & Yang, Y. Microbial Degradation and Valorization of Plastic Wastes. *Front Microbiol* **11**, 442, doi:10.3389/fmicb.2020.00442 (2020).

138 Nikel, P. I. & de Lorenzo, V. Pseudomonas putida as a functional chassis for industrial biocatalysis: From native biochemistry to trans-metabolism. *Metab Eng* **50**, 142-155, doi:10.1016/j.ymben.2018.05.005 (2018).

139 Borrero-de Acuna, J. M. *et al.* Production of medium chain length polyhydroxyalkanoate in metabolic flux optimized Pseudomonas putida. *Microb Cell Fact* **13**, 88, doi:10.1186/1475-2859-13-88 (2014).

140 Meijnen, J. P., de Winde, J. H. & Ruijssenaars, H. J. Establishment of oxidative D-xylose metabolism in Pseudomonas putida S12. *Appl Environ Microbiol* **75**, 2784-2791, doi:10.1128/AEM.02713-08 (2009).

141 Elmore, J. R., Furches, A., Wolff, G. N., Gorday, K. & Guss, A. M. Development of a high efficiency integration system and promoter library for rapid modification of Pseudomonas putida KT2440. *Metab Eng Commun* **5**, 1-8, doi:10.1016/j.meteno.2017.04.001 (2017).

142 Calero, P. & Nikel, P. I. Chasing bacterial chassis for metabolic engineering: a perspective review from classical to non-traditional microorganisms. *Microb Biotechnol* **12**, 98-124, doi:10.1111/1751-7915.13292 (2019).

143 Haynes, A. R. *Characterization of extremophilic bacteria for potential in the biofuel and bioprocess industries* Ph.D. dissertation, Texas A&M University, College Station, (2014).

144 Schafer, A. *et al.* Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum. *Gene* **145**, 69-73, doi:10.1016/0378-1119(94)90324-7 (1994).

145 Johnson, C. W. & Beckham, G. T. Aromatic catabolic pathway selection for optimal production of pyruvate and lactate from lignin. *Metab Eng* **28**, 240-247, doi:10.1016/j.ymben.2015.01.005 (2015).

146 Brown, W. R., Lee, N. C., Xu, Z. & Smith, M. C. Serine recombinases as tools for genome engineering. *Methods* **53**, 372-379, doi:10.1016/j.ymeth.2010.12.031 (2011).

147 Stark, W. M. The Serine Recombinases. *Microbiol Spectr* **2**, doi:10.1128/microbiolspec.MDNA3-0046-2014 (2014).

148 Smith, M. C. M. Phage-encoded Serine Integrases and Other Large Serine Recombinases. *Microbiol Spectr* **3**, doi:10.1128/microbiolspec.MDNA3-0059-2014 (2015).

149 Xu, Z. & Brown, W. R. Comparison and optimization of ten phage encoded serine integrases for genome engineering in Saccharomyces cerevisiae. *BMC Biotechnol* **16**, 13, doi:10.1186/s12896-016-0241-5 (2016).

150 Kim, A. I. *et al.* Mycobacteriophage Bxb1 integrates into the Mycobacterium smegmatis groEL1 gene. *Mol Microbiol* **50**, 463-473, doi:10.1046/j.1365-2958.2003.03723.x (2003).

151 Caraway, E. B. *Sequencing, Pipeline Development, and Select Comparative Analysis of 64 High-Quality Draft Genomes of Extremophilic Bacteria Isolated from Communities in Carboxylate Platform Fermentations* Master of Science thesis, Texas A&M University, College Station, (2016).

152 Shcherbo, D. *et al.* Bright far-red fluorescent protein for whole-body imaging. *Nat Methods* **4**, 741-746, doi:10.1038/nmeth1083 (2007).

153 Chaves, J. E. *et al.* Evaluation of chromosomal insertion loci in the Pseudomonas putida KT2440 genome for predictable biosystems design. *Metab Eng Commun* **11**, e00139, doi:10.1016/j.mec.2020.e00139 (2020).

154 Shcherbo, D. *et al.* Far-red fluorescent tags for protein imaging in living tissues. *Biochem J* **418**, 567-574, doi:10.1042/BJ20081949 (2009).

155 Ku, H. H. Notes on the Use of Propagation of Error Formulas. *J. Res. Nat. Bur. Stand* **70C**, 263-273 (1966).

156 Yin, J., Chen, J. C., Wu, Q. & Chen, G. Q. Halophiles, coming stars for industrial biotechnology. *Biotechnol Adv* **33**, 1433-1442, doi:10.1016/j.biotechadv.2014.10.008 (2015).

157 Bashir, Y., Pradeep Singh, S. & Kumar Konwar, B. Metagenomics: An Application Based Perspective. *Chinese Journal of Biology* **2014**, 1-7, doi:10.1155/2014/146030 (2014).

158 Amoozegar, M. A., Safarpour, A., Noghabi, K. A., Bakhtiary, T. & Ventosa, A. Halophiles and Their Vast Potential in Biofuel Production. *Front Microbiol* **10**, 1895, doi:10.3389/fmicb.2019.01895 (2019).

159 Jonsson, L. J. & Martin, C. Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects. *Bioresour Technol* **199**, 103-112, doi:10.1016/j.biortech.2015.10.009 (2016).

160 de Oliveira, D. M. *et al.* Ferulic acid: a key component in grass lignocellulose recalcitrance to hydrolysis. *Plant Biotechnol J* **13**, 1224-1232, doi:10.1111/pbi.12292 (2015).

161 Seo, J. S., Keum, Y. S. & Li, Q. X. Bacterial degradation of aromatic compounds. *Int J Environ Res Public Health* **6**, 278-309, doi:10.3390/ijerph6010278 (2009).

162 Plaggenborg, R., Overhage, J., Steinbuchel, A. & Priefert, H. Functional analyses of genes involved in the metabolism of ferulic acid in Pseudomonas putida KT2440. *Appl Microbiol Biotechnol* **61**, 528-535, doi:10.1007/s00253-003-1260-4 (2003).

163 Ravi, K., Garcia-Hidalgo, J., Gorwa-Grauslund, M. F. & Liden, G. Conversion of lignin model compounds by Pseudomonas putida KT2440 and isolates from compost. *Appl Microbiol Biotechnol* **101**, 5059-5070, doi:10.1007/s00253-017-8211-y (2017).

164 Romero-Steiner, S., Parales, R. E., Harwood, C. S. & Houghton, J. E. Characterization of the pcaR regulatory gene from Pseudomonas putida, which is required for the complete degradation of p-hydroxybenzoate. *J Bacteriol* **176**, 5771-5779, doi:10.1128/jb.176.18.5771-5779.1994 (1994).

165 Buchan, A., Collier, L. S., Neidle, E. L. & Moran, M. A. Key aromatic-ring-cleaving enzyme, protocatechuate 3,4-dioxygenase, in the ecologically important marine Roseobacter lineage. *Appl Environ Microbiol* **66**, 4662-4672, doi:10.1128/aem.66.11.4662-4672.2000 (2000).

166 Iwagami, S. G., Yang, K. & Davies, J. Characterization of the protocatechuic acid catabolic gene cluster from Streptomyces sp. strain 2065. *Appl Environ Microbiol* **66**, 1499-1508, doi:10.1128/aem.66.4.1499-1508.2000 (2000).

167 Romero, P. & Karp, P. PseudoCyc, a pathway-genome database for Pseudomonas aeruginosa. *J Mol Microbiol Biotechnol* **5**, 230-239, doi:10.1159/000071075 (2003).

168 Taylor, M. P., Esteban, C. D. & Leak, D. J. Development of a versatile shuttle vector for gene expression in Geobacillus spp. *Plasmid* **60**, 45-52, doi:10.1016/j.plasmid.2008.04.001 (2008).

169 Clarkson, S. M. *et al.* Construction and Optimization of a Heterologous Pathway for Protocatechuate Catabolism in Escherichia coli Enables Bioconversion of Model Aromatic Compounds. *Appl Environ Microbiol* **83**, doi:10.1128/AEM.01313-17 (2017).

170 Basta, D. W. *et al.* Characterization of the ospZ promoter in Shigella flexneri and its regulation by VirB and H-NS. *J Bacteriol* **195**, 2562-2572, doi:10.1128/JB.00212-13 (2013).

171 Vogt, S. L. *et al.* Characterization of the Citrobacter rodentium Cpx regulon and its role in host infection. *Mol Microbiol* **111**, 700-716, doi:10.1111/mmi.14182 (2019).

172 Kim, D. *et al.* Comparative analysis of regulatory elements between Escherichia coli and Klebsiella pneumoniae by genome-wide transcription start site profiling. *PLoS Genet* **8**, e1002867, doi:10.1371/journal.pgen.1002867 (2012).

173 Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research* **47**, W636-W641, doi:10.1093/nar/gkz268 (2019).

174 Yansura, D. G. & Henner, D. J. Use of the Escherichia coli lac repressor and operator to control gene expression in Bacillus subtilis. *Proc Natl Acad Sci U S A* **81**, 439-443, doi:10.1073/pnas.81.2.439 (1984).

175 Takagi, H., Yamagata, H. & Udaka, S. Gene expression in Bacillus brevis under control of the Lac repressor-operator system from Escherichia coli. *Journal of Fermentation and Bioengineering* **74**, 1-6, doi:10.1016/0922-338x(92)90258-v (1992).

176 Henkin, T. M. & Sonenshein, A. L. Mutations of the Escherichia coli lacUV5 promoter resulting in increased expression in Bacillus subtilis. *Mol Gen Genet* **209**, 467-474, doi:10.1007/BF00331151 (1987).

177 Natori, Y., Kano, Y. & Imamoto, F. Characterization and promoter selectivity of Lactobacillus acidophilus RNA polymerase. *Biochimie* **70**, 1765-1774, doi:10.1016/0300-9084(88)90036-3 (1988).

178 Brunschwig, E. & Darzins, A. A two-component T7 system for the overexpression of genes in Pseudomonas aeruginosa. *Gene* **111**, 35-41, doi:10.1016/0378-1119(92)90600-t (1992).

179 Bethesda-Research-Laboratories (1986) BRL pUC host: Escherichia coli DH5a competent cells. Bethesda Research Laboratories Focus 8:9-10. .

180 Hols, P. *et al.* Efficient secretion of the model antigen M6-gp41E in Lactobacillus plantarum NCIMB 8826. *Microbiology (Reading)* **143 (Pt 8)**, 2733-2741, doi:10.1099/00221287-143-8-2733 (1997).

181 Bagdasarian, M. *et al.* Specific-purpose plasmid cloning vectors II. Broad host range, high copy number, RSF 1010-derived vectors, and a host-vector system for gene cloning in Pseudomonas. *Gene* **16**, 237-247, doi:10.1016/0378-1119(81)90080-9 (1981).

182 Pfleger, B. F., Fawzi, N. J. & Keasling, J. D. Optimization of DsRed production in Escherichia coli: effect of ribosome binding site sequestration on translation efficiency. *Biotechnol Bioeng* **92**, 553-558, doi:10.1002/bit.20630 (2005).

183 Elmore, J. R. *et al.* The SAGE genetic toolkit enables highly efficient, iterative site-specific genome engineering in bacteria. *bioRxiv*, doi:10.1101/2020.06.28.176339 (2020).

184 Shapiro, J. A. The significance of bacterial colony patterns. *BioEssays* **17**, 597-607 (1995).

185 Liu, C. *et al.* Sequential establishment of stripe patterns in an expanding cell population. *Science* **334**, 238-241, doi:10.1126/science.1209042 (2011).

186 Rauprich, O. *et al.* Periodic phenomena in Proteus mirabilis swarm colony development. *J Bacteriol* **178**, 6525-6538, doi:10.1128/jb.178.22.6525-6538.1996 (1996).

187 Lominski, I. & Lendrum, A. C. The mechanism of swarming of proteus. *J Pathol Bacteriol* **59**, 688-691, doi:10.1002/path.1700590423 (1947).

188 Williams, F. D. & Schwarzhoff, R. H. Nature of the swarming phenomenon in Proteus. *Annu Rev Microbiol* **32**, 101-122, doi:10.1146/annurev.mi.32.100178.000533 (1978).

189 Kearns, D. B. A field guide to bacterial swarming motility. *Nat Rev Microbiol* **8**, 634-644, doi:10.1038/nrmicro2405 (2010).

190 Fujikawa, H. Diversity of the growth patterns of Bacillus subtilis colonies on agar plates. *Fems Microbiol Ecol* **13**, 159-168, doi:10.1111/j.1574-6941.1994.tb00062.x (1994).

191 Ben-Jacob, E. *et al.* Generic modelling of cooperative growth patterns in bacterial colonies. *Nature* **368**, 46-49, doi:10.1038/368046a0 (1994).

192 Swiecicki, J. M., Sliusarenko, O. & Weibel, D. B. From swimming to swarming: Escherichia coli cell motility in two-dimensions. *Integr Biol (Camb)* **5**, 1490-1494, doi:10.1039/c3ib40130h (2013).

193 Chilcott, G. S. & Hughes, K. T. Coupling of flagellar gene expression to flagellar assembly in Salmonella enterica serovar typhimurium and Escherichia coli. *Microbiol Mol Biol Rev* **64**, 694-708, doi:10.1128/mmbr.64.4.694-708.2000 (2000).

194 Beagle, S. D. *Characterization of a Prokaryotic K+ Channel from Escherichia coli*, Ph.D. dissertation, Texas A&M University, College Station, (2019).

195 Lehnen, D. *et al.* LrhA as a new transcriptional key regulator of flagella, motility and chemotaxis genes in Escherichia coli. *Mol Microbiol* **45**, 521-532, doi:10.1046/j.1365-2958.2002.03032.x (2002).

196 Peterson, C. N., Carabetta, V. J., Chowdhury, T. & Silhavy, T. J. LrhA regulates rpoS translation in response to the Rcs phosphorelay system in Escherichia coli. *J Bacteriol* **188**, 3175-3181, doi:10.1128/JB.188.9.3175-3181.2006 (2006).

197 Gibson, K. E. & Silhavy, T. J. The LysR homolog LrhA promotes RpoS degradation by modulating activity of the response regulator sprE. *J Bacteriol* **181**, 563-571, doi:10.1128/JB.181.2.563-571.1999 (1999).

198 Blumer, C. *et al.* Regulation of type 1 fimbriae synthesis and biofilm formation by the transcriptional regulator LrhA of Escherichia coli. *Microbiology (Reading)* **151**, 3287-3298, doi:10.1099/mic.0.28098-0 (2005).

199 Martinez-Garcia, E., Calles, B., Arevalo-Rodriguez, M. & de Lorenzo, V. pBAM1: an all-synthetic genetic tool for analysis and construction of complex bacterial phenotypes. *BMC Microbiol* **11**, 38, doi:10.1186/1471-2180-11-38 (2011).

200 Lodge, J. K., Westonhafer, K. & Berg, D. E. Transposon Tn5 Target Specificity - Preference for Insertion at G/C Pairs. *Genetics* **120**, 645-650 (1988).

201 Ishizuka, H., Hanamura, A., Inada, T. & Aiba, H. Mechanism of the down-regulation of cAMP receptor protein by glucose in Escherichia coli: role of autoregulation of the crp gene. *EMBO J* **13**, 3077-3082 (1994).

202 Soutourina, O. *et al.* Multiple control of flagellum biosynthesis in Escherichia coli: role of H-NS protein and the cyclic AMP-catabolite activator protein complex in transcription of the flhDC master operon. *J Bacteriol* **181**, 7500-7508, doi:10.1128/JB.181.24.7500-7508.1999 (1999).

203 Soutourina, O. A. & Bertin, P. N. Regulation cascade of flagellar expression in Gram-negative bacteria. *FEMS microbiology reviews* **27**, 505-523, doi:10.1016/S0168-6445(03)00064-0 (2003).

APPENDIX A

SUPPLEMENTAL SECTION

Table 6 List of strains used in this study.

Sr. No.	Organism	Strain	Genotype/Features	Purpose in this study	Reference
1	*E. coli*	BW25113	*lacI*$^+$ Δ*lacZ4787::rrnB-3 hsdR514* Δ(*araBAD*)*567* Δ(*rhaBAD*)*568 rph-1,* λ^-, F^-	Used to perform the mRNA analysis	[127]
2	*E. coli*	BW25141	*lacI*$^+$ Δ*lacZ4787::rrnB-3 hsdR514* Δ(*araBAD*)*567* Δ(*rhaBAD*)*568* Δ*phoBR580 rph-1 galU95* Δ*endA9::FRT* Δ*uid*A3::*pir*$^+$ *recA1,* λ^-, F^-	Used to maintain all the pLA2 derivatives	[127]
3	*E. coli*	DH5alpha	F^- *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^-m_K^+$), λ^-	Used to maintain the constructed plasmids	[179]

Table 6 Continued

Sr. No.	Organism	Strain	Genotype/Features	Purpose in this study	Reference
4	*L. plantarum*	ATCC BAA793	Isolated from human saliva	Used to amplify the cps3 cluster	[180]
5	*P. putida*	KT2440	$r^{-}m^{+}$ (r = host-specific restriction; m = host specific modification of DNA)	Used to test mKate2 expression	[181]

Table 7 List of plasmids used in this study.

Plasmid No.	Description	Function	Reference
1	pBAD-RFP(EC)	The plasmid was used as a backbone for cloning the heterologous DNA	[182]
2	pLA2	The plasmid was altered to introduce the gene encoding λQ	[126]
3	pINT-ts	The CRIM helper plasmid used to integrate the CRIM plasmid pLA2	[126]

Table 7 Continued

Plasmid No.	Description	Function	Reference
4	pJKK1	The plasmid was used to introduce the gene encoding λQ on the *E. coli* genome at the λ phage attachment site	This work
5	pJKK2	The plasmid carries the lac promoter in addition with the cps3 cluster	This work
6	pJKK3	The plasmid carries the P_R' promoter in addition with the cps3 cluster	This work
7	pGW31	The plasmid was used to transform and express the BxB1 integrase	[183]
8	pK18mobsacB	The plasmid was used to integrate λQ encoding gene using homologous recombination	[144]
9	pJKK4	Suicide vector to replace the PP_RS17485-PP_RS17495 with Q encoding gene in *P. putida* KT2440	This work

Table 8 List of promoter and gene sequences used in this study.

Sr. No.	Description	Sequence
1	Anti-terminator λQ encoding gene for use in *E. coli*	5'- ATGAGACTCGAAAGCGTAGCTAAATTTCATTCGCCAAAAAGCCCGATGATGAGCGACTCACCAC GGGCCACGGCTTCTGACTCTCTTTCCGGTACTGATGTGATGGCTGCTATGGGGATGGCGCAATCA CAAGCCGGATTCGGTATGGCTGCATTCTGCGGTAAGCACGAACTCAGCCAGAACGACAAACAAA AGGCTATCAACTATCTGATGCAATTTGCACACAAGGTATCGGGGAAATACCGTGGTGTGGCAAA GTTTGAAGGAAATACTAAGGCAAAGGTACTGCAAGTGCTCGCAACATTCGCTTATGCGGATTAT TGCCGTAGTGCCGCGACGCCGGGGGCAAGATGCAGAGATTGCCACGGTACAGGCCGTGCGGTTG ATATTGCCAAAACAGAGCTGTGGGGGAGAGTTGTCGAGAAAGAGTGCGGAAGATGCAAAGGCG TCGGCTATTCAAGGATGCCAGCAAGCGCAGCATATCGCGCTGTGACGATGCTAATCCCAAACCT TACCCAACCCACCTGGTCACGCACTGTTAAGCCGCTGTATGACGCTCTGGTGGTGCAATGCCACA AAGAAGAGTCAATCGCAGACAACATTTTGAATGCGGTCACACGTTAG-3'
2	P_R' promoter	5'-TTGACTTATTGAATAAAATTGGGTAAATTTGACTCAACGAT-3'
3	P_{lac} promoter	5'-TTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAA-3'
4	Truncated PP_RS17495 – PP_RS17485 (pJAK2)	5'- AATTCCAACAGTTCAGTCCTTTCGTCATCCGAGAGGTTGGATGTCGCCTGGTGATCGCTGTCGGC GGCGATGTGCTTCAACTCCTTCAACAGGGCCTCGCCCGATTTGCTCAGGAAAATCCCATACGAGC GCTTGTCCGGTTTGCACCGCACGCGCACGGCCAGCGCGCGCTCTTCCAGTTTGTTCAGCATTGGC ACGACTTGCGGGGGGTCGATTGCCAACGCACGCGCCAGGTCGGCCTGCATCAGCCCGGGGTTCT GGTCGATGATCGCCAGGGCGGAAAATTGGGCCGGGCGCAGGTCATGGGCCGAGAGCCGGGCAA TAAGGTTCTGAAACAGCTTCAATTGCGCGCGTCGCAAGGCGTAGCCGATCAAGTCTTCAAGCAC CCCTTCTCCCACAGGAGCAGCCACGCAAGCGTCGTCGGTACTACGGGCAGACCTAGCCATCAGG TTTCATGCTCCTCGATCATGGGTAATAAAGTTACCTATTTTGCCTGTCCTTATGCGATTCGGCTAG AGAGGTTCTGGAAAAAGGCAGCGCGCCTAACCCCAGGACAACGGTAAAACTGTTAATGGCTGAT TGACATAACTAATTTGACCCGTTAGCGTGGCCCCATCACCTCGAACAACAAGAGTGCAGAACCA TGAGCAAATACCTGTACCGTGGCGAAGACCCCGACGCATTGCACGCCAAGGTGCGGCCTTGAGT GCGGCCGGGCTGTGCACCGTCTTCGAGGACACGGCCATCGTCCAGATGGTGCGCACGCCCTGGG TCGACCTGGGTGGCGCATTGGCCGCGATCTCGCCCATCGACCTGGGCATTCAAGCAGGGCGCGC TGTCCTCGCTCGCGCTGCTGCTGCTCCTCAGGCGGTAGACAGTGTGCTCGCCGGCAGCATGGCTC AGGCCAGCTTCGACGCCTACATGCTGCCACGCCATGTCGGCCTCTACTGTGGCGTGCCGCAGGCG GTACCGGCACTGGCGGTGCAGCGCATCTGTGGCACGGGCCTTGAACTGTTGCGTCAGGCCGGTG AACAGCTGCGCAGTGGCGCCCGCCAGGTGCTGTGCGTGAAGCTTGGCACTGGCCGTCGTTTTAC AACGTCGTGACTGGGAAAACCCT-3'

Table 8 Continued

Sr. No.	Description	Sequence
5	mKate2_KanR_Bxb1-attP (pJAK3)	5'- GCTAGCCTTAAGATTAACTCACACAGGAGATATCATATGGTCAGCGAACTCATTAAAGAGAACA TGCACATGAAACTCTACATGGAGGGCACGGTGAATAATCACCATTTTAAGTGCACCTCCGAAGG TGAAGGCAAGCCGTACGAGGGGACGCAGACGATGCGTATTAAAGCGGTGGAGGGGGGCCCTCT CCCATTCGCGTTCGACATCCTCGCAACCAGCTTCATGTACGGGAGCAAGACGTTCATTAACCATA CCCAGGGCATCCCAGACTTTTTCAAACAATCCTTTCCAGAAGGGTTTACGTGGGAACGCGTCACC ACGTATGAGGATGGGGGGGTCCTCACGGCAACGCAGGACACGAGCCTCCAAGACGGCTGCCTG ATCTACAATGTGAAGATTCGTGGGGTGAATTTCCCATCGAATGGGCCCGTGATGCAAAAAAAGA CCCTCGGTTGGGAGGCTTCCACGGAGACCCTCTACCCAGCGGACGGTGGCTTGGAGGGCCGCGC GGATATGGCGCTCAAGTTGGTCGGTGGGGGTCATCTGATCTGCAACCTCAAAACGACGTATCGC TCGAAGAAGCCAGCGAAAAATCTCAAAATGCCCGGCGTCTACTACGTGGACCGTCGTCTGGAAC GGATCAAAGAAGCCGATAAAGAGACGTACGTCGAGCAGCACGAAGTGGCGGTCGCTCGCTATT GTGATTTGCCGAGCAAGCTCGGGCACCGCTGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGG AAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTGGGAGACCAGAAACAAAAAAAGGCCGCGT TAGCGGCCTTCAATAATTGGACCTGGCTCCTAGACGAACAATAAGGCCTCCCTAACGGGGGGCC TTTTTTATTGATAACAAAAAAGAATTCATCAGAAGAACTCGTCAAGAAGGCGATAGAAGGCGAT GCGCTGCGAATCGGGAGCGGCGATACCGTAAAGCACGAGGAAGCGGTCAGCCCATTCGCCGCC AAGCTCTTCAGCAATATCACGGGTAGCCAACGCTATGTCCTGATAGCGGTCCGCCACACCCAGC CGGCCACAGTCGATGAATCCAGAAAAGCGGCCATTTTCCACCATGATATTCGGCAAGCAGGCAT CGCCATGGGTCACGACGAGATCCTCGCCGTCGGGCATCCGCGCCTTGAGCCTGGCGAACAGTTC GGCTGGCGCGAGCCCCTGATGCTCTTCGTCCAGATCATCCTGATCGACAAGACCGGCTTCCATCC GAGTACGTGCTCGCTCGATGCGATGTTTCGCTTGGTGGTCGAATGGGCAGGTAGCCGGATCAAG CGTATGCAGCCGCCGCATTGCATCAGCCATGATGGATACTTTCTCGGCAGGAGCAAGGTGAGAT GACAGGAGATCCTGCCCCGGCACTTCGCCCAATAGCAGCCAGTCCCTTCCCGCTTCAGTGACAAC GTCGAGCACAGCTGCGCAAGGAACGCCCGTCGTGGCCAGCCACGATAGCCGCGCTGCCTCGTCT TGGAGTTCATTCAGGGCACCGGACAGGTCGGTCTTGACAAAAAGAACCGGGCGCCCCTGCGCTG ACAGCCGGAACACGGCGGCATCAGAGCAGCCGATTGTCTGTTGTGCCCAGTCATAGCCGAATAG CCTCTCCACCCAAGCGGCCGGAGAACCTGCGTGCAATCCATCTTGTTCAATCATGCGAAACGATC CTCATCCTGTCTCTTGATCAGATCTTGATCCCCTGCGCCATCAGATCCTTGGCGGCAAGAAAGCC ATCCAGTTTACTTTGCAGGGCTTCCCAACCTTACCAGAGGGCGCCCCAGCTGGCAATTCCGGTTC GCTTGCTGTCCATAAAACCGCCCAGTCTAGCTATCGCCATGTAAGCCCACTGCAAGCTACCTGCT TTCTCTTTGCGGCGGCCGCGTCGTGGTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGT ACAAACCCCGACGCTAG-3'

Table 8 Continued

Sr. No.	Description	Sequence
6	Anti-terminator λQ encoding gene (pJAK4) for use in *P. putida* KT2440	5'- CATATGTTGATTCAGGTAACAGGGAGAAAGGCGCATGAGACTCGAAAGCGTAGCTAAATTTCAT TCGCCAAAAAGCCCGATGATGAGCGACTCACCACGGGCCACGGCTTCTGACTCTCTTTCCGGTAC TGATGTGATGGCTGCTATGGGGATGGCGCAATCACAAGCGGGATTCGGTATGGCTGCATTCTGC GGTAAGCACGAACTCAGCCAGAACGACAAACAAAAGGCTATCAACTATCTGATGCAATTTGCAC ACAAGGTATCGGGGAAATACCGCGGTGTGGCGAAGCTTGAAGGAAATACTAAGGCAAAGGTAC TGCAAGTGCTCGCAACATTGCCTTATGCGGATTATTGCCGTAGTGGCGCGACGCCGGGGGCAAG ATGCAGAGATTGCCATGGTACAGGCCGTGCGGTTGATATTGCCAAAACAGAGCTGTGGGGGAGA GTTGTCGAGAAAGAGTGCGGAAGATGCAAAGGCGTCGGCTATTCAAGGATGCCAGCAAGCGCA GCATATCGCGCTGTGACGATGCTAATCCCAAACCTTACCCAACCCACCTGGTCACGCACTGTTAA GCCGCTGTATGACGCTCTGGTGGTGCAATGCCACAAAGAAGAGTCAATCGCAGACAACATTTTG AATGCGATCACACGTTAG-3'

Table 9 Sequences of all the primers used in this study.

Primer. No.	**Primer Name**	**Sequence**
P1	JK_NdeI_Q_F	5-GGCAATTC CAT ATGAGACTCGAAAGCGTAGC-3'
P2	JK_KpnI_Q_R	5'-TCT GGTACC CTAACGTGTGACCGCATTC-3-
P3	JK_invPBAD_Xho-Xba_F	5'-CAATTATGCTCGAGTCTAGA GATGAATTCCAGCTGAGC-3'
P4	JK_invPBAD_Xho-Xba_R	5'-AATTCATCTCTAGACTCGAG CATAATTGGTAACGAATCAGAC-3'

Table 9 Continued

Primer. No.	**Primer Name**	**Sequence**
P5	JK_P_R'OLP_cps3_Xho-2_F	5'-CCGCTCGAGTTGACTTATTGAATAAAATTGGGTAAATTTGACTCAACGATATGACAAAAATTGCATGTGTG-3'
P6	JK_cps3-1_Not-Xba_rev	5'-GC TCTAGAGTGCGGCCGCCATTCTTTTATTTGGTCGTTG-3'
P7	JK_inv_pBAD_NcoI-NheI_R	5'-GCCATGGCGCTAGCCTTCGCAACGTTCAAATC-3'
P8	JK_inv_pBAD_NcoI-NheI_F	5'-GGCTAGCGCCATGGCATAAACTGCCAGGCATC-3'
P9	JK_NheI_cps3FHI_F	5'-ATC GCTAGC ATGATGAGACGTAGGGGAG-3'
P10	JK_NcoI_cps3FHI_R	5'-TAG CCATGG TTAACGACGTTGTAGCCG-3'
P11	JK_lac_cps3_F	5'-CGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAA ATGACAAAAATTGCATGTG-3'

Table 9 Continued

Primer. No.	**Primer Name**	**Sequence**
P12	JK_lac_cps3_R	5'-CCACACAACATACGAGCCGGAAGCATAAAGTGTAAA CATAATTGGTAACGAATCAGAC-3'
P13	JK_RT_Q_F	5'-GTAAGCACGAACTCAGCC-3'
P14	JK_RT_Q_R	5'-GGCAATCTCTGCATCTTG-3'
P15	JK_RT_*cps3A*_F	5'-GCATGTGTGATTGTCACTTAC-3'
P16	JK_RT_*cps3A*_R	5'-CACATAGTGCCATTGTCG-3'
P17	JK_RT_*cps3I*_F	5'-CGCAAAAGAATTGAGTGG-3'
P18	JK_RT_*cps3I*_R	5'-CTGACATTCTTGTCGTGC-3'
P19	JK_RT_*hcaT*_F	5'-CACTGCTGACACTTCTCTTTG-3'
P20	JK_RT_*hcaT*_R	5'-CTGCTTTTGCCACGTATTC-3'
P21	pJAK7_Bam-Nde-Xho_For	5'-GGATCCACATATGGTCTCGAGTAG CTGTACCGTGGCGAAGAC-3'
P22	pJAK7_Bam-Nde-Xho_Rev	5'-CTACTCGAGACCATATGTGGATCC GTATTTGCTCATGGTTCTGC-3'
P23	oJAK4_XhoI_Q_Rev	5'-CTA CTCGAG CTAACGTGTGATCGCATTC-3'

Table 9 Continued

Primer. No.	**Primer Name**	**Sequence**
P24	oJAK4_BamHI_Q_For	5’-CG GGATCC CATATGTTGATTCAGGTAACAGG-3’
P25	pJAK3_RBS-mKate2_For	5’-GCTAGCCTTAAGATTAACTCACAC-3’
P26	pJAK3_lac_UnP_Rev	5’-CTAGCGTCGGGGTTTGTAC-3’
P27	env_lac_FA_For	5’-TTTACACTTTATGCTTCCGGCTCGTATGTTG ATAGCGGAAATGATGAGAGAG-3’
P28	env_FA_Rev	5’-CTACCGTTTGATCTTCCTTG-3’
P29	env_P_R'_FA_For	5’-GCTTGACTTATTGAATAAAATTGGGTAAATTTG ACTCAACGAT ATAGCGGAAATGATGAGAGAG-3’

Table 10 Calculations for the qRT-PCR analysis.

Gene Name	*cps3A*			*cps3I*			*hcaT*		
Sample name	1	2	3	1	2	3	1	2	3
Plac -0mM - tech. rep. 1	21.485	21.427	22.095	17.631	17.323	17.371	26.907	27.002	28.138
Plac -0mM - tech. rep. 2	22.106	21.423	21.808	18.795	18.096	18.137	28.807	28.069	27.982
Plac - 1mM - tech. rep. 1	21.156	21.581	21.184	17.043	16.979	18.315	28.755	28.724	29.278
Plac - 1mM - tech. rep. 2	20.602	20.524	20.819	18.199	17.914	18.449	27.248	26.917	27.342

ΔCq	*cps3A*			*cps3I*		
Sample name	1	2	3	1	2	3
Plac -0mM - tech. rep. 1	-5.422	-5.575	-6.043	-9.276	-9.679	-10.767
Plac -0mM - tech. rep. 2	-6.701	-6.646	-6.174	-10.012	-9.973	-9.845
Plac - 1mM - tech. rep. 1	-7.599	-7.143	-8.094	-11.712	-11.745	-10.963
Plac - 1mM - tech. rep. 2	-6.646	-6.393	-6.523	-9.049	-9.003	-8.893

ΔCq average	*cps3A*	*cps3I*
Sample name	*cps3A-P*$_{lac}$	*cps3I-P*$_{lac}$
Plac - 0mM	-6.09	-9.93
Plac - 1mM	-7.07	-10.23

SD	*cps3A*	*cps3I*
Sample name	*cps3A-P*$_{lac}$	*cps3I-P*$_{lac}$
Plac - 0mM	0.53	0.49
Plac - 1mM	0.67	1.39

ΔCq expression	*cps3A*	*cps3I*
Sample name	*cps3A-P*$_{lac}$	*cps3I-P*$_{lac}$
Plac - 0mM	68.29	972.35
Plac - 1mM	134.02	1198.90

Standard error	*cps3A*	*cps3I*
Sample name	*cps3A-P*$_{lac}$	*cps3I-P*$_{lac}$
Plac - 0mM	0.22	0.20
Plac - 1mM	0.28	0.57

ΔΔCq w.r.t control	*cps3A*	*cps3I*
Sample name lac	*cps3A-P*$_{lac}$	*cps3I-P*$_{lac}$
Plac - 0mM	1.00	1.00
Plac - 1mM	1.96	1.23

error	*cps3A*	*cps3I*
Sample name	*cps3A-P*$_{lac}$	*cps3I-P*$_{lac}$
Plac - 0mM	10.24	135.19
Plac - 1mM	25.56	472.94

error	*cps3A*	*cps3I*
Sample name	*cps3A-P*$_{lac}$	*cps3I-P*$_{lac}$
Plac - 0mM	0.21	0.20
Plac - 1mM	0.48	0.52

p-value	*cps3A*	*cps3I*
Sample name	*cps3A-P*$_{lac}$	*cps3I-P*$_{lac}$
Plac - 0mM	-	-
Plac - 1mM	0.02	0.63

Table 10 Continued

Gene Name	*cps3A*			*cps3I*			*hcaT*		
Sample name	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
PR'-Q - tech. rep 1	18.063	18.337	17.354	17.816	16.071	15.756	25.390	26.060	24.555
PR'-Q - tech. rep 2	18.268	18.357	17.417	16.805	16.707	16.230	25.714	25.517	23.811
PR'-Q - tech. rep 3	18.636	18.773	17.349	17.805	17.335	16.425	26.151	25.835	24.382
PR'+Q - tech. rep 1	16.235	17.237	17.542	15.433	15.828	15.775	25.571	25.846	25.656
PR'+Q - tech. rep 2	16.813	17.573	17.720	15.934	16.537	16.333	25.242	25.228	25.475
PR'+Q - tech. rep 3	16.792	17.405	17.425	16.068	17.254	17.575	25.643	25.931	25.779

ΔCq	*cps3A*			*cps3I*		
Sample name	**1**	**2**	**3**	**1**	**2**	**3**
PR'-Q - tech. rep 1	-7.327	-7.723	-7.201	-7.574	-9.989	-8.799
PR'-Q - tech. rep 2	-7.446	-7.160	-6.394	-8.909	-8.810	-7.581
PR'-Q - tech. rep 3	-7.515	-7.062	-7.033	-8.346	-8.500	-7.957
PR'+Q - tech. rep 1	-9.336	-8.609	-8.114	-10.139	-10.018	-9.881
PR'+Q - tech. rep 2	-8.429	-7.655	-7.755	-9.308	-8.691	-9.142
PR'+Q - tech. rep 3	-8.851	-8.526	-8.354	-9.575	-8.677	-8.204

ΔCq average	*cps3A*	*cps3I*
Sample name	*cps3A-P*$_{R'}$	*cps3I-P*$_{R'}$
PR' - Q	-7.21	-8.50
PR' + Q	-8.40	-9.29

SD	*cps3A*	*cps3I*
Sample name	*cps3A-P*$_{R'}$	*cps3I-P*$_{R'}$
PR' - Q	0.38	0.76
PR' + Q	0.52	0.67

ΔCq expression	*cps3A*	*cps3I*
Sample name	*cps3A-P*$_{R'}$	*cps3I-P*$_{R'}$
PR' - Q	147.73	361.06
PR' + Q	338.54	627.15

Standard error	*cps3A*	*cps3I*
Sample name	*cps3A-P*$_{R'}$	*cps3I-P*$_{R'}$
PR' - Q	0.13	0.25
PR' + Q	0.17	0.22

ΔΔCq w.r.t control	*cps3A*	*cps3I*
Sample name	*cps3A-P*$_{R'}$	*cps3I-P*$_{R'}$
PR' - Q	1.00	1.00
PR' + Q	2.29	1.74

error	*cps3A*	*cps3I*
Sample name	*cps3A-P*$_{R'}$	*cps3I-P*$_{R'}$
PR' - Q	12.92	63.19
PR' + Q	40.98	97.42

error	*cps3A*	*cps3I*
Sample name	*cps3A-P*$_{R'}$	*cps3I-P*$_{R'}$
PR' - Q	0.12	0.25
PR' + Q	0.34	0.41

p-value	*cps3A*	*cps3I*
Sample name	*cps3A-P*$_{R'}$	*cps3I-P*$_{R'}$
PR' - Q	-	-
PR' + Q	0.00006	0.032

Table 11 Data for the fluorescence assay normalized to OD_{600nm} .

Average of fluorescence/absorbance of all biological and technical replicates										
Time	**attP - Q**		**attP + Q**		**Plac - Q**		**P_R'- Q**		**P_R' + Q**	
mins	**Avg**	**SD**	**Avg**	**SD**	**Avg**	**SD**	**Avg**	**SD**	**Avg**	**SD**
600	153	57	155	81	839	404	1	13	4381	1175

Table 12 Statistical calculations for the fluorescence assay .

P-values of fluorescence/O.D. w.r.t time of P_R'+Q strain compared to		
Time	**Plac - Q**	**P_R'-Q**
600	1.2E-06	3.0E-07

APPENDIX B

IDENTIFYING GENES INVOLVED IN AN UNEXPECTED GROWTH PATTERN IN *ESCHERICHIA COLI*

Overview

Multi-cellular behavior is observed in all domains of life. Like higher eukaryotes, bacteria form complex patterns of growth which is predicted to enhance cell-cell communication, protection from environment and efficient utilization of resources. Recently, the Lockless lab observed that a mutant *Escherichia coli* BW25113 strain transformed the colonies from non-motile to motile cells that form a distinct bull's eye pattern when grown in nutrient rich motility media. This work reports results from a Tn5 mutagenesis screen, where 32,004 mutant cells were screened, and 20 isolates were identified that lead to bull's eye phenotype.

Introduction

Multi-cellular eukaryotes have tissues composed of cells that work to complete a task efficiently. Bacteria appear in the evolutionary tree about a billion years before the eukaryotes. It is known that single celled bacteria switch into organized arrangements in time and space to carry out activities like cell-cell communication, survival, and nutrient transport [184,185]. There are many examples of bacteria forming a multicellular pattern on surfaces such as *Proteus mirabilis* swarming on semi-solid agar surface creating concentric patterns of growth [186-188] or *Bacillus subtilis* forming a variety of patterns including wrinkled branch spreading [189-191] . *Escherichia coli* (*E. coli*) is a widely used model for its simple structure and ease to perform genetic manipulations to study a pathway. It has a genome of ~4.6 Mbp, encoding ~4,300 genes, and peritrichously arranged flagella on its cell surface for locomotion. The orientation of flagellar rotation directs cells to follow along two trajectories. Flagella either rotate in counterclockwise (CCW) direction pushing them forward in a constant direction, or a clockwise (CW) direction causing them to tumble and change their direction [192].

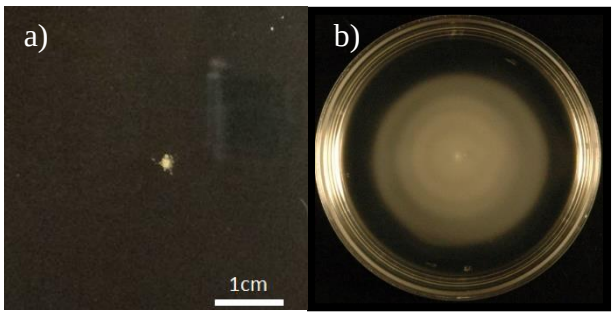


Figure 18 Motility plates stabbed with wild type (a) and mutant *E. coli* (b).

Flagellar expression is controlled by a heterodimer complex FlhDC, a transcriptional regulator of more than 50 flagellar genes. Flagella synthesis is expensive and are only expressed when a necessity is sensed. Strains of *E. coli*, such as BW25113[127] , are non-motile on due to suppressed FlhDC expression [193]. However, the Lockless lab discovered that the cells of this strain form a remarkable phenotype when the only known potassium selective channel was deleted (Figure 18)[194]. The BW25113Δ*kch* cells were found to spread on soft agar in concentric rings of growth, referred to as bull's eye pattern. The bull's eye pattern consists of successive bands of dense growth, and a region of lower density cells referred to as inter-bands (Figure 17). Upon restoring the wild type *kch* gene in Δ*kch* background, the wild type non-motile phenotype was not observed (unpublished work). This suggested that there could be other mutations that could be a cause of the bull's eye banding phenotype.

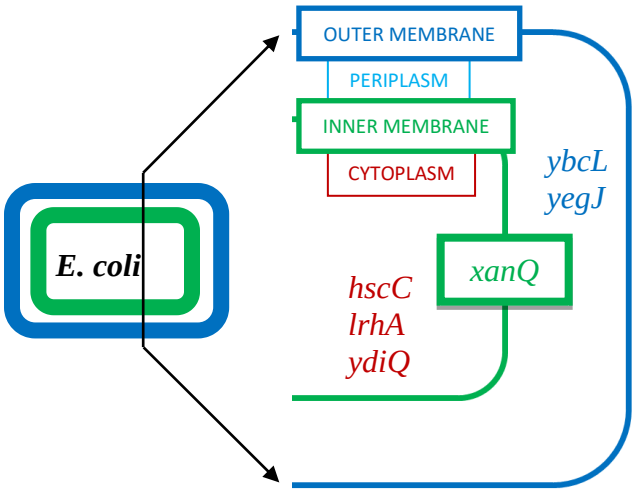


Figure 19 Localization of identified genes which form bull's eye phenotype on deletion.[3]

To identify genes that cause the banding phenotype, a forward genetics Tn5 mutagenesis screen was used, which aided in identifying 20 isolates that gave bull's eye pattern on motility media. Of these, six genes - *lrhA*, *ydiQ*, *yegJ*, *ybcL*, *xanQ* and *hscC* (Figure 19), were tested and confirmed for formation of bull's eye phenotype using strains with deletions in these genes in the Keio collection. LysR Homologue A (LrhA) belongs to a *lysR* family of transcription factor and is shown to negatively regulate *flhDC* (master regulator of motility expression genes) by binding to the promoter of *flhDC* [195]. It also indirectly affects *rpoS* translation (σ38; stress response gene) by suppressing the expression of small RNA *rprA*, a known regulator of RpoS [196], and promotes degradation of RpoS by stimulating expression of *rssB* (*sprE*)[197]. LrhA negatively regulates the synthesis of type I fimbriae by binding to the promoter of *fimE*, a recombinase responsible for switching off expression of fimbriae [198]. Thus, LrhA appears to be a potential candidate to extend this study and identify the cause of pattern formation due to its diverse role as a regulator. The following work describes the forward genetics screen carried out to identify knockouts that could lead to a bull's eye pattern phenotype in *E. coli* BW25113.

Methods

Tn5 mutagenesis to identify genes involved in bull's eye phenotype formation

Approach in brief

Since the bull's eye phenotype was observed on a plate with motility media, the approach was to perform random mutagenesis and screen a library of mutants for formation of bull's eye phenotype. The screening was performed in two stages. First, a primary screen was performed on larger plates where the motility phenotype was observed for a short incubation time. Then, the positives from the primary screen were scrapped and using a secondary screen, on smaller motility plates, individual phenotype was observed for longer incubation. Mutants from the secondary screen were isolated and the mutations responsible for the observed phenotype were identified using inverse PCR strategy.

Transposon for screening

Tn5 transposon mutagenesis approach was used to randomly disrupt genes in *Escherichia coli* BW25113. For this purpose, a plasmid carrying mini-transposon, pBAM-1 [199] was used. pBAM-1 was opted for transposon mutagenesis because of multiple significant features. pBAM-1 carries IS50 insertion element, which is a non-replicating transposon. The plasmid carries a R6K origin of replication, which permits replication of the plasmid in the host cell only in presence of a pi initiator protein, encoded by *pir* gene. Additionally, the plasmid carries the Kanamycin marker flanked by the transposase recognition terminal ends in trans-arrangement with the transposase

encoding *tnpA* gene. This feature is helpful in performing such screens, where BW25113 is *pir*$^{-}$ and thus transposase is expressed, hopping the kanamycin marker from the plasmid and inserting in the genome without any possibility of re-arrangement due to lack of expressing transposase again from the plasmid. The IS50 element in pBAM-1 randomly integrates the kanamycin marker into G/C rich sites and duplicates 9-bp nucleotide of the sites where it has integrated [200].

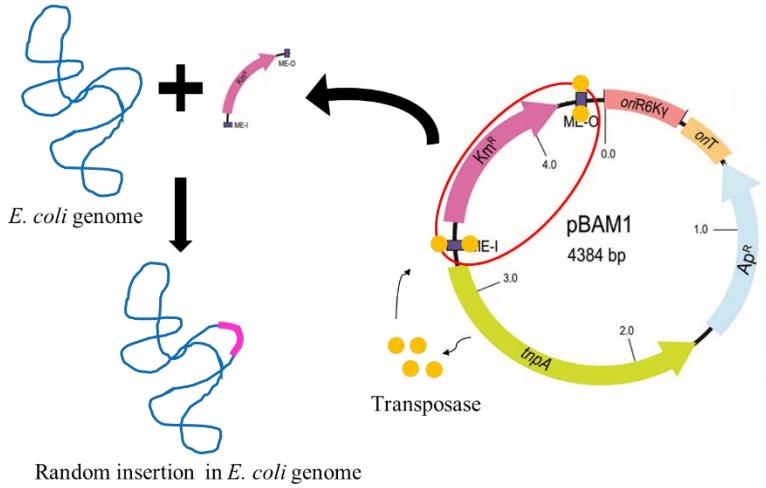
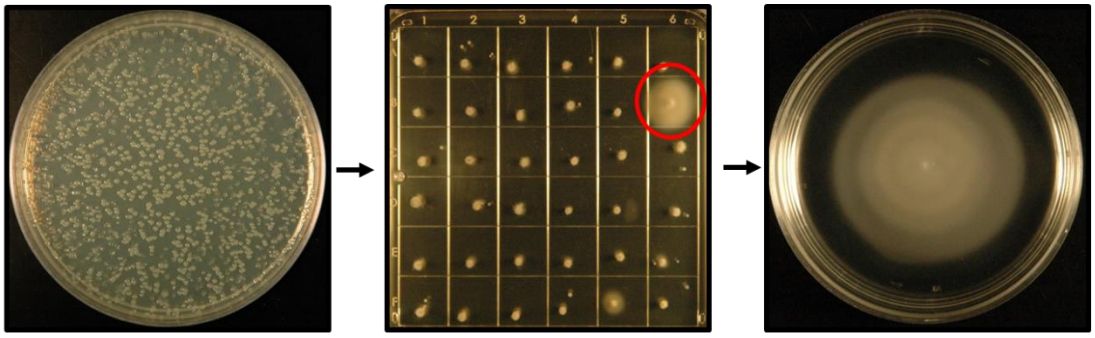


Figure 20 Schematic of pBAM1 use to disrupt genes in *E. coli* using Tn5 transposon mutagenesis.

Motility media composition for screening

Bull's eye phenotype formation by *Escherichia coli* was tested by stabbing cells on motility plates which were composed of low agar concentration (0.25 – 0.27% w/v). In order to make the screening faster and efficient, the screen was divided into two phases (Figure 21). In the first phase, 100 x 100 x 15 mm square petri dishes (referred to as square motility plates from here on) were used to partially screen for isolates which show spreading. In the second phase, the selected isolates from the first phase were tested individually for forming bull's eye phenotype on small petri plates (60 x 15 mm).



Total number of isolates screened = **32,004**
Total number of isolates that spread in primary screening = **45 (0.14%)**
Total number of isolates that show bull's eye pattern = **20 (0.06%)**
Total number of isolates that did not spread = **31,959 (99.86%)**

Figure 21 Schematic of Tn5 transposon mutagenesis to identify mutants that form bull's eye phenotype on motility plates.

The initial work was to design the media composition for the first phase, so that a multiple of Tn5 mutagenized isolates could be stabbed on square petri plates. A set of modifications were made to the ingredients, such as varying tryptone, yeast extract, NaCl/KCl, agar concentration, in order to obtain a concise but distinct bull's eye phenotype.

Methodology to screening from bull's eye phenotype forming Tn5 mutants

pBAM-1 was a gift from Victor de Lorenzo (Addgene plasmid # 600487). The plasmid pBAM-1 was extracted from cells which were grown overnight in LB + kanamycin broth and transformed into BW25141 using electroporation. The vector pBAM-1 carries the R6K origin of replication (ori). Hence, *E. coli* BW25141 was used, for propagation of pBAM1, since the strain carries gene that encode for pi protein required for replication of vector with R6K ori. The kanamycin resistant isolates were cultured overnight, and presence of plasmid was confirmed by extracting the plasmid and digesting with various restriction sites. The expected size of the plasmid using various restriction enzymes confirmed the correctness of pBAM-1. The cell carrying pBAM-1 was grown in a large batch of media (250 ml) for 16 hours. Using maxiprep kit from Qiagen, the plasmid was harvested and stored at -20^{o}C for further transformation purposes.

A general outline of the protocol followed to screen for isolates is as follows:

Day 1 - Performing electroporation of BW25113 strains with pBAM-1:

~500 ng of purified pBAM-1 was used to transform electrocompetent BW25113 cells. The electroporated cells were incubated in SOC media for 1 hour at 37^{o}C. Post

incubation, the culture was plated on LB + kanamycin plates in 10^{-6} and 10^{-7} fold serial dilutions. The transformation efficiency was calculated to be ~4 x10^{9} CFU/μg of DNA.

Day 2 - Screening colonies on motility plates with first pass:

The transformed isolates were stabbed on square plates (100 x 100 x 15mm petri plates). Each square motility plates could hold 36 isolates. [The composition of motility agar plates is: 3% tryptone + 1.5% yeast = ~170mM KCl in 0.27% agar. The agar depth in the plates should be ~0.5 cm, approx. media volume is 40ml]. The stabbed cells were incubated at 37ºC for 9 hours. After 9 hours, images of the square motility plates were taken and the isolates which showed spreading, were streaked for isolation on LB + Kanamycin plates.

Day 3 - Confirming bull's pattern in spreading colony on baby plates:

Isolates of transformed mutants, which showed spreading on square motility plates, were stabbed on small motility plates (60 x 15 mm petri plates) to confirm the pattern formation [the composition of motility agar plates is: 1% tryptone + 0.5% yeast = ~170mM KCl in 0.25% agar. The media volume for these plates is 7.5 ml]. The plates were incubated at 37ºC and imaged after 12 hours. The isolates which formed bull's eye phenotype after growing on small baby plates were cultured overnight in 5 mls of LB + kanamycin broth.

Day 4 – Extraction of genome of isolates that were confirmed to create bull's eye pattern:

An aliquot of the culture was stocked for future and the remaining was used to extract genomic DNA using the DNeasy Blood and Tissue kit from Qiagen. Approximately 1μg

of genomic DNA was digested overnight using *Bss HII* (~2474 restriction sites on BW25113 genome) at 50°C.

Day 5 – Identification of gene knocked out by Tn5 transposon:

The digested genomic preparation was incubated at 65°C for 10 minutes to inactivate restriction enzyme. Approximately 0.3 µg of digested genomic DNA sample was ligated using T4 DNA ligase for 2 hours at room temperature (total 40 µl volume) and 1µl of the ligated mix was used to PCR amplify the transposon using primers which anneal to the known sequence on kanamycin resistance gene, which eventually amplify a fragment with the unknown gene sequence flanked by partial fragments of kanamycin resistance gene (Figure 22). The PCR amplified fragment was purified and sequenced from ETON Bioscience Inc.

The above procedure was used to screen 32,004 Tn5 mutagenized isolates. The screen resulted in 20 isolates that formed bull's eye phenotype. Of these 20 isolates, 12 unique genes were identified (9 isolates carried disruption in the same gene *lrhA*) that formed the pattern.

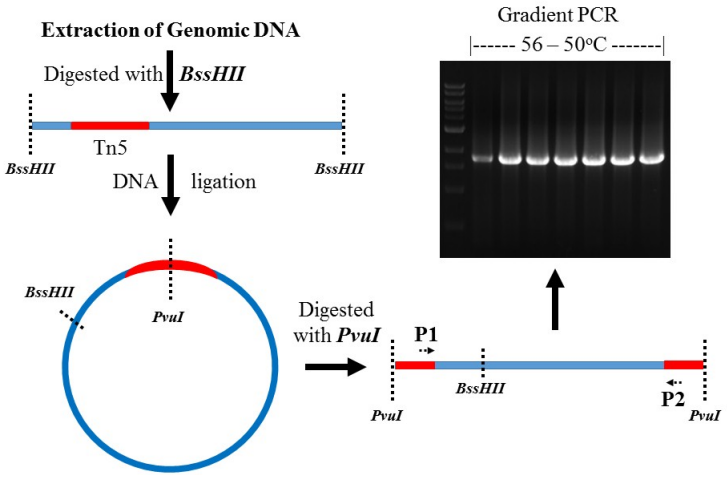


Figure 22 Schematic demonstrating steps for identification of gene disrupted by Tn5 transposon mutagenesis.

Results

Genes identified in the screening

Upon screening 32,004 Tn5 mutagenized isolates, 20 isolates formed bull's eye phenotype. Using molecular techniques, 12 unique genes (Figure 23) were identified from the 20 isolates, which upon disruption resulted in the pattern formation. The illustrations in Table 13 show where the disruption occurred on the genome. Isolate names refer to the location of the isolate on the square petri plates, for example, 121D4 strains suggests that this strain was isolated from the square petri plate 121 and is stabbed in the column D and row 4.

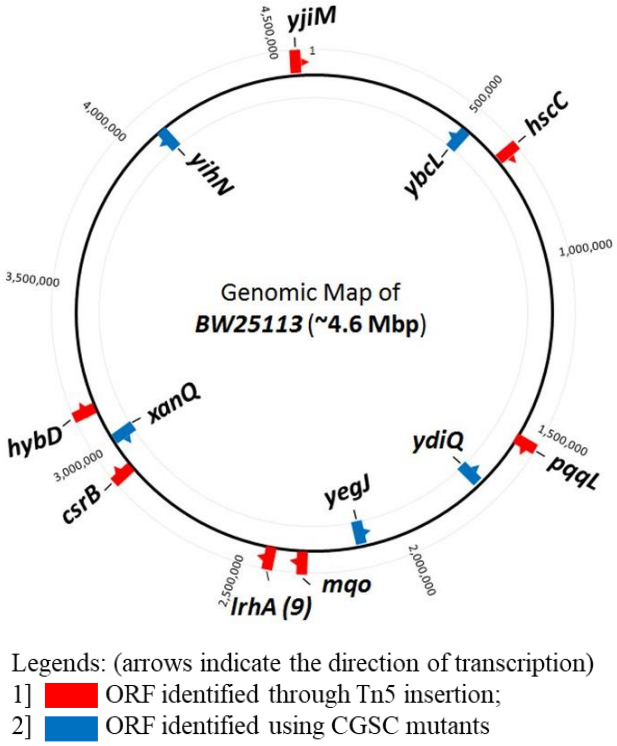


Figure 23 Mapping the location of genes disrupted using Tn5 transposon mutagenesis.

Table 13 Illustration of gene disruption location on the genome for strains that showed bulls eye pattern formation.

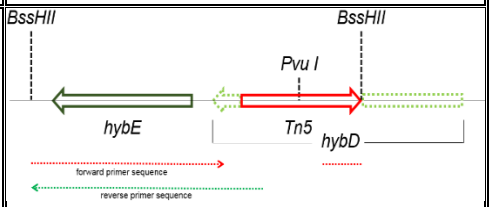
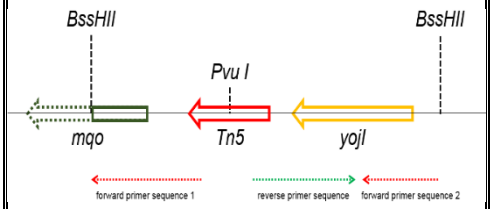
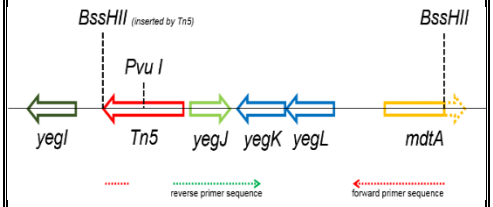
Isolate name	**Site of insertion**	**Description of gene**	**Schematic**
4E4 (*hybD)*	3134213	predicted hydrogenase 2 maturation protease	
5F2 (*Mqo promoter? In between mqo promoter and yojI terminatio n)*	2300417 - 2300425	*Mqo* – malate quinone oxidoreductase; *yojI* – putative ABC superfamily transporter	
19B2 (*yegI/J promoter? Between yegJ promoter and gene)*	2144624	No predicted role of YegJ	

Table 13 Continued

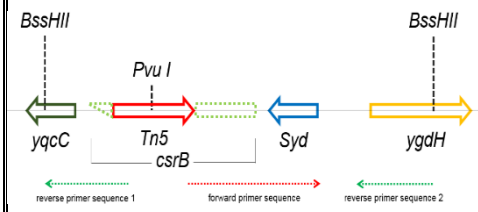
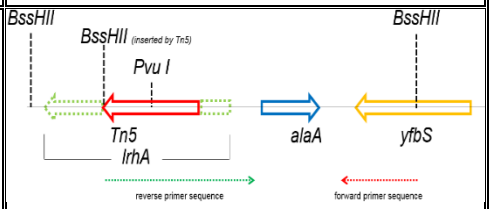
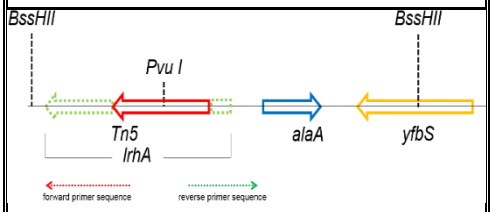
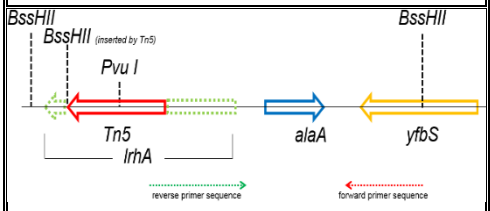
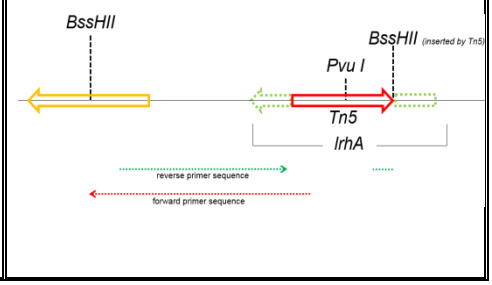
Isolate name	Site of insertion	Description of gene	Schematic
24E2 (*csrB*)	2917584 - 2917592	*csrB* Small regulatory RNA Of carbon storage regulator (csr) operon	
121D4 (*lrhA*)	2400006	Lysine Regulator Homologue A	
174F6 (*lrhA*)	2400008 - 2400016		
187E3 (*lrhA*)	2399594		
373C4 (*lrhA*)	2399294 (added *Bss HII* site)		

Table 13 Continued

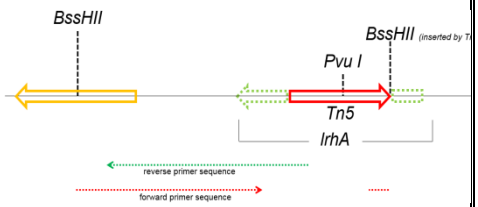
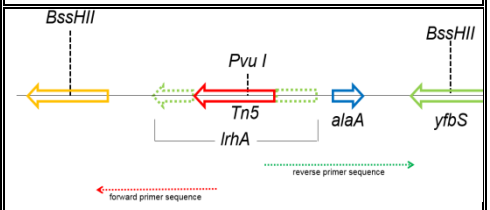
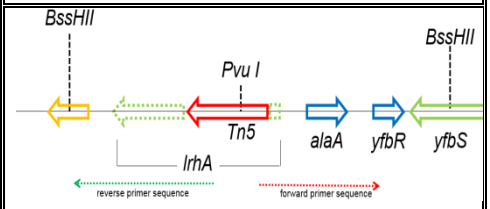
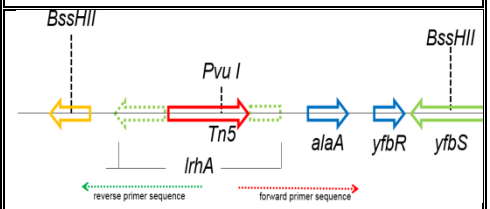
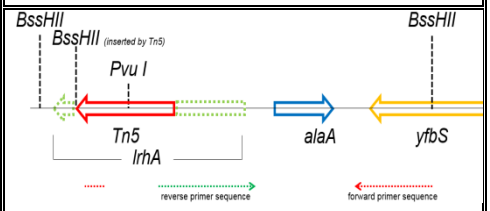
Isolate name	Site of insertion	Description of gene	Schematic
419D6 (*lrhA)*	2399608 (added *Bss HII* site)	Lysine Regulator Homologue A - Regulates transcription of gene involved in type 1 fimbriae and expression of *flhDC*. Also, LrhA negatively and partially regulates expression of *rpoS*	
463A2 (*lrhA)*	2399608		
492E2 (*lrhA)*	2399584 - 2399592		
571D5 (*lrhA)*	2400008 - 2400016		
805B6 (*lrhA)*	2399468 (added *Bss HII*)		

Table 13 Continued

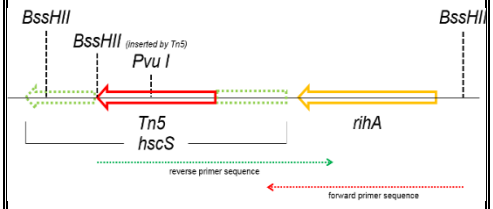
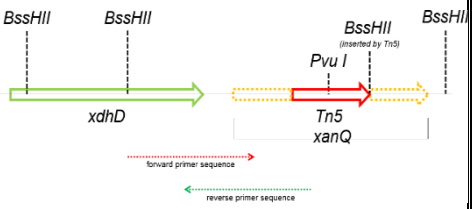
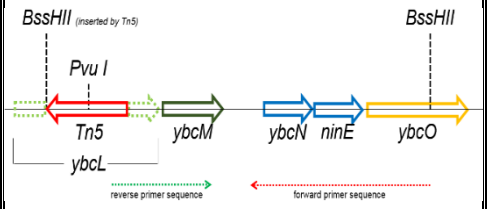
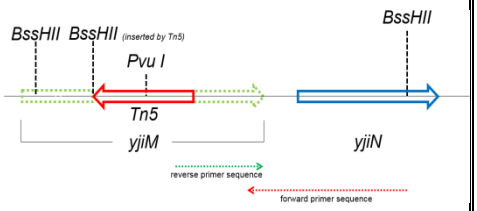
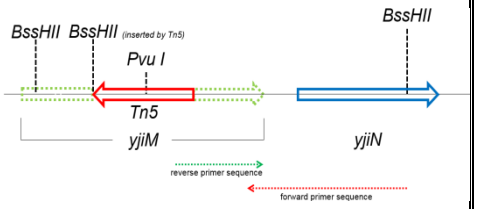
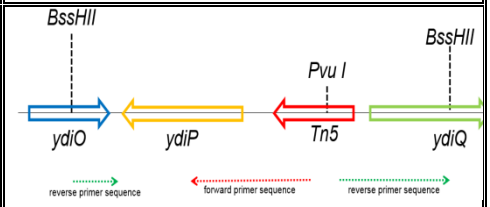
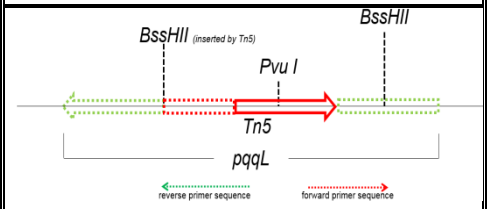
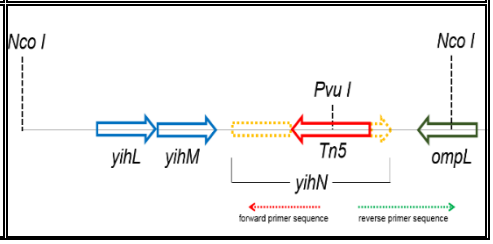
Isolate name	**Site of insertion**	**Description of gene**	**Schematic**
121F4(*hscC)*	678560	*hscC/62* – Hsp70 family chaperone binds to RpoD and inhibits transcription of it	
305A1 (*xanQ* available as *ygfO)*	3018315	Xanthine : H^+ symporter. Proton motive force dependent xanthine transporter	
323C3 (*ybcL)*	566688	periplasmic protein, YbcL from *E.coli strain* UT189 suppresses acute neutrophil migration. Has sequence similarity to rat and human Raf kinase inhibitor protein	

Table 13 Continued

Isolate name	**Site of insertion**	**Description of gene**	**Schematic**
368D3 (*yjiM)*	4555120 -4556599	Predicted to be responsible in uptake of proline rich antimicrobial peptides	
368D3 (*yjiM)*	4555120 -4556599	Predicted to be responsible in uptake of proline rich antimicrobial peptides	
749B2 (ydiQ)	1773766 - 1773774	Predicted subunit of an electron transport flavoprotein.	
18F2 (*pqqL)*	1566834	Predicted zinc peptidase	
20C4 (*yihN)*	4056597	Putative transport protein, major facilitator superfamily	

Confirming the genes identified by Tn5 mutagenesis screening using CGSC knockouts to form bull's eye phenotype

After identification of genes disrupted in 20 of the Tn5 mutants, CGSC knockouts of the genes were ordered from the Yale Stock center and tested for formation of bull's eye pattern phenotype. After arrival of the mutants from the stock center, three biological replicates from each stock were stabbed on motility media plates. The results from the CGSC knockouts are briefly summarized in Table 14.

Table 14 Confirming bull's eye phenotype using CGSC knockout strains.

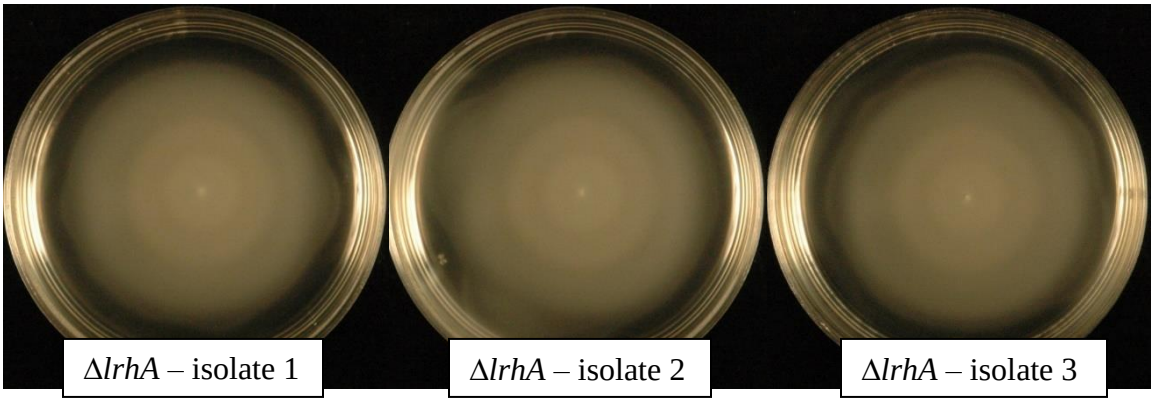
Gene disrupted	Bull's eye phenotype formation
hscC	Yes/faint bands
lrhA	Yes
xanQ/ygfO	Yes
ybcL	Yes
ydiQ	Multiple phenotype
yegJ	Yes
hybD	Spread confluently
mqo	Spread confluently
pqqL	Spread confluently
csrB (created the knockout in lab)	No
yihN	No
yjiM	No

It can be concluded from the results that there were only 6 of the 12 genes which upon disruption formed bull's eye pattern and can be replicated using the CGSC knockouts.

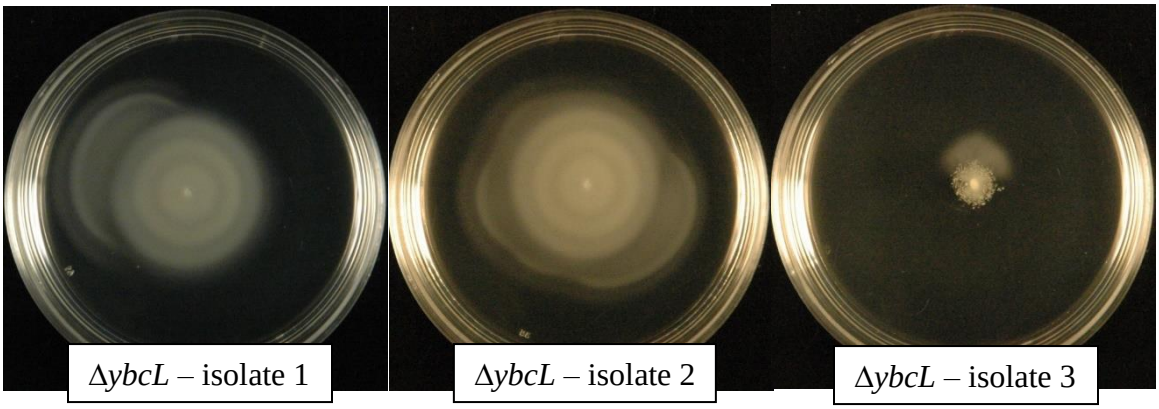
The possible reasons for other genes which did not replicate the bull's eye phenotype could be because, the genes identified through the screening may not be the only alteration on the genome and perhaps another mutation elsewhere could have impacted the formation of phenotype. This is possible by observing multiple amplification bands after an inverse PCR reaction in certain isolates. Another possibility could be that during the attempted disruptions certain spontaneous mutations could have led to mutations in other genes on the genome.

Plate images of CGSC knockouts when tested for bull's eye pattern formation

Δ*lrhA* imaged after 13 hours:-



Δ*ybcL* imaged after 13 hours:-



Δ*yegJ* imaged after 13 hours:-

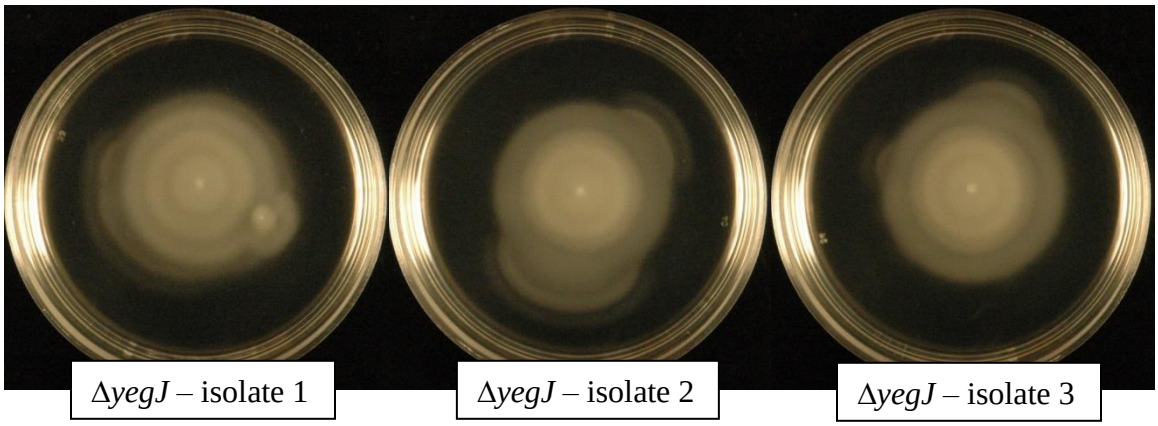
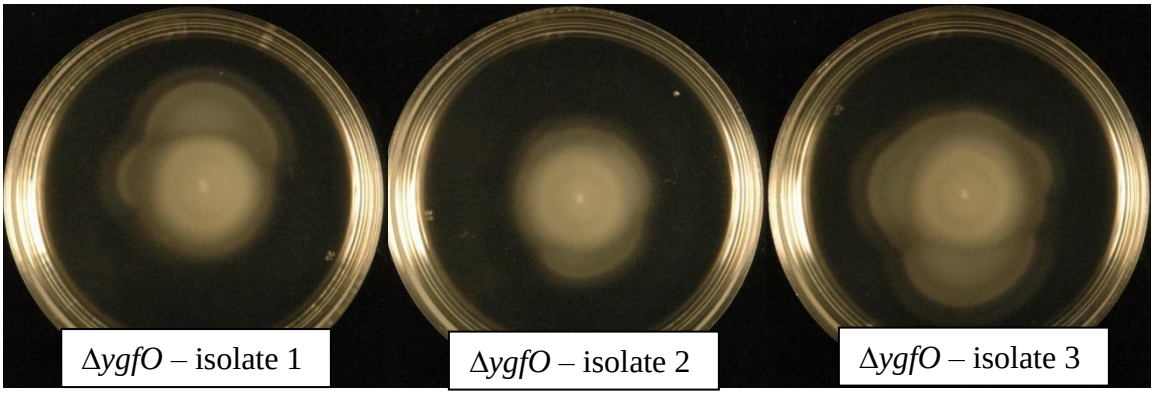
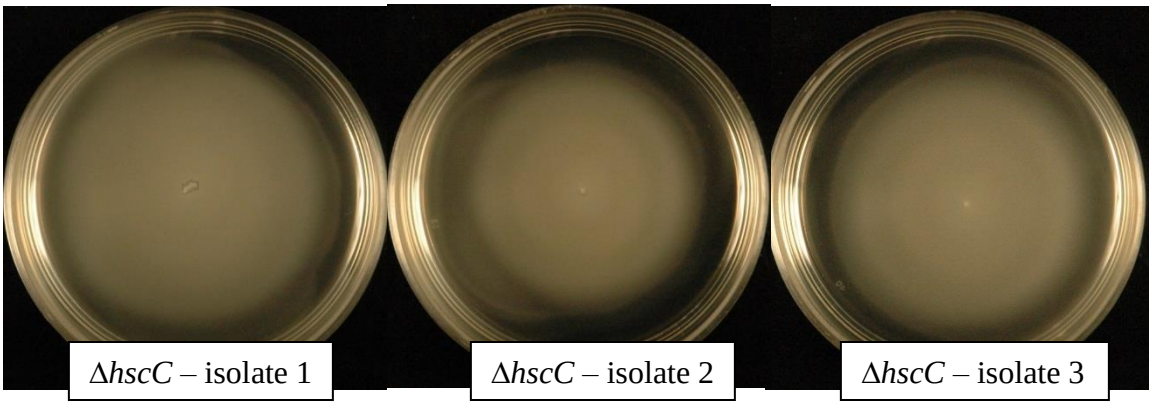


Figure 24 Image of motility plates stabbed with CGSC knockouts to test formation of bull's eye patter formation.

Δ*ygfO* imaged after 13 hours:-



Δ*hscC* imaged after 13 hours:-



Δ*hybC* imaged after 13 hours:-

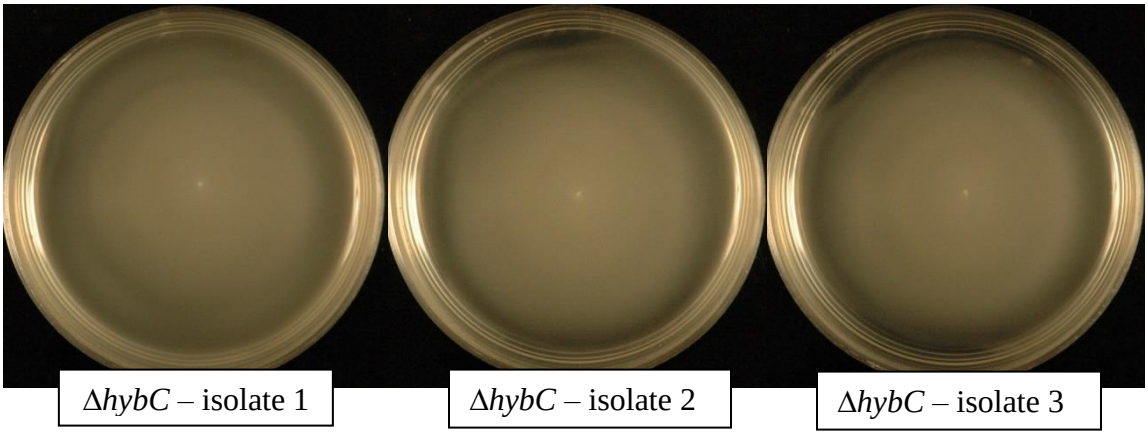
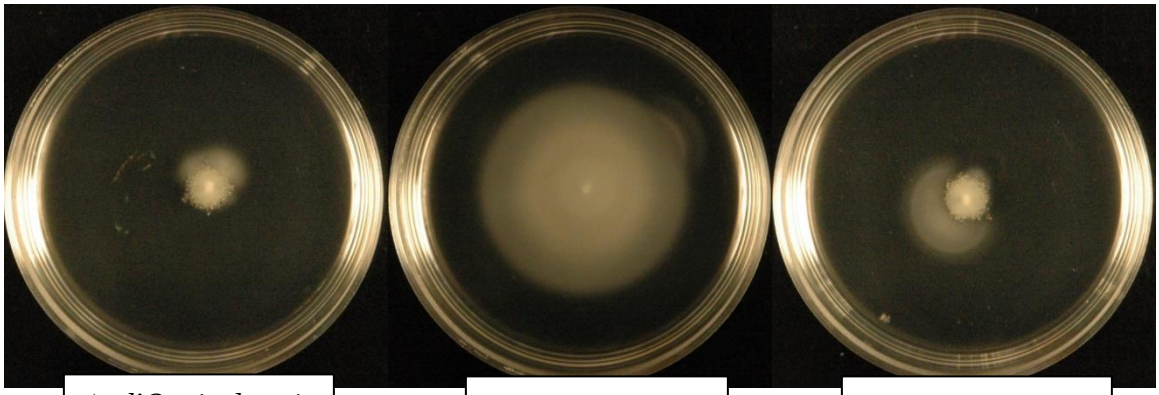


Figure 24 Continued

Δ*ydiQ* imaged after 13 hours:-



Δ*ydiQ* – isolate 1 | Δ*ydiQ* – isolate 2 | Δ*ydiQ* – isolate 3

Δ*hybD* imaged after 13 hours:-

Δ*hybD* – isolate 1 | Δ*hybD* – isolate 2 | Δ*hybD* – isolate 3

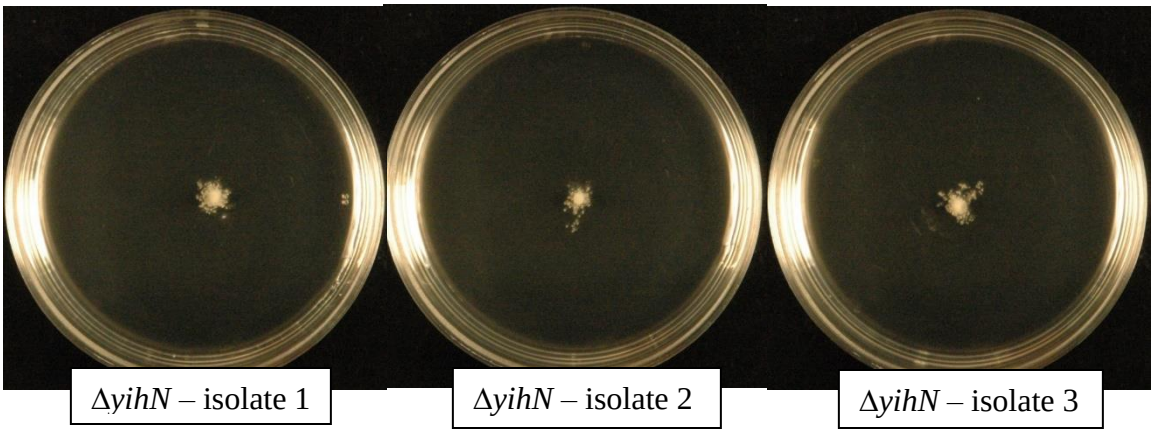
Δ*mqo* imaged after 13 hours:-

Δ*mqo* – isolate 1 | Δ*mqo* – isolate 2 | Δ*mqo* – isolate 3

Figure 24 Continued

Δ*pqqL* imaged after 13 hours:-

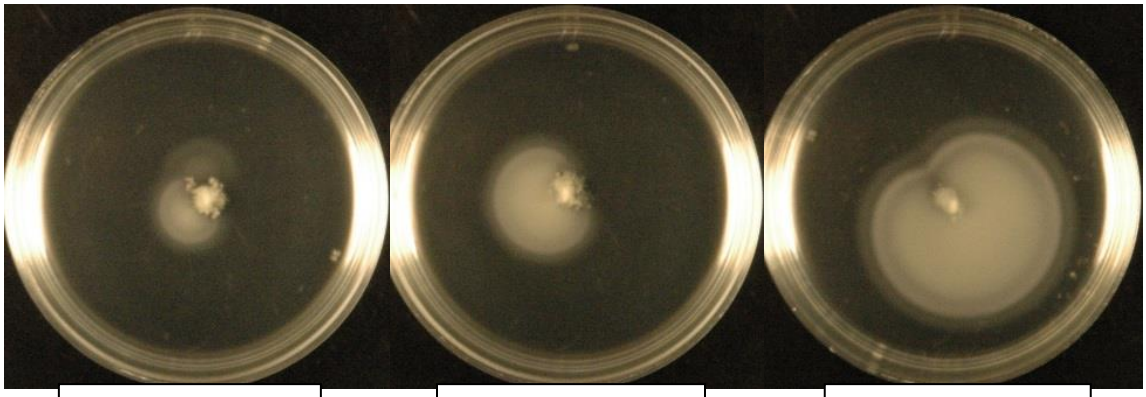
Δ*yihN* imaged after 13 hours:-



Δ*yjiM* imaged after 13 hours:-

Figure 24 Continued

Δ*csrB* imaged after 13 hours:-



Δ*csrB* – isolate 1 | Δ*csrB* – isolate 2 | Δ*csrB* – isolate 3

Figure 24 Continued

Gel images of inverse PCR and petri plate of bull's eye pattern forming 20 isolates

This section is included in order to show whether the amplification yielded in a single amplicon or multiple amplicons (which was in the case of Δ*ydiQ*) while running the inverse PCR.

4E4 – *hybD::IS(KanR)*

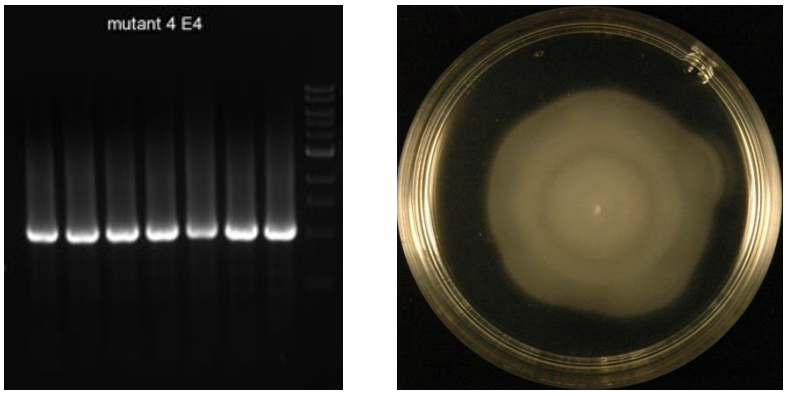
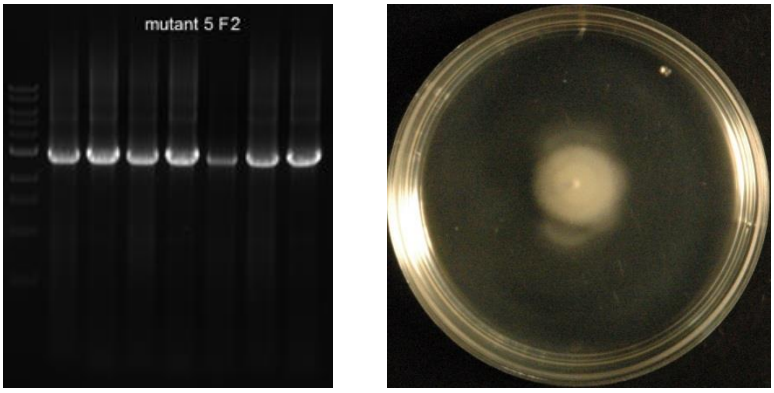
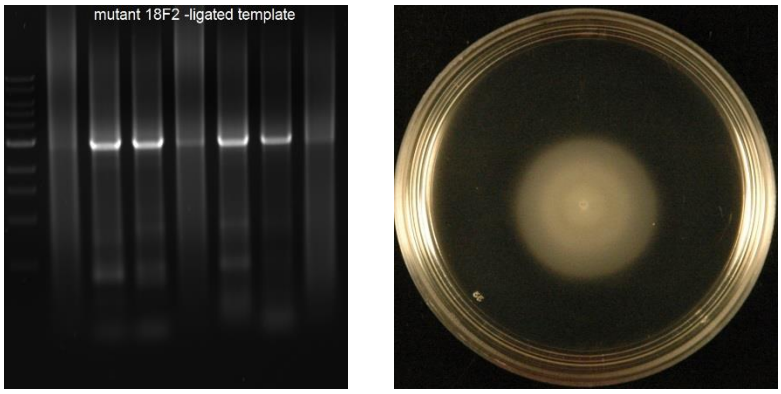


Figure 25 Gel and motility plate images of isolates from the Tn5 mutagenesis screening that formed bull's eye phenotype.

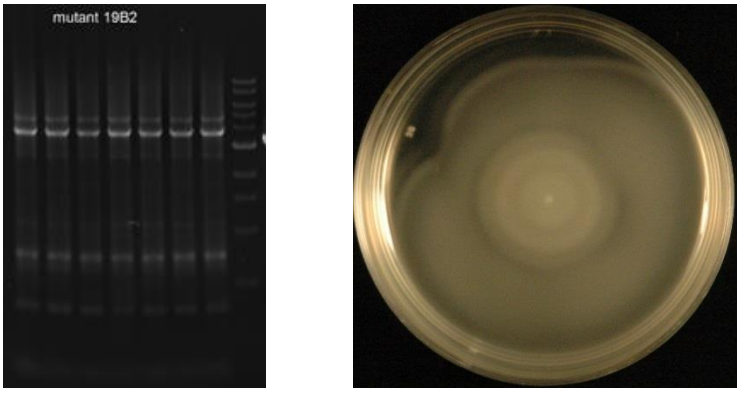
5F2 – *mqo::IS(KanR)*



18F2 – *pqqL::IS(KanR)*



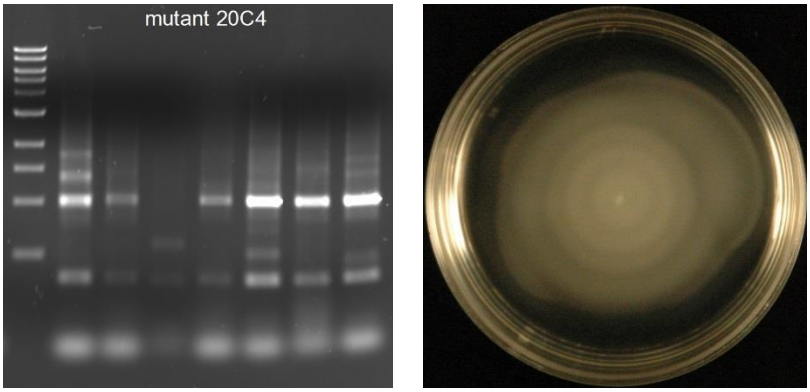
19B2 – *yegJ::IS(KanR)*



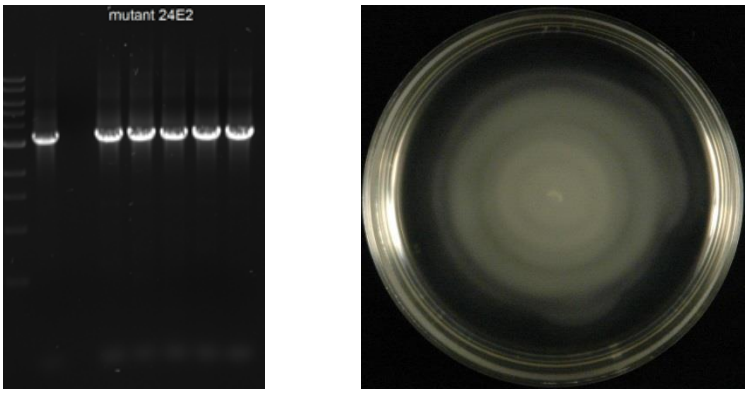
.

Figure 25 Continued

20C4 – *yihN::IS(KanR)*



24E2 – *csrB::IS(KanR)*



121D4 – *lrhA::IS(KanR)*

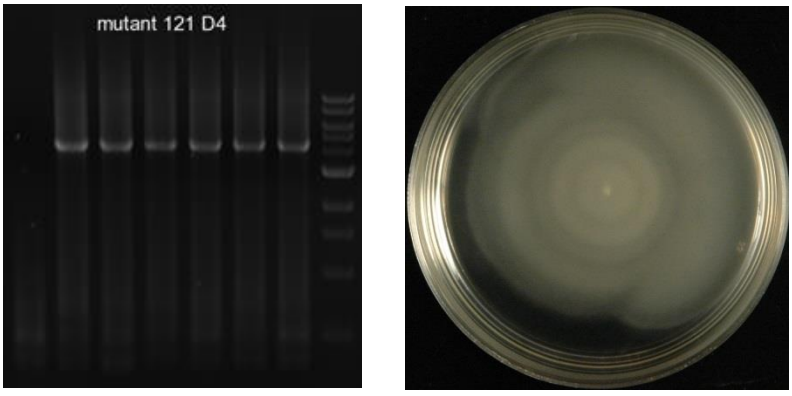
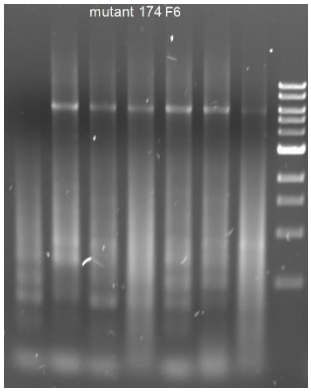
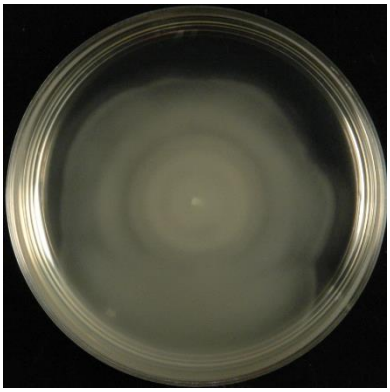


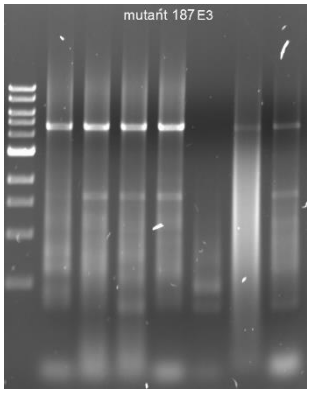
Figure 25 Continued

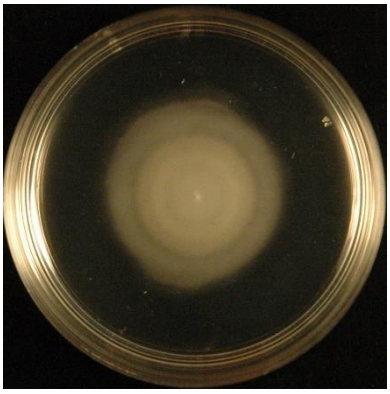
174F6 – *lrhA::IS(KanR)*



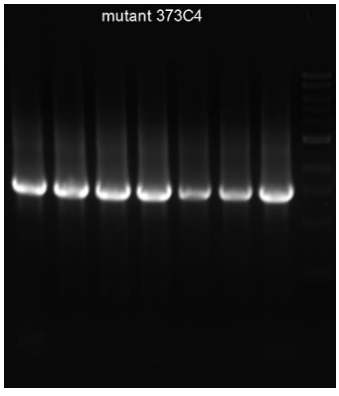


187E3 – *lrhA::IS(KanR)*





373C4 – *lrhA::IS(KanR)*



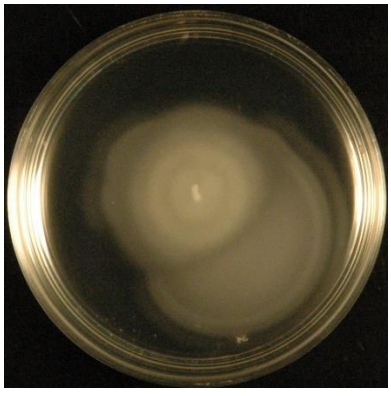
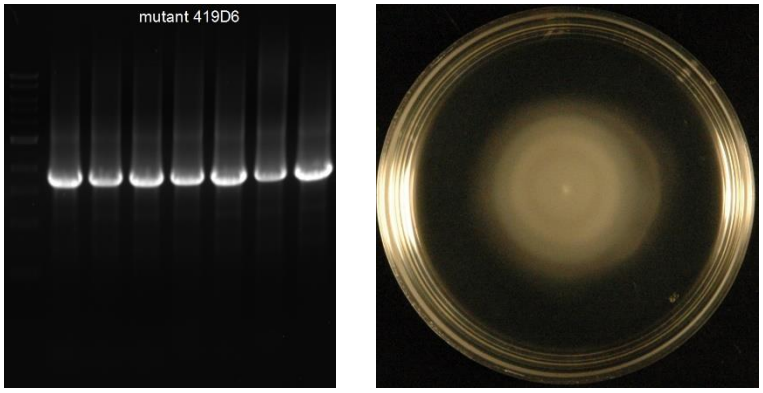
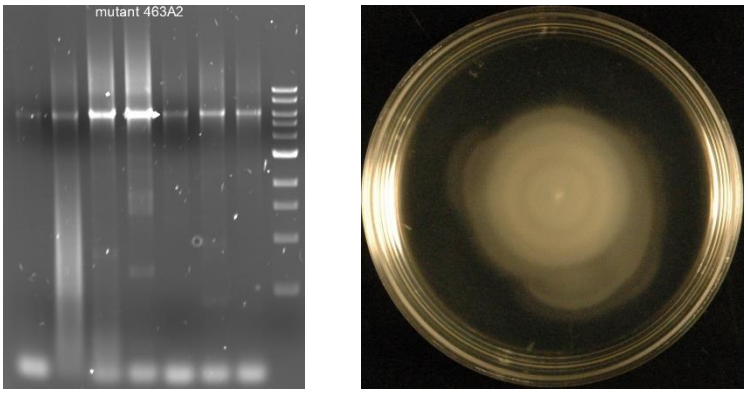


Figure 25 Continued

419D6 – *lrhA::IS(KanR)*



463A2 – *lrhA::IS(KanR)*



492E2 – *lrhA::IS(KanR)*

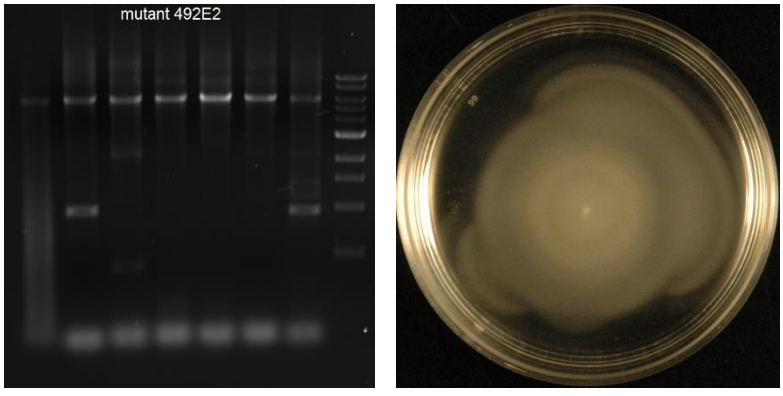
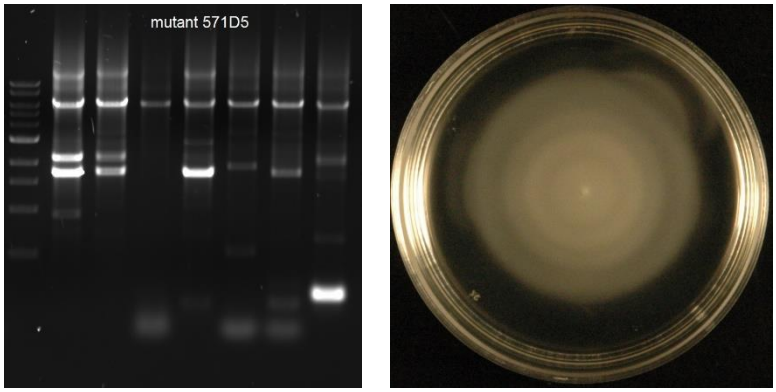
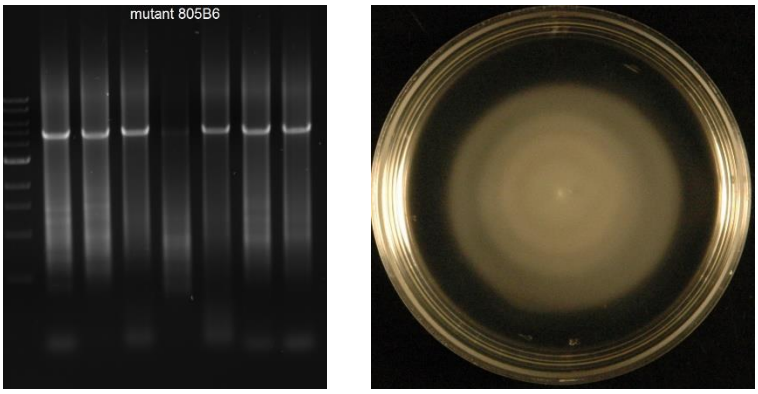


Figure 25 Continued

571D5 – *lrhA::IS(KanR)*



805B6 – *lrhA::IS(KanR)*



121F4 – *hscC/ybeW::IS(KanR)*

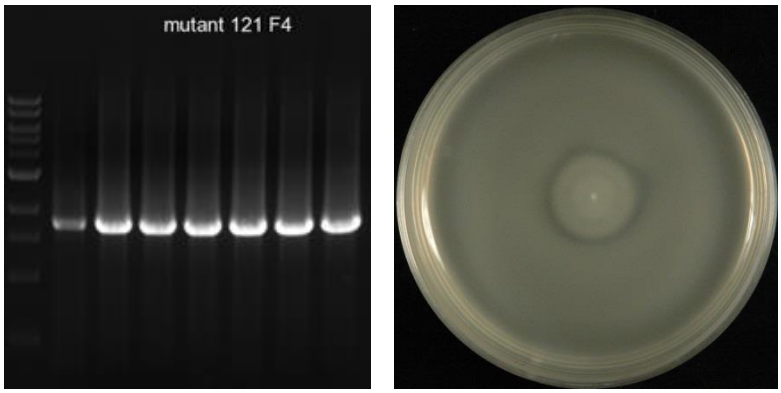
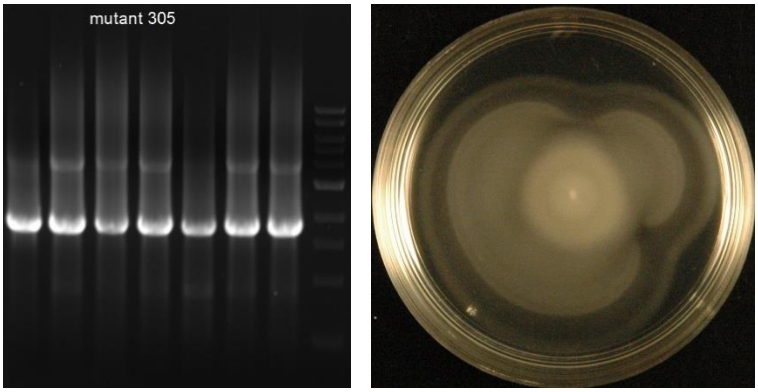
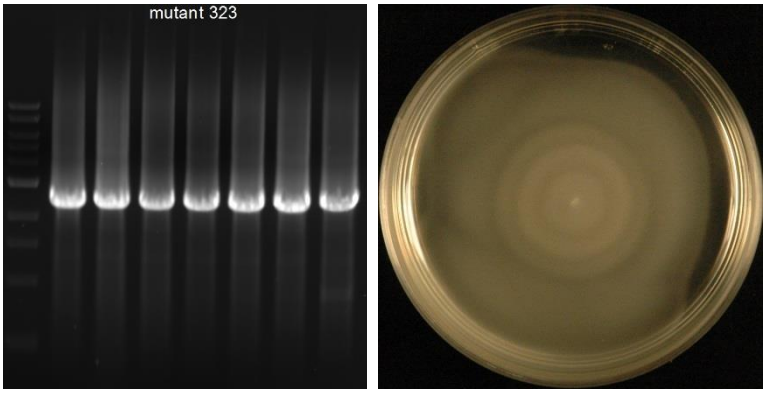


Figure 25 Continued

305A1 – Δ*xanQ/ygfO::IS(KanR)*



323C2 – *ybcL::IS(KanR)*



368D3 – *yjiM::IS(KanR)*

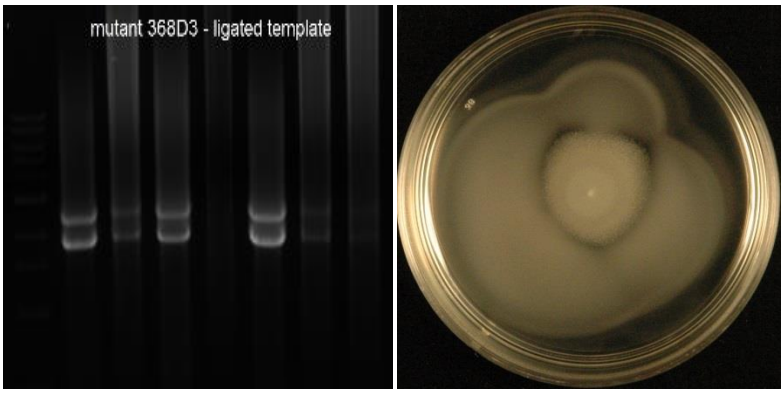


Figure 25 Continued

759B2 – *ydiQ::IS(KanR)*

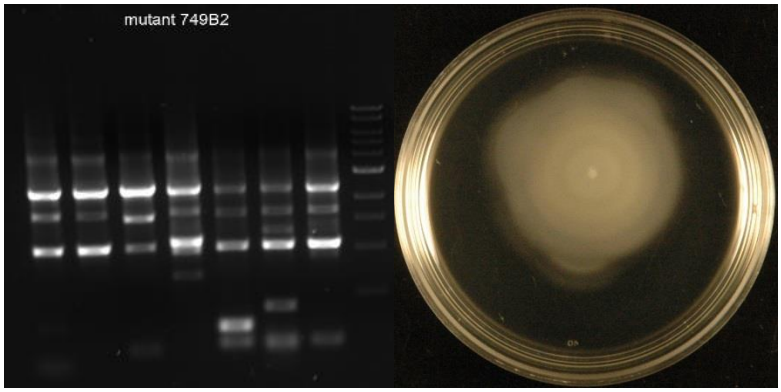


Figure 25 Continued

Variations in CGSC Δ*ydiQ* phenotypes

CGSC knockout Δ*ydiQ* showed multiple phenotypes when confirming formation of bull's eye pattern. In order to understand if the multiple phenotypes observed is due to a random chance, a series of 10 random isolates from the CGSC knockout (The *E. coli* Genetic Stock Center) stock of Δ*ydiQ* were stabbed and observed for phenotype on motility plates.

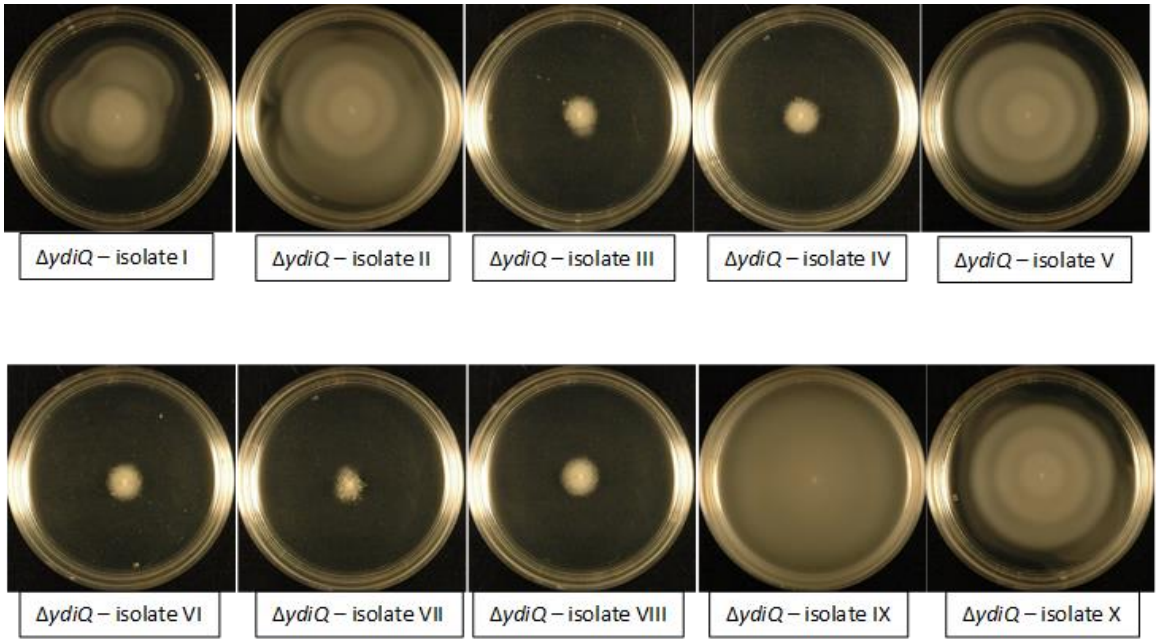


Figure 26 Motility plate images of 10 isolates of CGSC Δ*ydiQ* strain.

The isolates were deleted with the wild type *ydiQ* gene, as was confirmed by a PCR reaction (Figure 27) where the control wild type *ydiQ* gene is ~ 750 bp whereas the *ydiQ* gene disrupted with FRT-Kan-FRT cassette [127] is expected to be ~ 1300 bp.

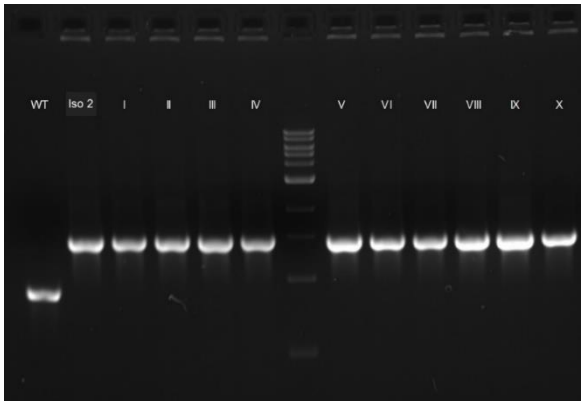


Figure 27 PCR image confirming deletion of ydiQ gene in the CGSC knockouts.

Genome sequencing results of non-motile Δ*ydiQ* and bull's eye pattern forming Δ*ydiQ*

Since isolates from the CGSC Δ*ydiQ* strains showed multiple phenotypes, a complete genome sequencing was done for the CGSC Δ*ydiQ* isolate V (bull's eye pattern forming isolate) and VI (non-motile) (Figure 26) to identify the molecular basis for this variation.

The genome sequencing results were helpful in identifying that there were two SNPs in isolate V of the CGSC Δ*ydiQ*. The first SNP was identified in *hemB* (C343598T leads to GLN171LEU mutation). Second, there was a SNP in *lrhA* (glutamine 278 STOP). There were no additional mutations found in CGSC Δ*ydiQ* isolate VI, besides the deletion in gene *ydiQ*.

Rescue experiments conducted on CGSC Δ*lrhA* and Δ*ydiQ*

Since multiple *lrhA* mutants were identified during the Tn5 transposon screening and in addition isolate V of Δ*ydiQ* also contained a SNP in *lrhA*, we decided to focus experiments based on these two mutants. First, we attempted to rescue the phenotype using the CGSC Δ*lrhA* and CGSC Δ*ydiQ* by expressing the wild-type lrhA gene from the pQE60 (Qiagen) expression vector.

The pQE60 vector was constructed by inserting two variations of the *lrhA* gene, one with the wild type *lrhA* (pSLJK13, i.e. pQE60 + $lrhA^{wt}$); and the other with SNP on *lrhA* identified and amplified from the CGCS Δ*ydiQ* isolate V. (pSLJK15, i.e. pQE60 + $lrhA^{Q278\text{-}Stop}$)

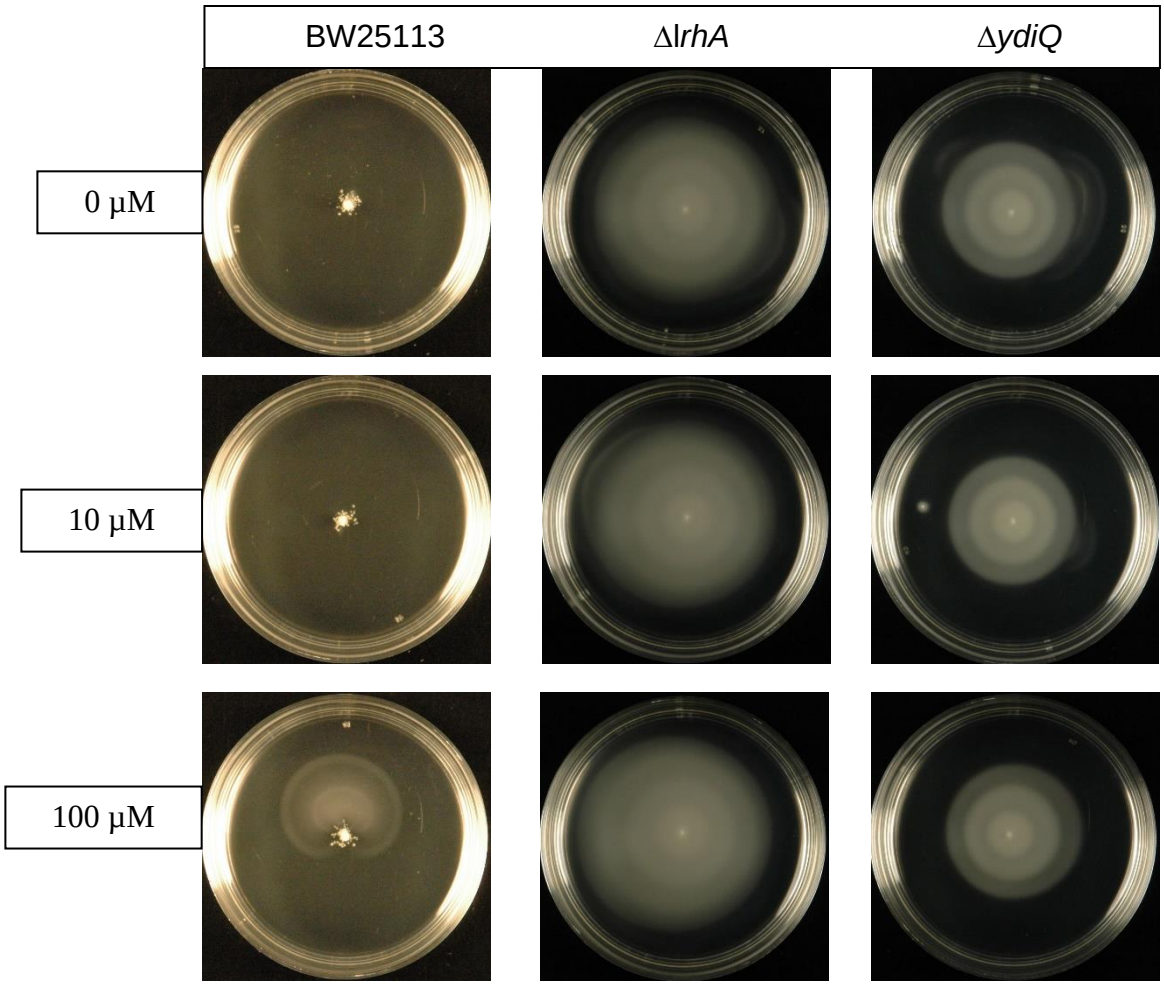


Figure 28 Motility assay when wild type *E. coli* BW25113, Δ*lrhA* and Δ*ydiQ* strains were transformed with empty vector pQE60 when induced with varying concentration of IPTG

The results from the expression studies are shown in Figure 28-30.

To test bull's eye pattern formation in wild type *E. coli* BW25113, Δ*lrhA* and Δ*ydiQ*

mutants carrying empty vector pQE60 at varying IPTG concentration.

To test bull's eye pattern formation in wild type *E. coli* BW25113, Δ*lrhA* and Δ*ydiQ* mutants carrying complete gene $lrhA^{wt}$ at varying IPTG concentration.

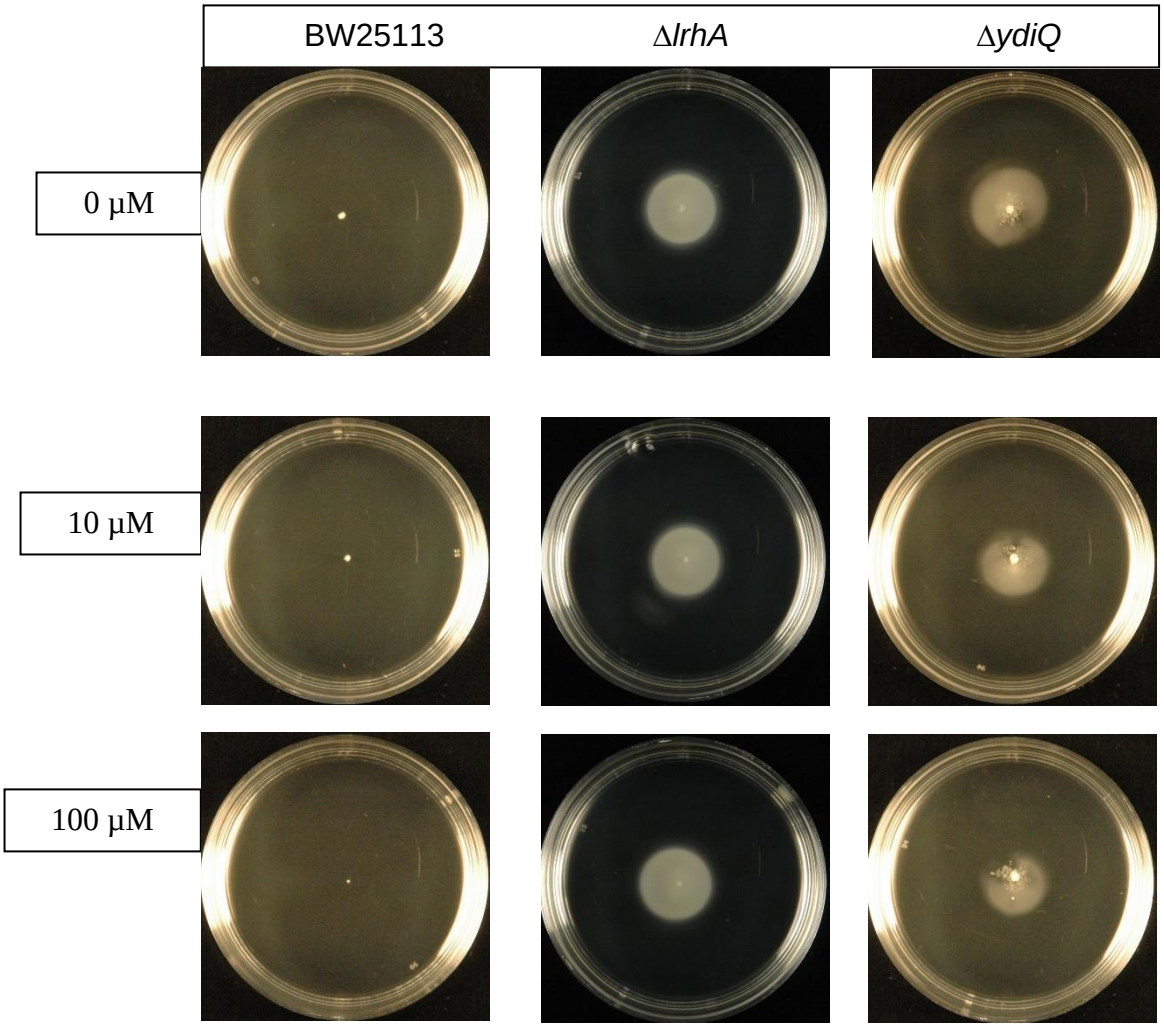


Figure 29 Motility assay when wild type *E. coli* BW25113, Δ*lrhA* and Δ*ydiQ* strains were transformed with vector pQE60 carrying wild type *lrhA* gene and induced with varying concentration of IPTG.

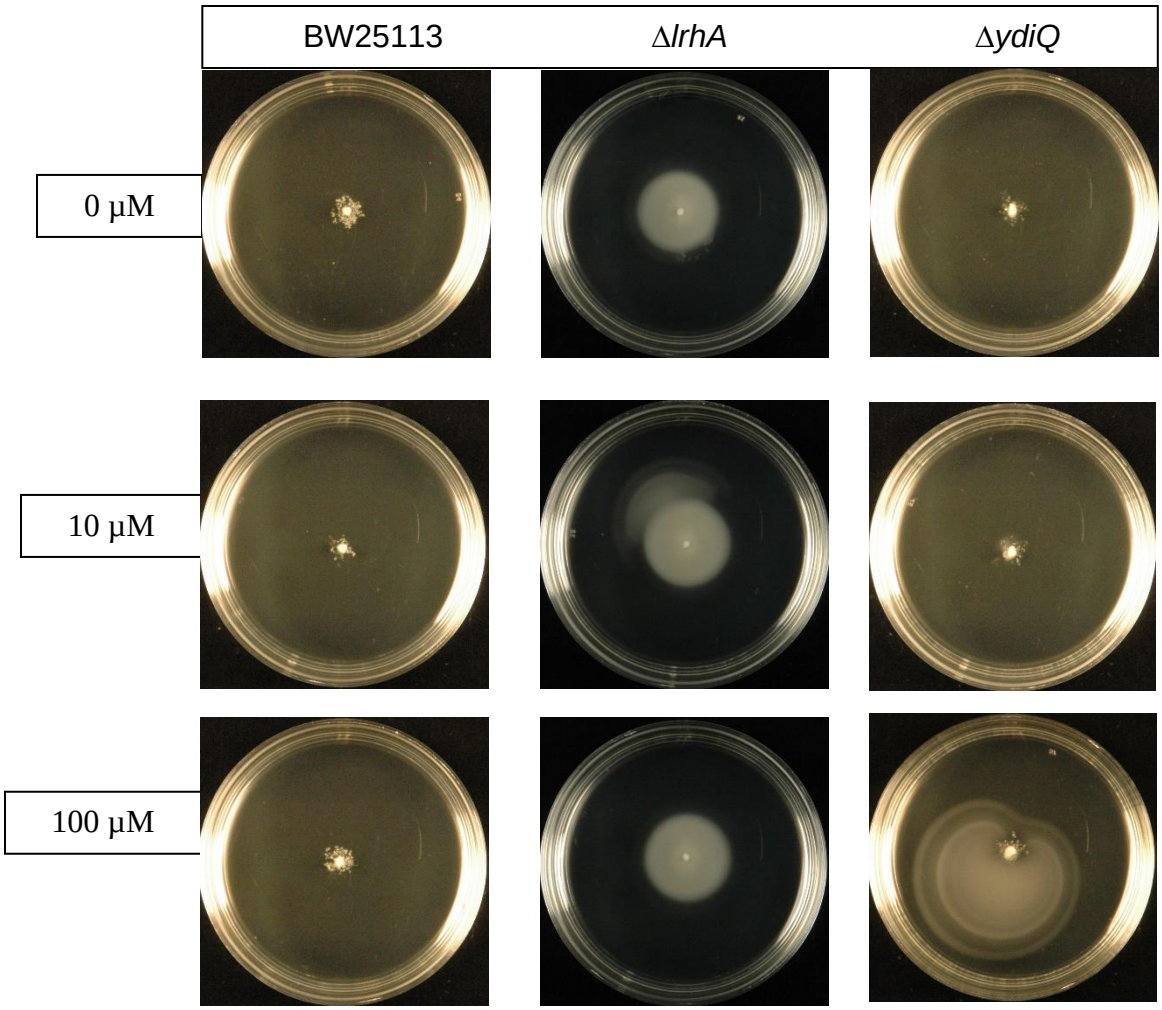


Figure 30 Motility assay when *E. coli* BW25113, Δ*lrhA* and Δ*ydiQ* strains were transformed with vector pQE60 carrying SNP (GLN 277 Stop) *lrhA* gene and induced with varying concentration of IPTG.

To test bull's eye pattern formation in WT, Δ*lrhA* and Δ*ydiQ* mutants carrying *lrhA* gene with SNP at position 278 changing encoded Glutamine to a STOP codon ($lrhA^{Q278\text{-}STOP}$) at varying IPTG concentration.

The results suggest that there was leaky expression of the *lrhA* variants from the pQE60 vector. This can be concluded based on the observation that in absence of IPTG, the strains carrying $lrhA^{wt}$ or $lrhA^{Q278\text{-}Stop}$ genes (Figure 29 and 30) showed the rescue - wild type phenotype.

Upon inducing expression of $lrhA^{wt}$ or $lrhA^{Q278\text{-}Stop}$ in *ΔlrhA* mutant, we observe the similar rescue phenotype, suggesting that both variants of *lrhA* gene (wild type and SNP carrying) are similar in function and result in suppression of bull's eye phenotype. On the other hand, induction of either variants of *lrhA* gene in *ΔydiQ* background, rescues the wild type non-motile phenotype. This could be due to higher expression of flagellar synthesis negative regulator, *lrhA, because* Δ*ydiQ* (carrying a functional copy of *lrhA* gene) in this study should natively express *lrhA* encoded on the genome, complimented by additional levels of *lrhA* from to the expression vector. In some plates, we do observe hyper-motile variants which instead of showing bull's eye phenotype start spreading on the agar surface. This could potentially be a result of spontaneous mutations in flagellar synthesis negative regulator such as *lrhA*, leading to de-repression of flagellar component synthesis resulting in motility of a subset of cells.

Effect of glucose on formation of bull's eye pattern in aerobic environment

Experiments were conducted to test the effect of glucose on bull's eye pattern formation of CGSC Δ*ydiQ* (carrying $lrhA^{Q278\text{-}Stop}$ SNP) and Δ*lrhA* in aerobic conditions. There have been reports suggesting that in presence of glucose (due to lower CRP/cAMP), *flhDC* expression is suppressed and hence flagella synthesis is affected [195]. Thus, we tested the potential of the mutants Δ*lrhA* and Δ*ydiQ* (carrying $lrhA^{Q278\text{-}Stop}$ SNP) to form bull's eye phenotype when SMMK was added with glucose or sucrose *(0.02 and 0.1% w/v).* The mutant strains were grown till 14 hours.

A pH indicator, phenol red (pH 8.0 > magenta red color; pH 6.8 < yellow), was used in the media. Yellow color indicates acidic pH which is due to fermentation of glucose by the bacterium.

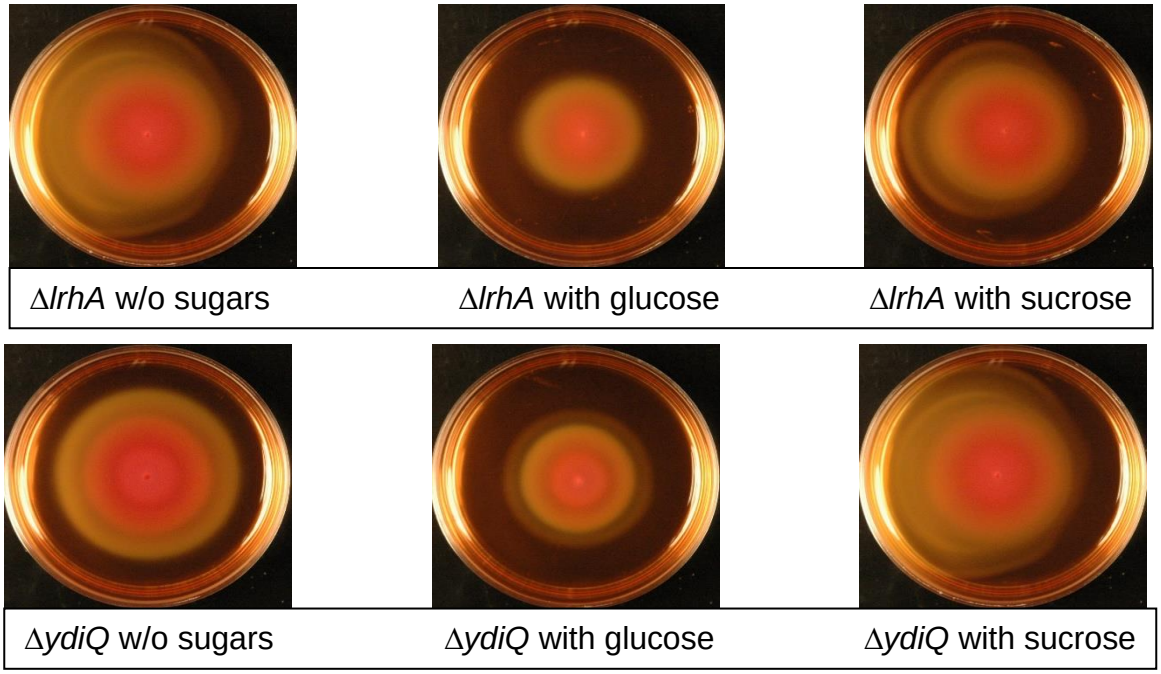


Figure 31 Bull's eye phenotype of Δ*lrhA* and Δ*ydiQ* in presence of 0.02% glucose or sucrose.

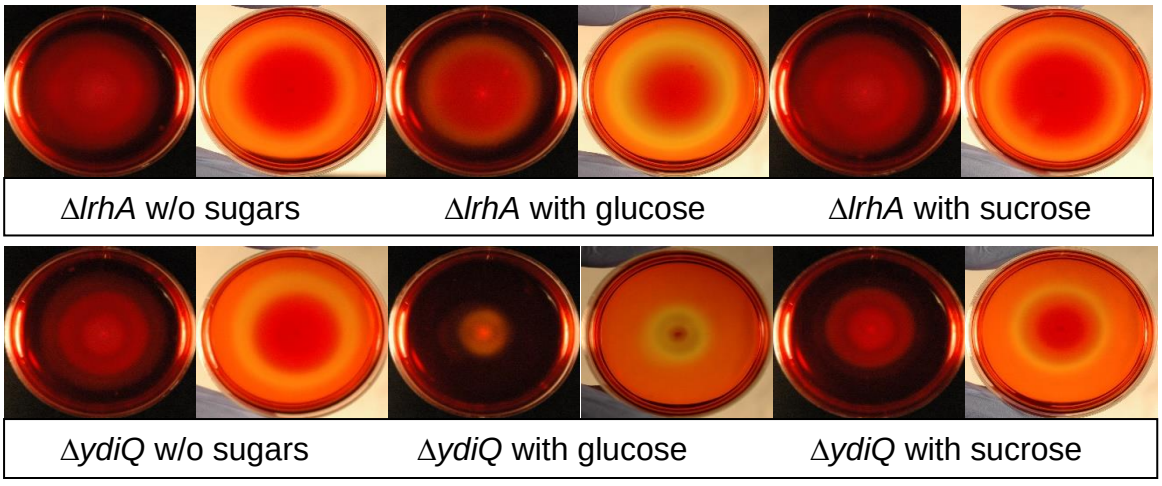


Figure 32 Bull's eye phenotype of Δ*lrhA* and Δ*ydiQ* in presence of 0.1% glucose or sucrose.

The Δ*ydiQ* mutant strains are suppressed in spreading and forming the bull's eye phenotype in presence of higher glucose concentration (shown in 0.1% w/v glucose presence, Figure 32).

The Δ*lrhA* strains are still motile, but they do not form the bull's eye pattern formation. Motility in the presence of glucose is unexpected, since glucose lowers the cyclic AMP/CRP levels [201], which are known to positively regulate expression of *flhDC* [202,203].

Conclusion

Predicted model describing the formation of the bull's eye pattern

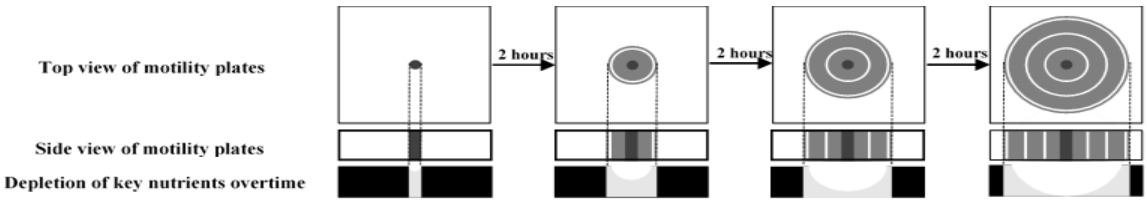


Figure 33 Predicting the bull's eye phenotype formation based on nutrient depletion on a motility plate.

E. coli switch between cellular division and flagellar synthesis, since both processes require high energy. Protein LrhA negatively regulates the expression of flagella synthesis by binding to the promoter of *flhDC*. Thus, there is an increase in flagellar expression in Δ*lrhA* cells. The model (Figure 33) describes how nutrient depletion drives the Δ*lrhA* strains to spread and form the bull's eye pattern on motility plates. When cells are inoculated, they divide by utilizing nutrients in the motility media. When depletion of certain nutrients is sensed, a subset of cells spread with the help of high flagellar assembly. Once the cells spread and have access to additional nutrients, they resume normal growth and divide to form the band region. While the cells divide, a subset of the cells are constantly moving forward accessing fresh resources to divide and grow. There is a small subset of these motile cells forming the inter-band region, which do not divide due to nutrient depletion, adjacent to the band region. This constant movement and growth, depending on availability of nutrients, is predicted to be the cause for the band and inter-band regions in the bull's eye pattern.