# Enzyme Evolution across Bacterial Species

By JohnCarlo Kristofich
MCDB, University of Colorado at Boulder

Defense date:
April 4th, 2016

Thesis Advisor:
Dr. Shelley Copley, MCDB

Defense Committee:
Dr. Xiang Wang, CHEM
Dr. Jennifer Martin, MCDB

## Table of Contents

Acknowledgements

I would like to thank my thesis advisor, Dr. Shelley Copley, for being my intellectual guide on the project and supporting all of my academic endeavors. I also want to thank my supervisors, Dr. Jamie Kershner and Dr. Juhan Kim, for teaching me essential laboratory skills and advising me on experimental design. Furthermore, I want to thank Andrew Morgenthaler and Jake Flood for answering question and providing knowledge on the subject. Last but not least, I would like to thank my family for supporting my goals and being proud of my accomplishments.

**Abstract**

Catalytic promiscuity provides a starting point for the evolution of new enzymes. When a secondary "promiscuous" activity becomes physiologically relevant, selection will favor genetic changes that improve upon this promiscuous activity. When the ability to do both activities limits growth, it is an inefficient bi-functional enzyme. The Copley lab at the University of Colorado at Boulder has found several genetic changes that increase fitness in a Δ*argC* strain of *Escherichia coli* BW25113 when growth is limited by an inefficient bi-functional enzyme (ProA*). The experiments described in this thesis investigate the genetic changes that increase fitness in a Δ*argC* strain of *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str.* SL1344 when limited by a similar inefficient bi-functional enzyme. In addition to identifying specific genetic changes, I aim to investigate how differences in the genetic background and physiology of the organism affect what changes are selected for and the adaptive potential of such changes.

# Chapter 1

## Introduction

### Promiscuous activity and the emergence of new enzymes

Ohno was the first to suggest that duplication of a gene encoding an enzyme could allow extra copies of the gene to accumulate mutations leading to the emergence of a new enzymatic activity while one copy reserves the original function (19). If this new activity helps a cell outgrow other cells around it, then it has increased fitness and is under positive selection. The problem with this model is that gene duplication will not be maintained if there is no benefit from having multiple copies (3). In most cases, gene duplications are either selectively neutral or detrimental due to the fitness costs of their maintenance, limiting the ability of a cell to maintain duplicated copies long enough for beneficial mutations to occur (12). However, solutions to this problem have been proposed based on the observation that most new enzyme activities closely resemble the original activity of the enzyme.

Active sites of specialized enzymes often bind molecules that have some resemblance to their natural or physiological substrates, although usually with much lower affinity (13). If an enzyme can bind an unintended substrate and catalyze a secondary reaction that is physiologically irrelevant, it is said to have "catalytic promiscuity" (9). If this secondary activity becomes physiologically relevant by increasing fitness, selection will favor genetic changes that take advantage of this new activity (10). The promiscuous activity of essential enzymes or those that are constitutively expressed are usually the first to be recruited in the face of an

environmental change due to their active and selectable presence in the cell (10). Alternatively, mutations that cause constitutive expression of non-essential genes that can be subsequently recruited to produce essential metabolites via their promiscuous activities have been documented (4, 23). In either case, promiscuous activities must be efficient enough to be physiologically sufficient, that is, the must affect the fitness of the cell by allowing survival or increasing growth rate.

There are a multiple ways that a promiscuous activity can be raised to physiologically sufficient levels. On the one hand, duplication can increase the concentration of an enzyme with promiscuous activity in the cell allowing the unintended substrate to be catalyzed to physiologically sufficient levels. Furthermore, if this secondary activity becomes necessary for survival, then duplication would be maintained allowing the possible divergence of the duplicated copies (10). Overproduction of an enzyme resulting in the emergence of a new activity to replace critical enzymes, digest toxins, and facilitate alternative routes for synthesis of important metabolites has be shown to occur (15). If a new activity could be raised to physiologically sufficient levels via duplication, then this is really promising for the evolution of new enzymes. However, gene duplication could have varying fitness impacts depending on the genetic context of the duplication and the physiology of the organism. The idea that duplication interferes with highly constrained cellular systems is termed the dosage-balance hypothesis (8). Furthermore, many promiscuous activities are too inefficient to confer a fitness benefit even after duplication (8, 10).

Alternatively, a mutation could increase the efficiency of the promiscuous activity to physiological sufficient levels without duplication. Mutations can cause structural

changes in the enzyme making it less conforming to its natural substrate and allowing this unintended substrate to more effectively compete (21). While such a mutation is very likely to have an impact on the necessary original activity of the enzyme and decrease fitness compared to the wild type, it selected for if it allows survival (17). This hypothesis predicts that a period of gene sharing is generally required for the emergence of new enzymatic activity (12). When the ability of an enzyme to do both activities limits growth of the cell, it is an inefficient bi-functional enzyme.

Once a promiscuous activity is raised to physiologically sufficient levels, selection will favor changes that continue to improve fitness. Gene duplications can be selected for if increasing the concentration of the enzyme in the cell increases fitness. However, point mutations that increase the fitness of the cell by causing structural changes or increasing expression of the inefficient bi-functional enzyme are equally selectable if they increase fitness to the same degree. Most importantly, it is clear that gene duplication can occur and be maintained even after point mutations improve upon the inefficient bi-functional enzyme, as long as there is room to increase fitness (18).

The innovation, amplification, and divergence model describes this process and the adaptive potential of such an inefficient bi-functional enzyme. The Nasvall paper describes experiments studying this process in *Salmonella enterica*. They started with a *trpF* knockout that acquired an initial mutation in *hisA* allowing the recovery of the TrpF function that is necessary for survival. Both HisA and TrpF enzymes catalyze similar reactions involving isomerization of a phosphoribosyl compound, but each act on different substrates. Some of the mutants were able to increase fitness via amplification of the *hisA* gene while others acquired mutations increasing the efficiency of the bi-

functional enzyme before amplification. After ~3000 generations, amplification allowed some copies to lose the previous mutation restoring HisA activity, while other copies acquired mutations increasing either TrpF activity or both activities. The gene for this bi-functional enzyme was cloned onto a low-copy plasmid in order to decrease the time and fitness cost of amplification so that the process could be studied in a reasonable time (18).

## Specific Context and Aim of Project

The Copley lab at the University of Colorado Boulder is currently investigating the initial events after the emergence of an inefficient bi-functional enzyme that allows survival of a Δ*argC* strain of *E. coli* on minimal medium. Normally, this strain cannot grow on minimal medium (no amino acid supplements) because it lacks the ability to synthesize arginine. ProA is an enzyme encoded by *proA* in the *proBA* operon and is necessary for proline synthesis (1). While ProA has an initial "promiscuous" activity towards the substrate of ArgC, it is not physiologically sufficient to allow survival of a Δ*argC* strain of *E. coli* on minimal medium. The chemical transformations catalyzed by ProA and ArgC are identical and the substrates for ProA and ArgC differ only in the absence or presence of an acetyl moiety on the amino group (Fig 1.1). A spontaneous mutation that changes glutamate 383 to alanine in ProA (denoted hereafter as ProA*) allows the Δ*argC* strain to grow on minimal medium (21). Dr. Kershner observed that Δ*argC E. coli* grows slowly on minimal medium while using ProA* for both proline and arginine synthesis when compared to the wild type *E. coli* (Jamie et al., 2015).

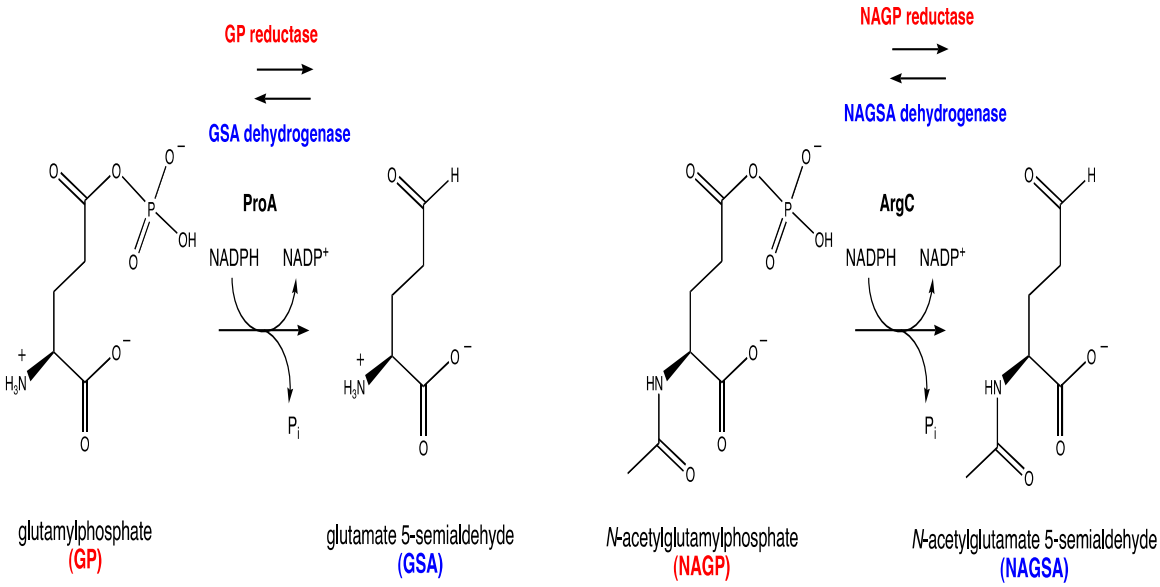



**Fig 1.1.** From Jamie et al., 2015. Shows the catalytic activity of ProA and ArgC; substrates differ by a single acetyl moiety group.

Dr. Kershner has found that after growth on minimal medium for as little as 6 days (~25 generations), mutants that grow faster than the progenitor (*E. coli* Δ*argC* E383A ProA) are observed. After further investigation, mutants were found to have one of several genetic changes that increase fitness. These include 3 different point mutations (M1, M2, and M3) and various amplifications of the operon (Fig 1.2, Table 1.1). M1 and M2 were found within the promoter region and determined to increase overall *proBA** transcript levels. M3, a synonymous mutation in the 3' end of the *proB* coding region, was found to increase transcript levels of *proA** without affecting transcript levels of *proB* by creating a new transcription start site in the middle of the operon (Jamie et al., 2015). What is more surprising is that M3 mutants grow nearly as fast as cells containing 50-fold amplifications, in which the level of transcript is markedly increased. Amplification strains were found to have varying 5' ends for the amplified region, but a conserved 3' end near a transposable element. Furthermore, amplification can occur in

Δ*argC proA** strains of *E. coli* containing M1, M2, or M3 if they are further adapted on minimal medium, albeit to a lesser degree (Jamie et al., 2015).

**Table 1.1.** Mutations identified by Dr. Jamie Kershner after growth of a Δ*argC proA** strain of *E. coli* on minimal medium.

| Mutation | Base change | Nucleotide Position (*proBA**) | Location |
|---|---|---|---|
| M1 | g->a transition | -30 | -35 region of the *proBA** promoter |
| M2 | c->t transition | -10 | -10 region of the *proBA** promoter |
| M3 | a->t transversion | 1044 | 3' end of *proB* |

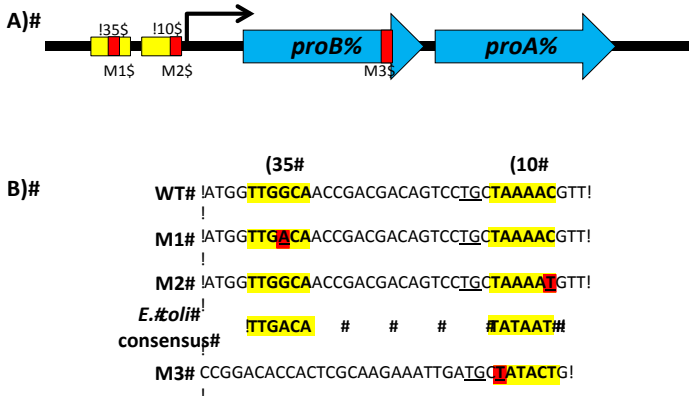



**Fig 1.2.** From Jamie et al., 2015. A) Representation of the *proBA* operon in *E. coli*. B) Mutations identified by Dr. Jamie Kershner after growth of a Δ*argC proA* strain of *E. coli* on minimal medium.

The pairwise sequence identity between the *E. coli* and *S. enterica* ProA enzymes is 86%. Neutral drift over time can result in sequence divergence among enzymes used to catalyze the same reaction that will have varying effects on the promiscuous activity of the enzyme (5). A study done by Khanal comparing the *in vitro* activity of ProA* from 9 different microbes, including *E. coli* and *S. enterica*, shows that the initial promiscuous activity of ProA varies depending on the species. Furthermore, the effect of changing ProA to ProA* varied by species and was not correlated with the initial level of the

promiscuous activity. Khanal hypothesized that the effects of sequence divergence among orthologs effects not only the initial levels of promiscuous activities, but the evolvability of promiscuous activities as well (13). In *S. enterica,* the glutamate corresponding to E383 of the *E. coli* enzyme is at position 382 of the *S. enterica* enzyme (16). A Δ*argC* strain of *S. enterica* can't grow on minimal medium because it lacks the ability to synthesize arginine. However, a Δ*argC proA** strain of *S. enterica* can grow slowly on minimal medium by using ProA* for both proline and arginine synthesis. The experiments described in this thesis investigate the initial genetic changes that increase fitness of a Δ*argC* strain of *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str.* SL1344 when limited by the inefficient bi-functional enzyme ProA*. In addition to identifying specific genetic changes, I aim to investigate how differences in the genetic background and physiology of the organism affect what changes are selected for and the adaptive potential of such changes.

In chapter two, I sought to identify the initial genetic changes that increase fitness of a Δ*argC* strain of *S. enterica* whose growth rate is limited by ProA*. First, I plated this strain on plates containing M9/glucose and identified faster-growing colonies. Next, these faster-growing colonies were analyzed by qPCR to determine whether amplification of the *proBA** operon had occurred. Mutants for which amplification of the *proBA** operon was not detected were screened using PCR and subsequent Sanger sequencing to identify mutations in the *proBA** operon (Macrogen). The promoter mutation, M2, was the only genetic change identified in faster-growing colonies of the Δ*argC proA** strain of *S. enterica* after growth on minimal medium.

The third chapter involves taking a deeper look at why M1 and M3 were not found in faster-growing colonies of the Δ*argC proA** strain of *S. enterica* after growth on minimal medium. First, M1, M2, and M3 were introduced into the Δ*argC proA** strain of *S. enterica.* Next, growth rates were determined for each mutant of the Δ*argC proA** strain of S. *enterica* in minimal media or minimal media supplemented either with arginine, proline, or both. Lastly, I tried to compare the data to a previously determined growth rates of M1, M2, and M3 mutants of the Δ*argC proA** strain of *E. coli* in order to elucidate why the mutations M1 and M3 were not identified in faster-growing colonies of the Δ*argC proA** strain of *S. enterica* after adaptation on minimal medium

In the fourth chapter, I explore whether extra time for adaptation would result in the amplification of the *proBA** operon. To do this, I adapted 4 cell lineages of the Δ*argC proA** strain of *S. enterica* for ~250 generations in M9/glucose liquid medium and looked for increases in growth rate over time. At the end of the adaptation, genomic DNA from the population for each cell lineage was collected during four time points and analyzed using qPCR in order to check for amplification of the *proBA** operon.

The final chapter explores why amplification was not observed in faster-growing colonies of the Δ*argC proA** strain of *S. enterica* after growth on minimal medium. To do this, I did preliminary experiments examining the effect of amplification on the expression of genes surrounding the *proBA** operon in amplification mutants of the Δ*argC proA** strain of *E. coli*. Two strains of the Δ*argC* strain of *E. coli* carrying the allele for ProA* with different degrees of amplification were selected. Label-free mass spectrometry and qPCR were used to determine if the expression of genes within the

amplified region increases as gene copy number increases or if homeostatic mechanisms will keep protein expression at a constant level.

## Chapter 2

## Identifying genetic changes that increase fitness of a Δ*argC* strain of *S. enterica* whose growth rate is limited by ProA*

Dr. Jamie Kershner identified several genetic changes that improve fitness of a Δ*argC* strain of *E. coli* whose growth rate is limited by an inefficient bi-functional enzyme, E383A ProA (ProA*). I chose to work with *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str.* SL1344 because of its previous use in the Copley lab and the amount of genetic manipulation techniques designed for it. Similar to a Δ*argC* strain of *E. coli*, a Δ*argC* strain of *S. enterica* can't grow on plates containing M9/glucose (no amino acid supplements) because it can't synthesize arginine. The change in ProA (E382A) to ProA* allows slow growth of a Δ*argC* strain of *S. enterica* on minimal medium.

This chapter describes experiments directed toward identifying the initial genetic changes that increase fitness in a Δ*argC proA* strain of *S. enterica* whose growth rate is limited by the inefficient ability of ProA* to catalyze the reduction of N-acteylglutamyl phosphate and γ–glutamyl phosphate (Fig 1.1). These initial changes are important because they effect the epistatic interactions of successive mutations thereby changing the possible evolutionary trajectories (13). I started by plated the Δ*argC proA** strain of *S. enterica* on M9/glucose and identified faster-growing colonies. The copy number of *proA** in 100 of these faster-growing colonies was analyzed using qPCR to determine whether amplification of the *proBA** operon had occurred. Lastly, mutants determined to

have no amplification were analyzed using PCR and subsequent Sanger sequencing to identify mutations in the *proBA** operon (Macrogen).

## Results

Growth of a Δ*argC proA** strain of *S. enterica* on M9/glucose (0.2%) for ~5-6 days yielded faster-growing colonies

To begin, I streaked an aliquot of a freezer stock of the Δ*argC proA** strain of *S. enterica* onto plates containing LB. Two individual colonies were used to prepare freezer stocks serving as genetic replicates (JK328 and JK329) for the following experiments. Both genetic replicates were used during these experiments and are referred to together as the "parental strain". (A list of all strains used in this thesis can be found in the Supplemental Information (Table S.1).) I collected 100 faster-growing colonies from two experiments that were done at different times. In both experiments, aliquots of a freezer stock of JK328 and JK329 were streaked onto plates containing LB and incubated overnight at 37 °C. A single colony for each genetic replicate was washed with sterile PBS before being spread onto plates containing M9/glucose. In addition, the same amount of cells was spread on plates containing LB to determine the average number of cells spread on M9/glucose plates. The plates were incubated at 37 °C for 5-6 days or until faster-growing colonies were observed (Fig 2.1). The number of colonies per plate varied between the two experiments, as did the number of faster-growing colonies per plate (Table 2.1).

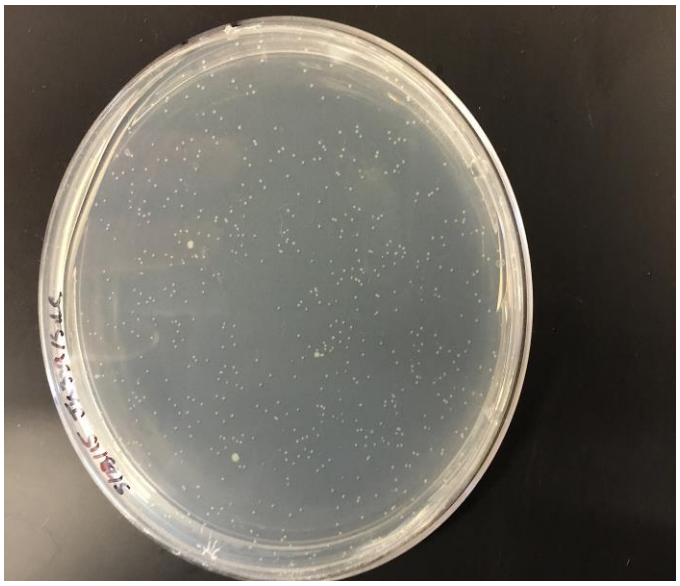


**Fig 2.1.** Growth of the Δ*argC proA** strain of *S. enterica* on plates containing M9/glucose for 5-6 days at 37 °C yields faster-growing colonies that can be distinguished from smaller colonies that do not have increased fitness.

**Table 2.1.** Average number of colonies per plate as determined from M9/glucose or LB plates. Average number of faster-growing colonies perM9/glucose plate.

| **A** | **Ave. number of colonies/plate (M9 plates) ± s.d.** | **Ave. number of colonies/plate (LB plates) ± s.d.** | **Ave. number of faster-growing colonies/plate (M9 plates) ± s.d.** |
|---|---|---|---|
| JK328 | 223 ± 47.7 | 338 ± 12.5 | 3.53 ± 1.7 |
| JK329 | 222 ± 9.2 | 384 ± 25.1 | 4.53 ± 2.0 |

| **B** | **Ave. number of colonies/plate (M9 plates) ± s.d.** | **Ave. number of colonies/plate (LB plates) ± s.d.** | **Ave. number of faster-growing colonies/plate (M9 plates) ± s.d.** |
|---|---|---|---|
| JK328 | 776.6 ± 37.2 | 833.3 ± 28.9 | 2.6 ± 1.3 |
| JK329 | 1180 ± 36.8 | 1150 ± 50 | 4.87 ± 2.1 |

qPCR reveals that none of the faster-growing colonies obtained after growth of the Δ*argC proA** strain of *S. enterica* on M9/glucose contains amplification of the *proBA** operon

To determine whether amplification of the *proBA** has occurred in faster-growing colonies observed after growth of the Δ*argC proA** strain of *S. enterica* on M9/glucose, qPCR was done to determine the *proA** copy number present in the cell lysates of faster-

growing mutants using *icdA* and *gyrB* as internal reference genes. All of the reagents used for qPCR were tested for contamination before doing any experiments by running a PCR reaction without template and loading the product on a 2% agarose gel (data not shown). I designed three primer sets for each of the three target genes. (A list of all primers used in this thesis can be found in the Supplemental Information (Table S.2).) A standard curve was generated for each primer set using four 1:10 serial dilutions of genomic DNA from the parental strain JK328 serving as my standards. Auto-threshold was used when calculating primer efficiencies from the standard curve, but then manually maintained thereafter. Primer sets that had the highest efficiencies were used to assess *proA** copy number of faster-growing colonies (Table 2.2). PCR reactions using each primer set with a lysate of the parental JK328 strain were run on a 2% agarose gel to check for side products or contamination. Furthermore, the products were sent for Sanger sequencing to make sure the sequence of the amplified products align to the expected target genes and not some other location in the genome (data not shown) (Macrogen).

**Table 2.2.** Primer sets designed for qPCR and their efficiencies determined by the standard curve. ***proA*set1**, ***icdA*set2**, and ***gyrB*set1** were used to determine *proA* copy number in cell lysates of faster-growing colonies.

| Primer set | Auto-threshold | Primer set efficiencies (%) |
|---|---|---|
| ***proA*set1** | 1.45 | 92.3 |
| *proA*set2 | 1.87 | 90.4 |
| *proA*set3 | 1.97 | 90.7 |
| *icdA*set1 | 1.19 | 91.3 |
| ***icdA*set2** | 0.99 | 93.1 |
| *icdA*set3 | 2.08 | 92.6 |
| ***gyrB*set1** | 1.05 | 96.0 |
| *gyrB*set2 | 2.72 | 91.6 |
| *gyrB*set3 | 2.17 | 91.8 |

Cell lysates of 100 faster-growing colonies were screened with the best primer set for each target gene on the same plate in order to eliminate the need for inter-run calibrators. Cell lysates of faster-growing colonies were screened initially without replicates and those that showed a *proA** copy number relative quantity (CNRQ) value greater than 1.5 or less than 0.5 were re-screened in duplicate. The first and last dilutions of the parental strain JK328 genomic DNA standards were used as the reference for determining the *proA** copy number of cell lysates from faster-growing colonies. Primer efficiencies and average Ct values for each reaction were exported to Excel for analysis with qBASE (11). None of the 100 faster-growing colonies had an increase in *proA** copy number (Fig 2.2).

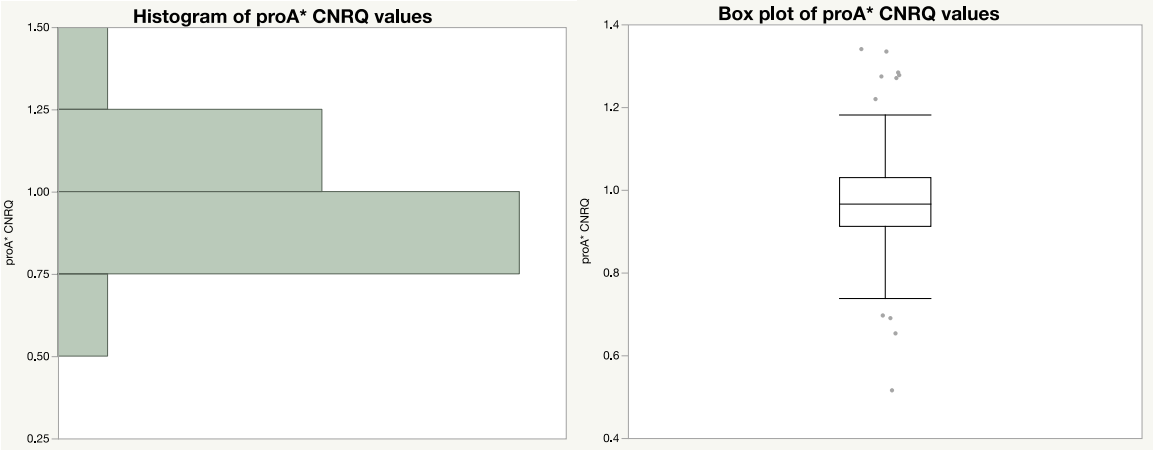



**Fig 2.2.** A histogram and box plot showing *proA** CNRQ values for 100 faster-growing colonies determined by qPCR of cell lysates.

All of the faster-growing colonies obtained after growth of the Δ*argC proA** strain of *S. enterica* on M9/glucose contain the promoter mutation M2

I sequenced the *proBA** operon using genomic DNA prepared from faster-growing colonies collected from M9/glucose plates containing the *S. enterica* parental strain. Two primer sets were designed to amplify the upstream and downstream segments of the operon with a ~500 bp overlap. These segments were PCR-amplified from the cell lysates of faster-growing colonies as well as the parental JK328 strain. The amplified PCR product using the cell lysate of my parental JK328 strain was run on a 2% agarose gel with a DNA ladder to assess the size of the amplified fragment and to check for side products that could interfere with Sanger sequencing (Fig 2.3). Sanger sequencing of the PCR-amplified products from cell lysates of faster-growing colonies was done to identify any mutations within the *proBA** operon (Macrogen). I sequenced the entire *proBA** operon for 20 of the 100 faster-growing colonies; only M2 in the promoter was found. For an additional 30 large colonies, only the promoter region was sequenced in order to reduce the cost of sequencing. In total, 50 out of 100 faster-growing colonies were determined to have M2.

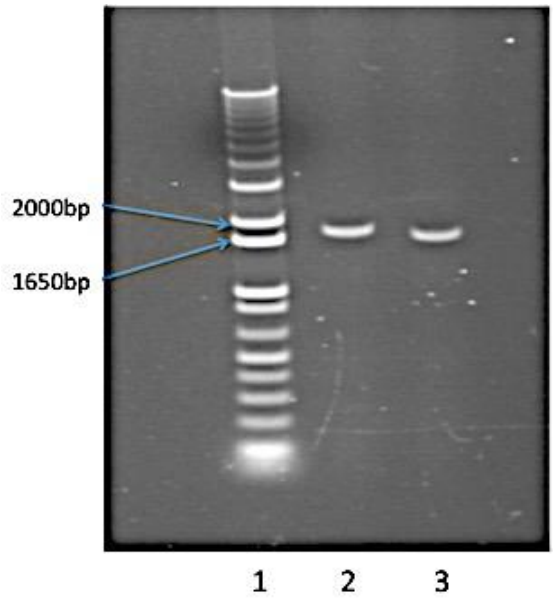



**Fig 2.3.** Visualization of PCR-amplified segments of the *proBA** operon from cell lysate of the parental strain (JK328) to check their size with a 1 kb plus DNA ladder (Invitrogen, lane 1). The 3' downstream segment (lane 2) was PCR-amplified with primers JC-P2 and JC-P3 and has an expected size of 1762 bp. The 5' upstream segment (lane 3) was PCR-amplified with primers JC-P25 and JC-P25 and has an expected size of 1657 bp.

## Discussion

In this chapter, I have discussed the identification of genetic changes that increase fitness in a Δ*argC proA** strain of *S. enterica* whose growth rate is limited by the inefficiency of ProA*. These findings suggest that amplification does not occur after adaptation of Δ*argC proA** *S. enterica* strain on M9/glucose. This is interesting considering that various extents of amplification are observed after adaptation of the Δ*argC proA** *E. coli* strain on M9/glucose (Jamie et al., 2015). On the one hand, it could be that amplification of *proBA** and the surrounding regions has a more detrimental fitness impact in *S. enterica* than in *E. coli* (8). This seems plausible given the differences in genetic context downstream of the *proBA* operon, but can't be concluded based on these results (Fig 2.4). Alternatively, It may be that the absence of an IS5 transposable element in *S. enterica* that is consistently observed at the 3' end of the amplified region in *E. coli* is making amplification of the region less frequent (Jamie et al., 2015). In either case, point mutations that do increase fitness do not prevent amplification from occurring at a later time point if fitness can still be improved (18). Experiments in the following chapters attempt to address these questions.

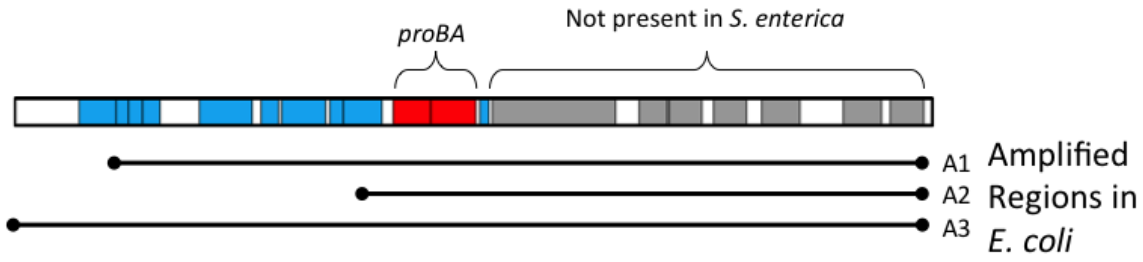



**Fig 2.4.** Map of the amplified regions identified in *E. coli* mutants. Top bar is a representation of the genome with colored boxes indicating genes. Red: *proBA* operon, Blue: genes found in both *E. coli* and *S. enterica*, Grey: genes found only in *E. coli*. Black bars indicate the region amplified in three different mutants (A1, A2 and A3). All amplifications identified in *E. coli* have the same 3' end, which is not present in *S. enterica*.

In the 50 faster-growing colonies I screened, only M2 was observed. M2 is a promoter mutation that was determined to increase the overall transcript levels of *proBA** in *E. coli*. The absence of M1 and M3 in faster-growing colonies may be due to a low frequency of these mutations, or a low impact of these mutations on fitness. Depending on the frequency of these mutations, screening enough faster-growing colonies to find these mutations could be costly and time consuming. The use of genomic editing to create these mutations will be used to assess their fitness effects in Chapter 3.

## Methods

*S. enterica ΔargC* E382A ProA strain (parental strain)

*Salmonella enterica* subsp. *enterica serovar* Typhimurium *str.* SL1344 was obtained from the Keio collection. *S. enterica* is a histidine auxotroph, Dr. Juhan Kim used the pSLTS method to fix a nonsense mutation in *hisG*, a gene required for the production of histidine. Furthermore, he used this method to knockout *argC* and change glutamate 382 to alanine in ProA (14). *S. enterica* has a natural resistance to streptomycin (strep), which was supplemented in the medium periodically to select against other bacterial species.

Plating the parental Δ*argC proA** strain of *S. enterica* on M9/glucose and collecting faster-growing colonies

An aliquot of a freezer stock of the parental Δ*argC proA** strain of *S. enterica* was streaked onto plated containing LB/strep (50 μg/mL) and incubated at 37 °C overnight. Two individual colonies serving as genetic replicates (JK328 and JK329) were

washed twice with 100 µL sterile PBS, suspended in 100 µL sterile PBS, and diluted 1:10000 (D1) for subsequent plating. Two µL of each suspension were used to inoculate 1 mL LB/strep (50 µg/mL) and grown at 37 °C until mid-log to make freezer stocks (10% DMSO (v/v)).

Fifteen, 50 µL aliquots of dilution D1 for each parental strain were spread on plates containing M9/glucose (0.2%)/strep (50 µg/mL) using sterile glass beads for a total of 30 plates. Plates were incubated at 37 °C for 5-6 days or until faster-growing colonies are observed. Three, 50 µL aliquots of dilution D1 for each parental strain were also spread on plates containing LB/strep (50 µg/mL) using sterile glass beads to evaluate the amount of cells spread on the M9/glucose plates.

After incubation of the M9/glucose plates at 37 °C for 5-6 days, 25 faster-growing colonies from each genetic replicate (JK328 and JK329) were each suspended in 20 µL sterile $H_2O$, streaked onto plates containing M9/glucose (0.2%)/strep (50 µg/mL), and incubated overnight at 37 °C. The remainder of each suspension was boiled at 98 °C for 10 min, diluted with 40 µL sterile $H_2O$, and stored at -20 °C for copy number determination by qPCR or Sanger sequencing.

Four sub-clones from each isolation streak were suspended in 20 µL sterile $H_2O$ (200 total). Two µL from each suspension were used to inoculate 0.5mL M9/glucose (0.2%)/strep (50 µg/mL) liquid media which was incubated overnight at 37 °C with shaking until mid-log and used to make freezer stocks (DMSO 10% (v/v)). The remainder of each suspension was boiled at 98 °C for 10 min, diluted with 40 µL sterile $H_2O$, and stored at -20 °C for copy number determination via qPCR or Sanger sequencing.

Designing primer sets for qPCR, making the standard curve, and determining *proA* CNRQ values from cell lysates of faster-growing colonies

Primer3Plus (Bioinformatics) was used to design primers that anneal to *proA*, *icdA*, and *gyrB* of *S. enterica* and can be used to amplify 50-150 bp fragments using PCR. NCBI was used to obtain the target sequences for this strain. Settings used were: Na+, 50 mM; Mg++, 1.5 mM; dNTPs, 0.6 mM; target type, DNA; and oligo concentration, 0.25 μM. mFold was used to estimate the ΔG of secondary structures in the amplified product that could interfere with accurate primer efficiencies, especially when secondary structures block where your primers bind. Settings were the same as those for Primer3Plus. IDT Oligo-analyzer was used to estimate the ΔG of self/hetero primer dimerization to assure that the primers would not anneal to each other and produce faulty primer efficiencies. Three primer sets for each of the target genes were made in order to increase the chance of getting a good primer set.

The genes *icdA* and *gyrB*, assumed to be present as a single copy in the genome, were used as internal reference genes for qPCR estimation of the copy number of *proA*. Cycling conditions for a standard run with Power SYBR Green in a 7500 Fast Real-Time PCR system (Applied Biosystems) were as follows: 95 °C for 10 min; 40 cycles of 95 °C for 15 s, 60 °C for 60 s. A melt curve was performed immediately following qPCR in order to test for multiple products or contamination of the samples.

A standard curve was generated for each primer set using four 1:10 serial dilutions of genomic DNA from the parental strain JK328 serving as my standards. Two μL of each parental strain (JK328) genomic DNA dilution were used in triplicate with 0.4

µL each of the 10 µM forward and reverse primers, 7.2 µL sterile $H_2O$, and 10 µL 2X SYBR Green master mix (Applied Biosystems). Reactions mixtures without template for each primer set were included during each run in order to check for contamination. Auto-threshold, cycling conditions, annealing temp, and auto-baseline settings were not changed when determining primer efficiencies from the standard curve. However, thresholds determined for a primer set during the standard curve experiment were used for further experiments using those primers. The Primer sets with the highest efficiency determined using the 7500 Software v2.3 were selected for mutant screening (Table 2.2).

Faster-growing colonies were screened with the best primer set for each target gene on the same plate in order to eliminate the need for inter-run calibrators. Reaction mixtures contained 2 µL template, 0.4 µL each of the 10 µM forward and reverse primers, 7.2 µL sterile $H_2O$, and 10 µL 2X SYBR Green master mix (Applied Biosystems). The first and last dilutions of the parental strain JK328 genomic DNA standards were used as the reference for determining the CNRQ of *proA** of faster-growing colonies. The average efficiencies for each primer set and the average Ct values for each reaction were exported into Excel for analysis with qBase in order to generate the copy number relative quantity (CNRQ) of *proA** for each mutant (11). Mutants were screened initially without replicates, and those that showed a CNRQ value greater than 1.5 or less than 0.5 were re-screened in duplicate. For mutants with confirmed copy number increase, sub-clones were used to prep genomic DNA and screened in triplicate to get a more accurate CNRQ for *proA**.

Identification of mutations in the *proBA** operon of faster-growing colonies using Sanger sequencing

Primers were designed to amplify two fragments that have an ~500 bp overlap and cover the entire *proBA* operon and its regulatory regions (~3 kb). Aliquots of freezer stocks for sub-clones prepped from mutants were streaked onto plates containing LB/strep (50 µg/mL) and incubated overnight at 37 °C. Single colonies were suspended in 50 µL sterile $H_2O$, boiled at 98 °C for 10 min, and stored at -20 °C prior to PCR amplification. Two µL of the mutant cell lysate were used in a 30 µL PCR reaction with 15 µL OneTaq HS 2X master mix, 0.6 µL of each 10 µM forward and reverse primers, and 11.8 µL sterile $H_2O$. Three µL of each PCR product was run on a 2% agarose gel in order to assess the size with a 1 kb plus DNA ladder (GeneRuler). The remainder of the PCR products were loaded and run on a 2% agarose gel for extraction using a GeneJET Gel Extraction Kit or simply purified using a GeneJET PCR Purification Kit (ThermoScientific) and eluted in 30 µL EB. Lastly, 100-200 ng of the eluted DNA was sent with each primer that was used to amplify it for Sanger sequencing (Macrogen).

Preparation of M9/glucose (0.2%) plates (1 L = 40 plates)

In 0.989 L deionized-$H_2O$, add and autoclave M9 salts

- 6.78 g $Na_2HPO_4$
- 3 g $KH_2PO_4$
- 1 g $NH_4Cl$
- 0.5 g NaCl

Add the following after sterilization

- 10 mL 20% glucose
- 1 mL 1 M $MgSO_4$
- 100 µL 1 M $CaCl_2$

## Chapter 3

## Analysis of the effect of M1, M2 and M3 on growth rate of the Δ*argC proA** strain of *S. enterica*

All of the faster-growing colonies obtained after growth of the Δ*argC proA** strain of *S. enterica* on M9/glucose contain the promoter mutation M2. This finding is in contrast to the identification of 3 different point mutations (M1, M2, and M3) in faster-growing colonies obtained after growth of the Δ*argC proA** strain of *E. coli* on M9/glucose. The first two mutations, M1 and M2, occur in the promoter region of the *proBA* operon and were determined to increase fitness in the *E. coli* parent strain by increasing overall transcript levels of the *proBA* operon (Jamie et al., 2015). The third mutation, M3, is a synonymous mutation at the 3' end of *proB* that was determined by RT-qPCR to increase transcript levels of only *proA* by creating a second promoter (Jamie et al., 2015). Synonymous mutations are implicated in adaptation and evolution by changing translation efficiency, creating new mRNA secondary structures, and affecting regulation of transcription (2, 6, 7). This chapter describes the introduction of comparable mutations into the genome of the Δ*argC proA** strain of *S. enterica* and assessment of their effects on growth rate to determine whether their absence in faster-growing colonies is due to a low frequency of these mutations, or a low impact of these mutations on fitness. In addition, M2 was inserted back into a Δ*argC proA** strain of *S. enterica* to test the fitness impact of this mutation in the absence of other mutations possibly acquired during growth on M9/glucose.

The pSLTS method for scarless genome editing by Dr. Juhan Kim was used to introduce these mutations back into a Δ*argC proA** strain of *S. enterica*. First, mutation template plasmids were made for each mutation (M1, M2, and M3) using Gibson assembly and were used to transform chemically competent *E. coli* MACH1 cells (Fig 3.1, A). These cells were cultured and harvested to obtain the mutation template plasmid. The mutation cassettes were amplified from harvested mutation template plasmids using PCR (Fig 3.1, B). Cells of the Δ*argC proA** strain of *S. enterica* containing the pSLTS plasmid were cultured in media containing arabinose to induce expression of λ-red recombinases necessary for mutation cassette transformation (Fig 3.1, C). Once cells were transformed with the mutation cassette and selected for having the selection marker, they were plated on anhydrotetracycline to remove the selection marker and restore the correct ORF of the gene with the intended mutation (Fig 3.1, D). A growth curve was generated for each clone in M9/glucose and M9/glucose supplemented either with proline, arginine, or both.

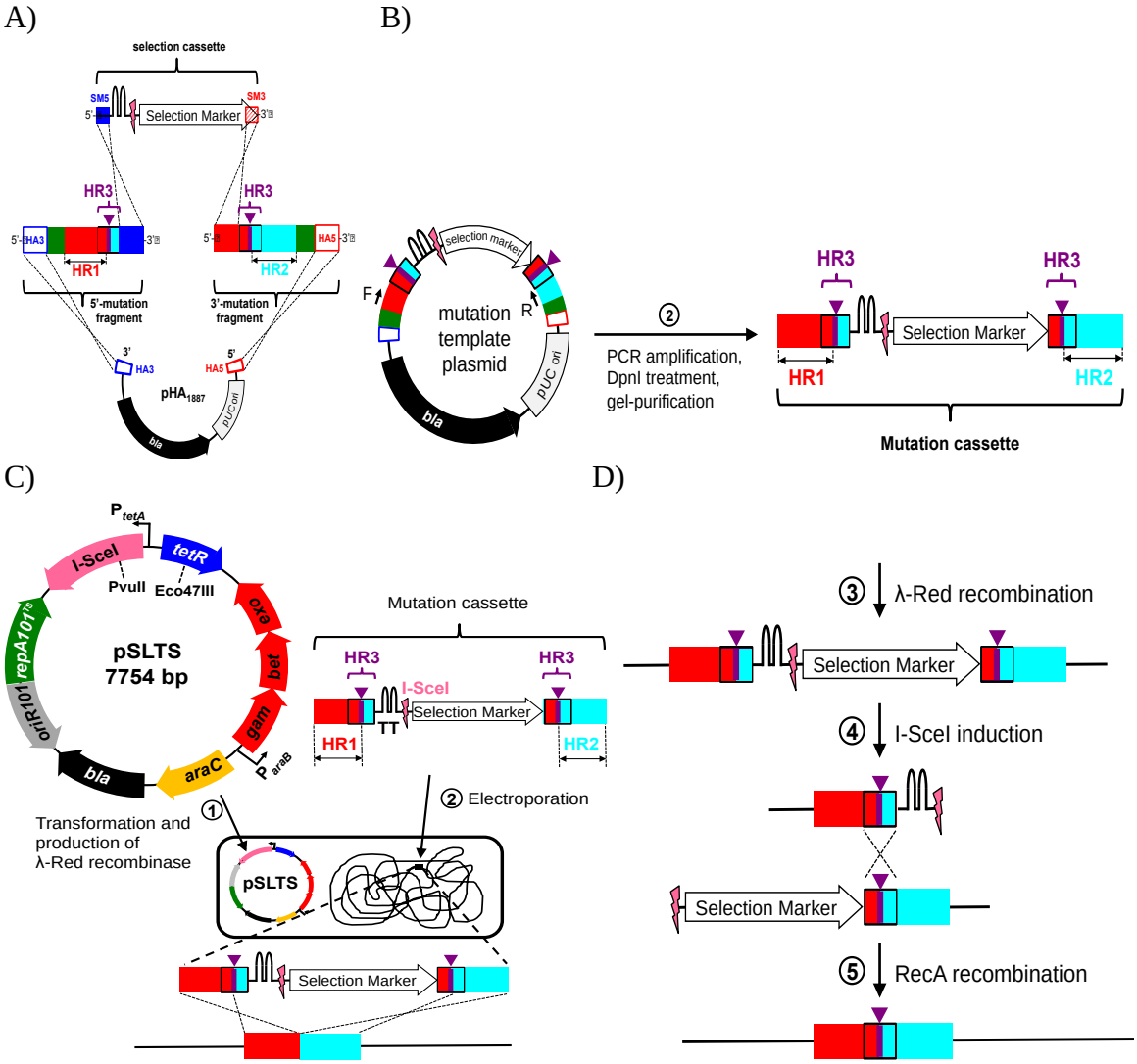



**Fig 3.1.** From Kim, 2014. Diagrams the pSLTS method developed by Dr. Juhan Kim to do scarless genome editing of the *S. enterica* genome that I used to introduce M1, M2 and M3 into the Δ*argC proA** strain of *S. enterica*. A) Construction of the mutation template plasmid B) amplification of the mutation cassette C) Introduction of the mutation cassette D) removal of the selection marker.

## Results

Introduction of M1, M2 and M3 into the Δ*argC proA** strain of *S. enterica*

The upstream and downstream fragments for the M2 mutation template plasmid were amplified by PCR from a Δ*argC proA* strain* of *S. enterica* containing M2 (JC67) using primers designed to add overhangs overlapping the selection marker and pUC19 backbone (Fig 3.2). These overhangs are necessary for Gibson assembly of the M2 mutation template plasmid. The PCR products containing the amplified upstream and downstream fragments were run on a 2% agarose gel in order to assess their size with a DNA ladder (Fig 3.3). gBLOCKs of the upstream and downstream fragments for the M1 and M3 mutation template plasmids were ordered from IDT. The Gibson assembly master mix was used to assemble each mutation template plasmid from the corresponding upstream and downstream fragments, a linear pUC19 backbone, and the CAT selection marker (NEB). The pUC19 backbone contains a gene for ampicillin (amp) resistance and the CAT selection marker contains a gene for chloramphenicol (chl) resistance. Chemically competent *E. coli* Mach1 cells obtained from Dr. Juhan Kim were transformed with the mutation template plasmid using the heat-shock method (NEB). Overnight growth on plates containing LB with ampicillin and chloramphenicol selects against cells that have not taken up the mutation template plasmid. Cells that were successfully transformed with the mutation cassette plasmid observed for each mutation template plasmid (Fig 3.4).

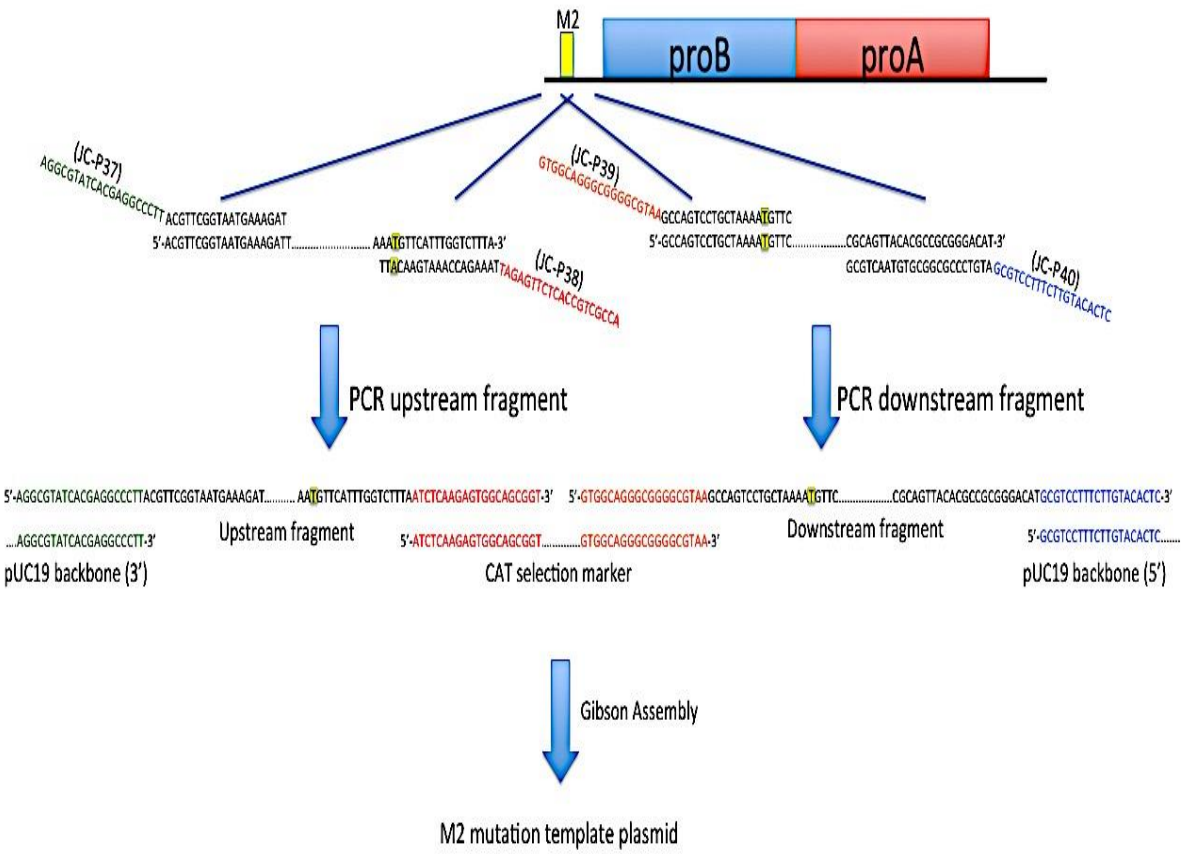



**Fig 3.2.** Primers were designed to amplify the upstream and downstream fragments using PCR from a Δ*argC proA** strain of *S. enterica* containing M2 (JC67) and add overhangs that overlap the selection marker and pUC19 backbone necessary for Gibson assembly of the M2 mutation template plasmid (NEB).

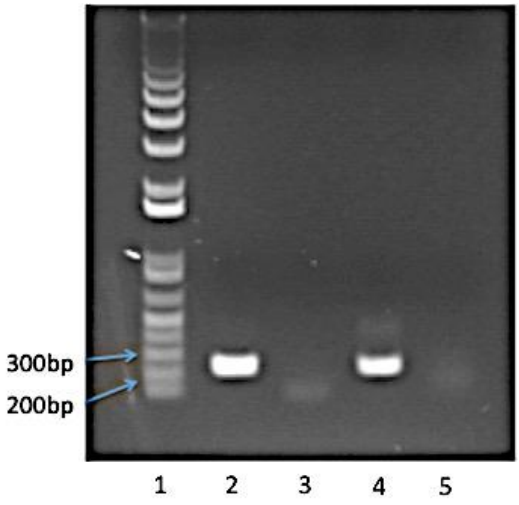



**Fig 3.3.** Visualization of PCR-amplified segments of the *proBA** operon from cell lysate of a Δ*argC proA** strain of *S. enterica* containing M2 (JC67) to check their size with a kb plus DNA ladder (Invitrogen, lane 1). The upstream fragment (lane 2) was PCR-amplified with primers JC-P37 and JC-P38 and has an expected size of 228 bp; a no template control was done for this reaction (lane 3). The downstream fragment (lane 4) was PCR-amplified with primers JC-P39 and JC-P40 and has an expected size of 227 bp; a no template control was done for this reaction (lane 5).

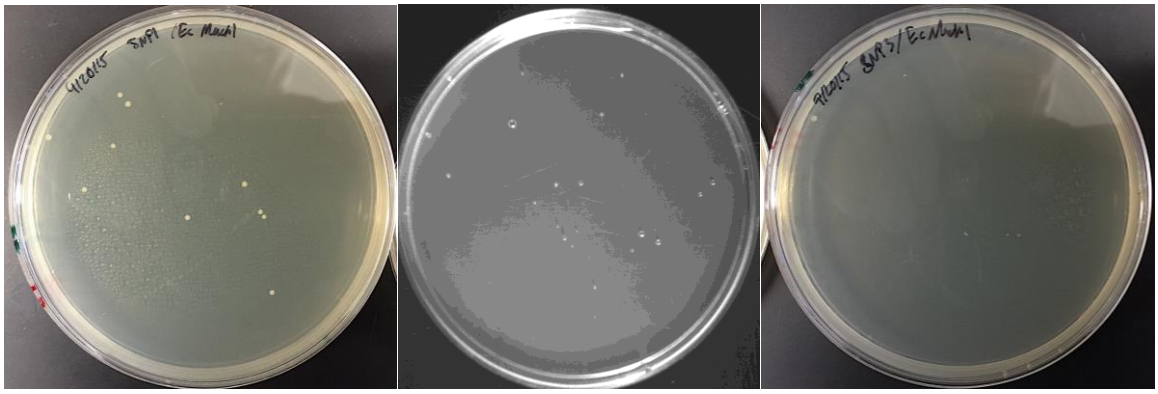


**Fig 3.4.** Overnight growth of *E. coli* MACH1 cells on plates containing LB, amp (100 µg/mL) and chl (20 µg/mL) after transformation with the mutation template plasmid using the heat-shock method (NEB). From left to right: M1, M2, and M3.

Primers that anneal to the pUC19 backbone of the mutation template plasmid and flank the mutation cassette were used to screen the cell lysate of colonies that survived overnight growth on LB/amp/chl for the correct assembly. The size of the amplified fragment was assessed on a 2% agarose gel using electrophoresis with a DNA ladder prior to sending it for Sanger sequencing (Fig 3.5). One colony from each mutation template plasmid transformation (M1, M2, and M3) that was confirmed to have the correct assembly was labeled and used to prepare freezer stocks: JC518 for M1, JC546 for M2, and JC519 for M3.

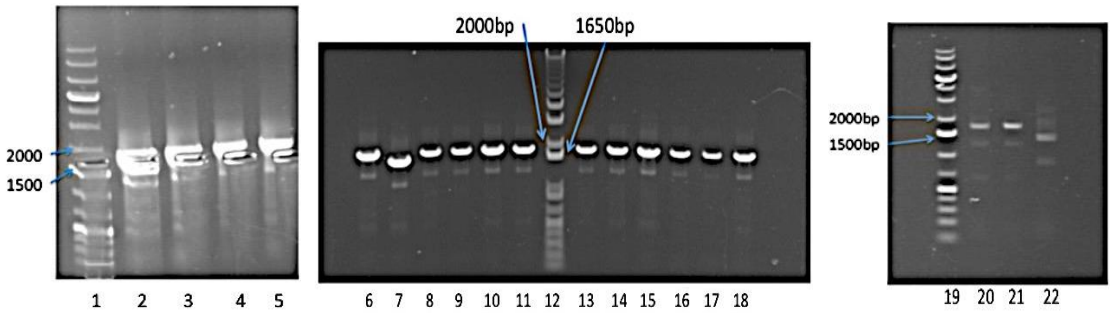



**Fig 3.5.** Visualization of PCR-amplified segments containing the mutation cassette from cell lysates of *E. coli* MACH1 colonies that survived overnight growth on plates containing LB/amp/chl after to assess their size with a 1 kb plus DNA ladder from either (Invitrogen, lane12) or (Gene ruler, lane 1 and 19). The segment was PCR-amplified with primers pH.seq.F and pH.seq.R and has an expected size of 1757 bp for M1 (lanes 2-5), 1723 bp for M2 (lanes 6-11 and 13-18), and 1757 bp for M3 (lanes 20-22).

Primers were designed to amplify the mutation cassettes from the mutant template plasmids using PCR. JC518, JC546, and JC519 cells containing the M1, M2, and M3 mutation template plasmids respectively were cultured in LB media containing amp and chl and harvested to obtain a large quantity of each mutation template plasmid. Primers designed to amplify each mutation cassette were used in a PCR reaction with the corresponding mutation template plasmid. The sizes of the amplified mutation cassettes were assessed by electrophoresis on a 2% agarose with a DNA ladder (Fig 3.6).

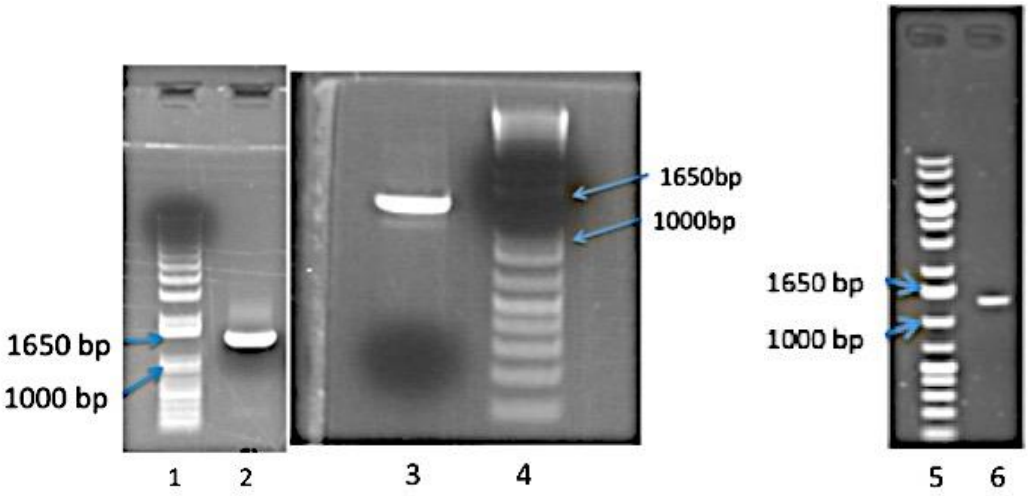
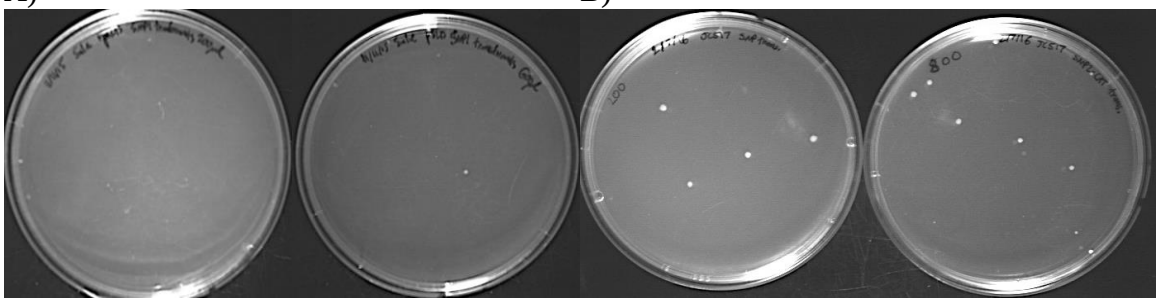



**Fig 3.6.** The PCR-amplified mutation cassettes were run on a 2% agarose gel to assess their size with a 1 kb plus DNA ladder (Invitrogen, lanes 1, 3 and 5). The M1 mutation cassette (lane 2) was PCR-amplified with primers JC-P29 and JC-P30 and has an expected size of 1289 bp. The M2 mutation cassette (lane 4) was PCR-amplified with primers JC-P41 and JC-P42 and has an expected size of 1305 bp. The M3 mutation cassette (lane 6) was PCR-amplified with primers JC-P31 and JC-P32 and has an expected size of 1306 bp.

Cells of a Δ*argC proA** strain of *S. enterica* containing the pSLTS plasmid were cultured in LB media containing arabinose to induce expression of λ-red recombinases from genes present on the pSLTS plasmid. The pSLTS plasmid contains a gene for ampicillin resistance and has a temperature sensitive origin of replication that allows replication of the plasmid at a temperature of 30 °C, but not at 37 °C. I mixed each mutation cassette with electrocompetent cells of a Δ*argC proA* strain* of *S. enterica*

+pSLTS (JC517) that were expressing λ-red recombinase and transformed them using electroporation. Cells were spread on plates containing LB/amp/chl and grown overnight at 30 °C to select for those that have integrated the mutation cassette containing the selection marker with a gene for chloramphenicol resistance (Fig 3.7). Up to four colonies from mutation cassette transformation were transferred onto fresh plates containing LB/amp/chl and incubated overnight at 30 °C to remove background colonies.

A) B)



C)

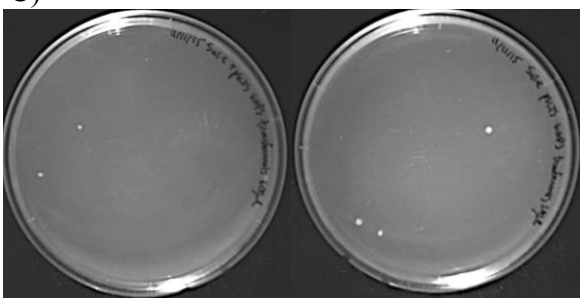
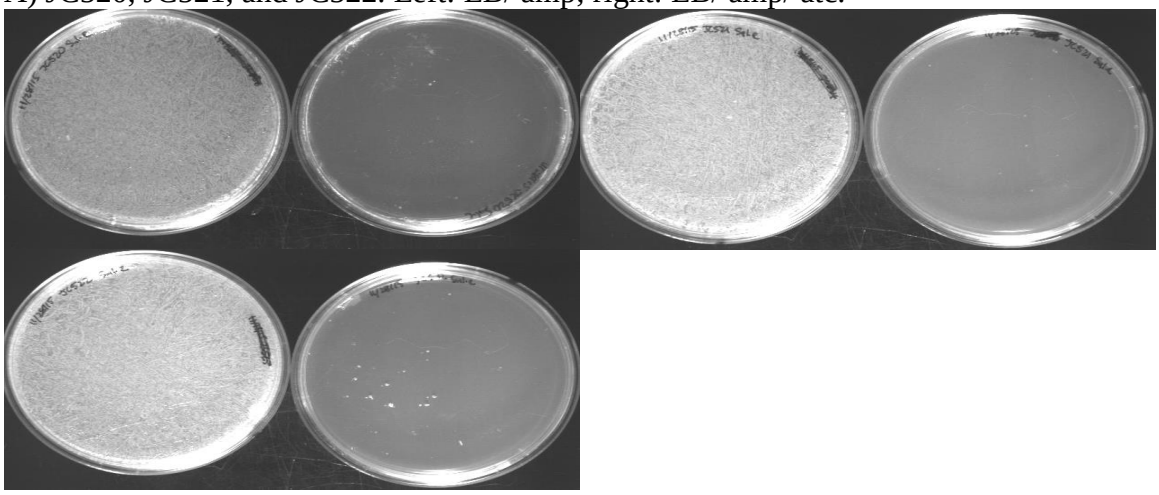


**Fig 3.7.** Overnight growth of JC517 cells on plates containing LB, amp (100 µg/mL) and chl (20 µg/mL) after transformation with the mutation cassette using electroporation (NEB). A) JC517 cells transformed with the M1 mutation cassette. B) JC517 cells transformed with the M2 mutation cassette. C) JC517 cells transformed with the M3 mutation cassette.

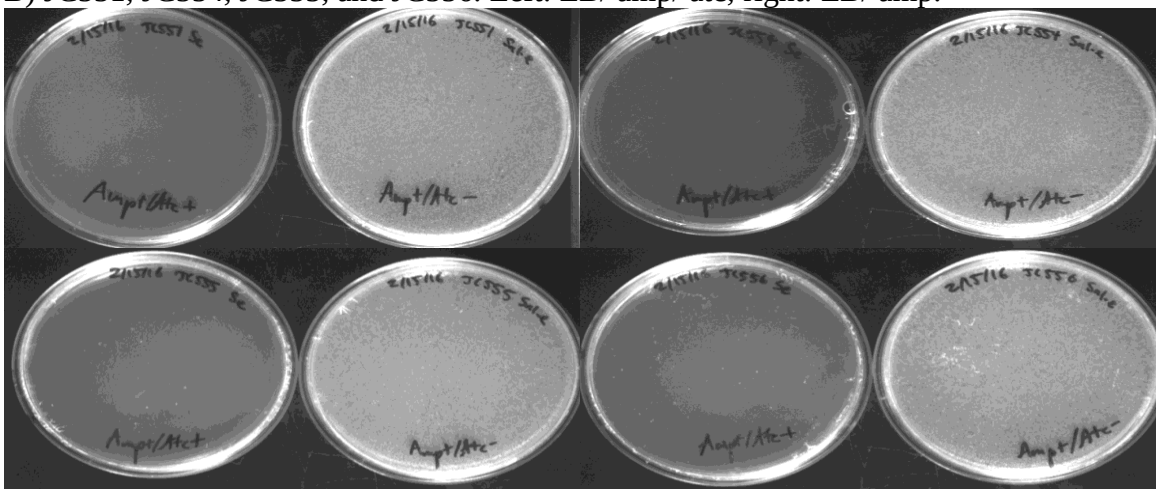
A single colony from each isolation streak was suspended in sterile PBS and spread onto plates containing LB/amp or LB/amp with anhydrotetracycline (atc). In addition, these suspensions were used to inoculate 2mL LB/amp/chl liquid medium. The cultures were incubated at 30 °C and used to make freezer stocks; JC520-JC522, M1 mutation cassette transformant; JC551, and JC554-JC556, M2 mutation cassette transformant; JC523-JC526, M3 mutation cassette transformant. Anhydrotetracycline

allows expression of the I-Sce1 meganuclease encoded on the pSLTS plasmid. The meganuclease will cut at a site within the selection marker, forcing the double-strand DNA repair mechanism to use homologous regions on the upstream and downstream fragments (HR3) to repair the damage thereby eliminating the selection marker in the process (Fig 3.1). The reduction in the number of colonies on plates containing LB/atc suggests that the expected double-strand breakage occurred (Fig 3.8). Five colonies from each plate containing LB/atc were patched onto plates containing either LB/amp or LB/amp/chl. Colonies that can't survive on plates containing LB/amp/chl have lost the selection marker restoring the complete and correct ORF of the *proA** (Fig 3.9).

A) JC520, JC521, and JC522. Left: LB/ amp; right: LB/ amp/ atc.



B) JC551, JC554, JC555, and JC556. Left: LB/ amp/ atc; right: LB/ amp.



C) JC523, JC524, JC525, and JC526. Left: LB/ amp; right: LB/ amp/ atc.

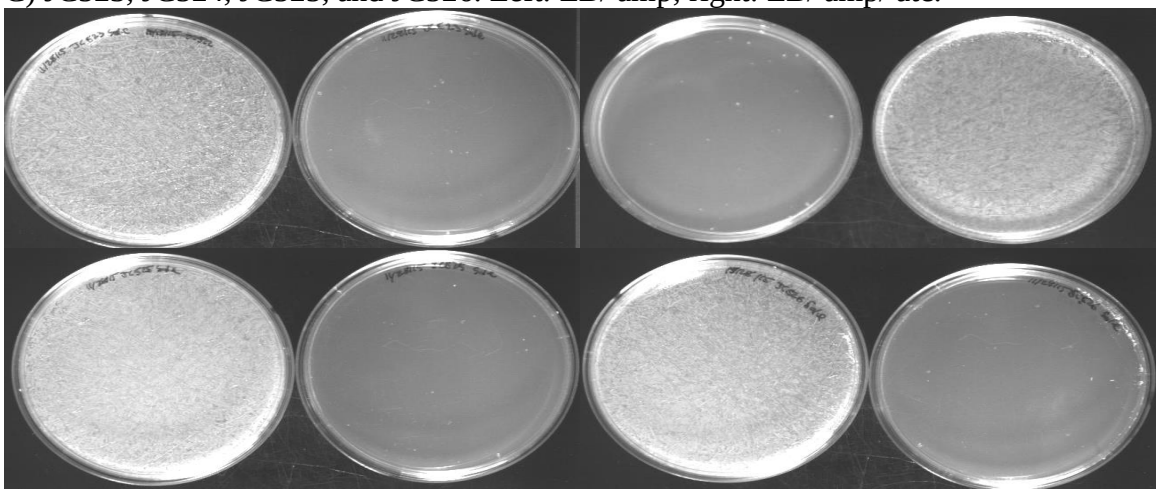
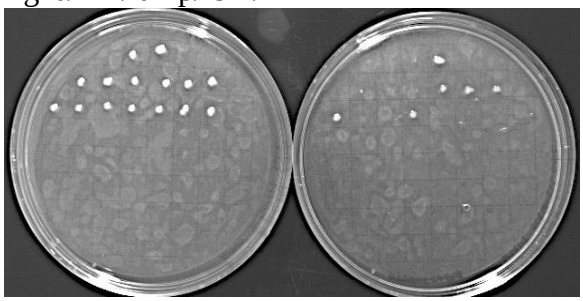
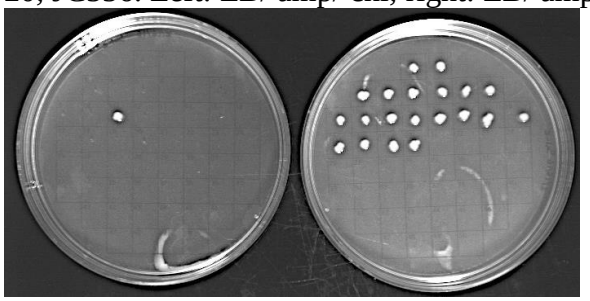


**Fig 3.8.** Overnight growth of isolated JC517 transformants with mutation cassette inserts on plates containing Lb with either amp (100 µg/mL) or amp (100 µg/mL) and atc (100 ng/mL). A) JC517 isolates with M1 mutation cassette insert. B) JC517 isolates with M2 mutation cassette insert. C) JC517 isolates with M3 mutation cassette insert.

A) quadrants 1-5, JC520; quadrants 6-10, JC521; quadrants 11-15, JC522. Left: LB/ amp; right: LB/ amp/ chl.



B) quadrants 1-5, JC551; quadrants 6-10, JC554; quadrants 11-15, JC555; quadrants 16-20, JC556. Left: LB/ amp/ chl; right: LB/ amp.



C) quadrants 1-5, JC523; quadrants 6-10, JC524; quadrants 11-15, JC525; quadrants 16-20, JC526. Left: LB/ amp; right: LB/ amp/ chl.

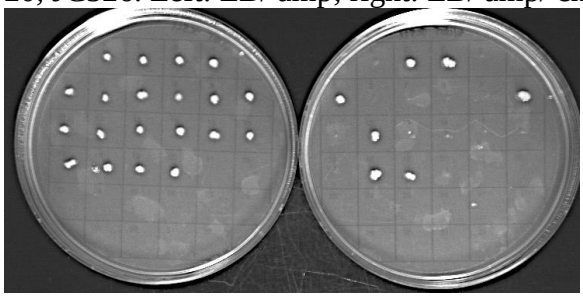


**Fig 3.9.** Overnight growth of individual colonies patched onto plates containing LB with either amp (100 µg/mL) or amp (100 µg/mL) and chl (20 µg/mL) that were taken from LB/amp/atc plates containing overnight growth of JC517 isolates transformed with the mutation cassettes (Fig 3.8). Cells that can't survive on plates containing chl have lost the selection marker that contains the gene for chl resistance. A) Colonies patched from LB/amp/atc plates containing overnight growth of JC517 isolates with M1 mutation cassette insert. B) Colonies patched from LB/amp/atc plates containing overnight growth of JC517 isolates with M2 mutation cassette insert. C) Colonies patched from LB/amp/atc plates containing overnight growth of JC517 isolates with M3 mutation cassette insert

PCR and subsequent Sanger sequencing was done using primers that flank the area where the mutation cassette was expected to integrate to screen colonies that could survive on LB/amp but not LB/amp/chl. The product of this PCR reaction was visualized on a 2% agarose gel with a DNA ladder to assess the size of the amplified fragment prior to sending it for Sanger sequencing (Fig 3.10). Colonies that were confirmed to have the intended mutation were labeled and used to make freezer stocks; JC543, M1; JC559 and JC560, M2; JC544 and JC545, M3.

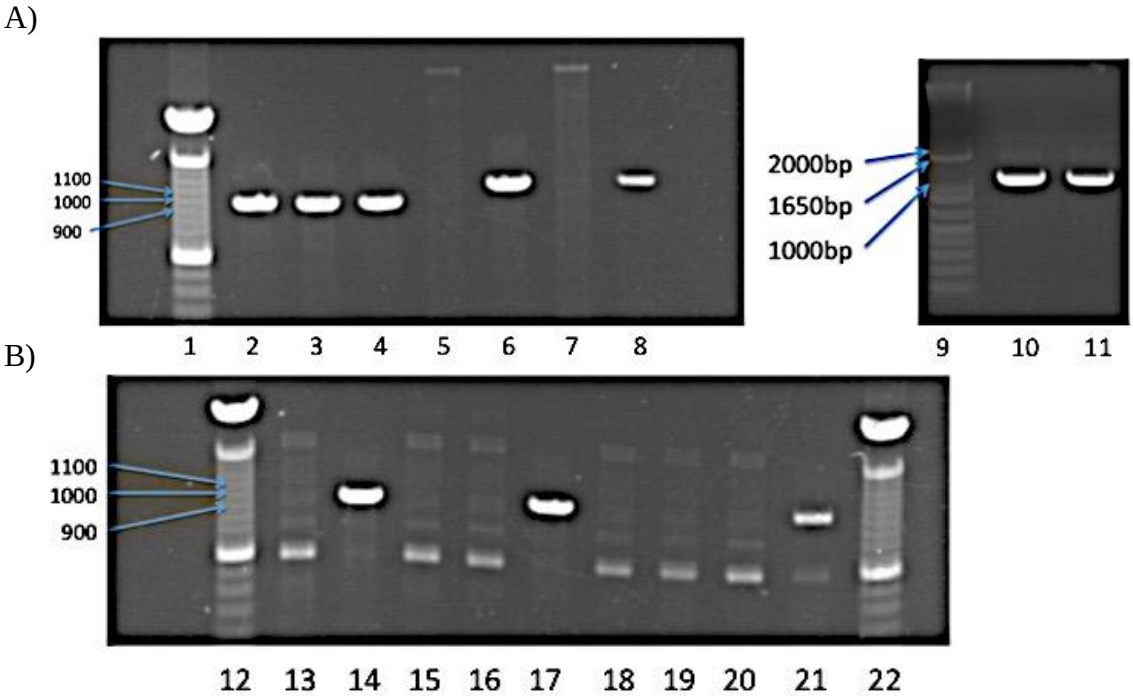



**Fig 3.10.** PCR-amplified segments of the *proBA** operon from cell lysate of colonies on patch plates were run on a 2% agarose gel with either a 1 kb plus DNA ladder (Invitrogen, lane 9) or a 100 bp DNA ladder (Invitrogen, lanes 1, 12 and 22) to assess their size prior to Sanger sequencing. A) The *proBA** operon was PCR-amplified from M1 utants (lanes 2-7) and M2 mutants (lanes 10 and 11) using primers JC-P33 and JC-P34 with an expected size of 1091 bp; positive control was done using JK328 cell lysate (lane 8). B) The *proBA** operon was PCR-amplified from M3 mutants (lanes 13-20) using primers JC-P35 and JC-P36 with an expected size of 972 bp; positive control was done using JK328 cell lysate (lane 21).

Point mutations M1, M2, and M3 increase the growth rate of the Δ*argC proA* S. enterica* strain on M9/glucose

In order to test the effect M1, M2, and M3 have on the growth rate of my *S. enterica* parent strain, all three clones were tested along with the parent strain (JK328) and WT (JK411) strain. The growth curve was done using a single biological replicate in triplicate for each strain in each growth condition. The growth conditions used were M9/glucose (0.2%), M9/glucose +0.4 mM proline, M9/glucose +5.2 mM arginine, and M9/glucose +0.4 mM proline and 5.2 mM arginine. The mean specific growth rate (λ) for each strain in each condition was estimated by non-linear regression using the modified Gompertz equation (24). The results from this experiment show that M1, M2, and M3 all increase the growth rate of the Δ*argC proA** strain on M9/glucose (Fig 3.11). Furthermore, cells of all strains except for those of the wild type version of *S. enterica* were more limited in their ability to synthesize arginine than proline. The growth rate of all strains was similar to the wild type version of *S. enterica* when grown on M9/glucose supplemented with both arginine and proline.

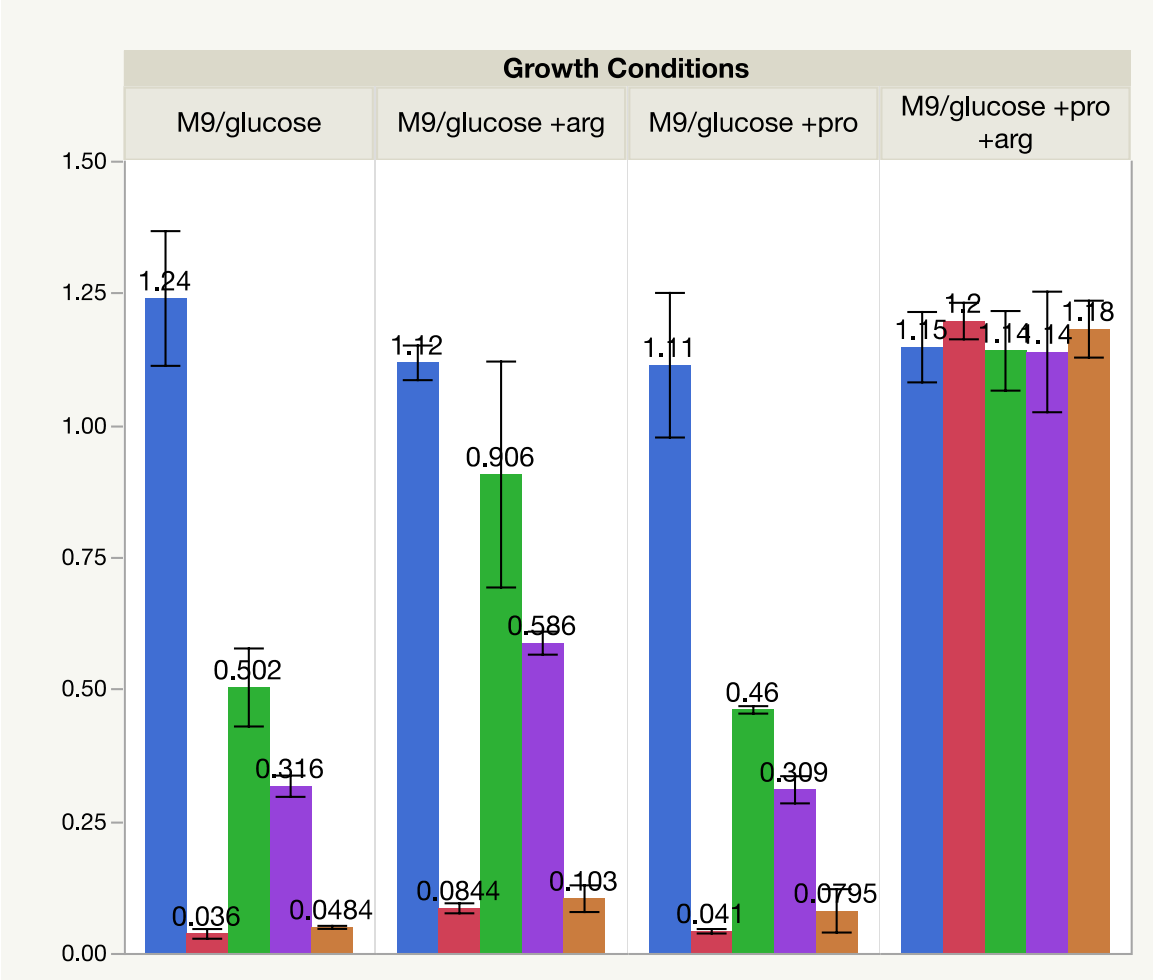



**Fig 3.11.** The mean specific growth rate (λ) of *S. enterica* strains in M9/glucose (0.2%), M9/glucose +0.4 mM proline, M9/glucose +5.2 mM arginine, and M9/glucose +0.4 mM proline and 5.2 mM arginine. Blue bars represent the mean specific growth rate of the wild type *S. enterica (*JK411) in the indicated medium. Red bars represent the mean specific growth rate of the Δ*argC proA** parental strain of *S. enterica* (JK328) in the indicated medium. Green, purple, and orange bars represent the mean specific growth rate of Δ*argC proA* S. enterica* clones containing mutations M1, M2, and M3 respectively in the indicated medium. This was done using 1 biological replicate in triplicate. Error bars represent one s.d. from the mean.

## Discussion

Both promoter mutations, M1 and M2, increased the specific growth rate of the Δ*argC proA** strain of *S. enterica* under all conditions tested (Fig 3.11). Furthermore, M1 seems to be more beneficial than M2, even though it was never observed in faster-growing colonies after growth of the Δ*argC proA** strain of *S. enterica* on M9/glucose.

The synonymous mutation M3 also increased growth rate, albeit very low. In order to be confident that this difference in growth rate is significant, I need more biological replicates. However, these data does suggest that its absence in faster-growing colonies after growth of the Δ*argC proA** strain of *S. enterica* on M9/glucose is due to a lack of its fitness impact rather than a lack of adequate sample size; a slight increases in growth rate will not be caught by eye.

The mean specific growth rate of all strains except for wild type *S. enterica* was lower when cells were grown in M9/glucose supplemented with proline compared to arginine (Fig3.11). These results suggest that the Δ*argC proA** strain of *S. enterica* is more limited in its ability to use ProA* for arginine synthesis than for proline synthesis. However, it is strange that the growth rates for all strain in M9/glucose supplemented with proline matches their growth rates in just M9/glucose. In the future, I will analyze more biological replicates in order to test if these results are reproducible and to account for biological variability. Furthermore, I will want to compare my results to those found in similar experiments involving the Δ*argC proA** strain *E. coli*. This will allow me to compare the fitness impact of M1, M2, and M3 in Δ*argC* strains of *E. coli* and *S. enterica* whose growth rates are limited by the ability to use a similar inefficient bi-functional enzyme (ProA*) for both proline and arginine synthesis.

## Methods

Assembling the M1, M2, and M3 mutation template plasmids and PCR amplification of the mutation cassettes

A gBLOCK of the upstream and downstream fragments for the M1 and M3 mutation template plasmid was ordered from IDT. The upstream and downstream

fragments for the M2 mutation template plasmid were amplified from a Δ*argcC proA** strain of *S. enterica* containing M2 (JC67) using primers that add overhangs overlapping the CAT selection marker and pUC19 backbone. To do this, an aliquot of a freezer stock of JC67 was streaked onto plates containing LB/ strep (50 μg/mL) and incubated overnight at 37 °C. A single colony was suspended in 50 μL sterile $H_2O$ and 2 μL of the suspension were used in a 100 μL PCR reaction with 50 μL OneTaq HS 2x MM, 2 μL of each 10 μM forward and reverse primers, and 44 μL sterile $H_2O$. Three μL of each PCR reaction mixttures were run on a gel in order to assess the size with a 1 kb plus DNA ladder (Invitrogen). The remaining PCR reaction mixtures were treated with an overnight dpn-1 digest, run on a 2% agarose gel, and extracted using the GeneJET Gel Extraction Kit (ThermoScientific) to remove any template still present that could interfere with Gibson assembly of the mutation template plasmid.

Gibson assembly of the three mutation template plasmids was done using the Gibson assembly 2x mastermix (NEB). The product of my Gibson assembly reactions were subsequently used to transform chemical competent *E. coli* Mach1 cells using the heat shock method (NEB). To do this, ~50ng of the mutation cassette is added to an aliquot of chemical competent *E. coli* Mach1 cells and incubated on ice for 30 minutes. Then, the tube containing the cells is submerged in a hot-water bath set to 42 °C for exactly 30 seconds before being placed back on the ice for another 5 minutes. Cells were allowed to recover in 1 mL SOC recovery broth for 1 hour at 37 °C with shacking. Lastly, cell were spread on plates containing LB/amp (100 μg/mL)/chl (20 μg/mL) and incubated overnight at 37 °C to select for those that have taken up the mutation template plasmid.

Primers that anneal to the pUC19 backbone of the mutation template plasmid and flank the region containing the mutation cassette were used to screen *E. coli* Mach1 cells that survived overnight growth on plates containing LB/amp (100 μg/mL)/chl (20 μg/mL). To do this, individual colonies were suspended in 50 μL sterile $H_2O$. Then, 2 μL of the suspension were used in a 30 μL PCR reaction with 15 μL OneTaq HS 2x MM, 0.6 μL of each 10 μM forward and reverse primers, and 11.8 μL sterile $H_2O$. The product of each PCR reaction was run on a gel in order to assess the size with a 1 kb plus DNA ladder (Invitrogen) before sending it for Sanger sequencing (Macrogen). Colony suspensions were stored at 4 °C until sequencing results confirming the correct assembly were received. A single colony suspension containing *E. coli* Mach1 cells determined to have the correct assembly for each mutation template plasmid transformation was used to inoculate 2 mL LB/amp (100 μg/mL)/chl (20 μg/mL) liquid medium, incubated at 37 °C overnight, and harvested for plasmid prep (Qiagen). In addition, this suspension was also used to inoculate 1 mL LB/amp (100 μg/mL)/chl (20 μg/mL) liquid medium, incubated at 37 °C until mid-log, and used to prepare freezer stocks (10% DMSO (v/v)).

Primers were designed to amplify mutation cassettes from the mutant template plasmids using PCR. First, 2 μL of the eluted plasmid were used in a 100 μL PCR reaction with 50 μL OneTaq HS 2x MM, 2 μL of each 10 μM forward and reverse primers, and 44 μL sterile $H_2O$. The product of each PCR reaction was run on a gel in order to assess the size with a 1 kb plus DNA ladder (Invitrogen). The amplified mutation cassettes were then treated with an overnight dpn-1 digest, run on a 2% agarose gel, and extracted using the GeneJET Gel Extraction Kit (ThermoScientific) to remove template still present in the PCR reaction mixtures.

Transformation of a Δ*argC proA** strain of *S. enterica* with the mutation cassette and removal of the selection marker

A Δ*argC* E383A ProA strain of *S. enterica* with the pSLTS plasmid already introduced was obtained from Dr. Juhan Kim. An aliquot of the freezer stock was used to inoculate 2 mL LB/amp (100 µg/mL), grown until mid-log at 30 °C, and used to make a freezer stock of my own (10% DMSO (v/v)). An aliquot of the freezer stock was streaked onto plates containing LB/amp (100 µg/mL) and incubated overnight at 30 °C with shaking. A single colony was used to inoculate 5 mL LB/amp (100 µg/mL) and incubated at 30 °C with shaking. At mid-log, 1 mL of the culture was used to inoculate a 100 mL LB/amp (100 µg/mL) and incubated at 30 °C. After 1 hour, L-arabinose was added to a final concentration of 2 mM and incubated for ~4 hours until the culture reached an $OD_{600}$ of .95. Twice the recommended amount of arabinose and a longer growth period after the addition of L-arabinose is needed when working with *S. enterica.* Two 50 mL aliquots of the culture were harvested via centrifugation at 4,500 xg for 10 min at 4 °C and washed with 25 mL ice-cold 10% glycerol via vortex. Aliquots were combined, harvested again via centrifugation, and wash with 25 mL ice-cold 10% glycerol via vortex. This was done once more with 12.5 mL ice-cold 10% glycerol before suspending the cell pellet in ~500 µL residual glycerol. Aliquots of the suspended cells were used immediately to introduce the mutation cassette; electrocompetent *S. enterica* cells do not work well when stored for later use.

The eluted mutation cassettes for M1, M2, and M3 were quantified using Qubit dsDNA BR. For each mutation cassette, about 50 µg of the eluted cassette was added to the E-comp cells prepared as described in the last section. Cells were transformed by

electroporation, suspended in 1 mL SOC recovery broth, and incubated for ~5 hours at 30 °C. Cells were spread onto plates containing LB/amp (100 μg/mL)/chl (20 μg/mL) and incubated at 30 °C overnight. Up to Four individual colonies were isolation streaked onto fresh plates containing LB/amp (100 μg/mL)/chl (20 μg/mL) and incubated at 30 °C overnight in order to remove background growth. A single colony from each of the 4 isolation streaks was suspended in 500 μl sterile PBS; 100 μL was spread onto plates containing LB/amp (100 μg/mL)/atc (100 ng/mL) and LB/amp (100 μg/mL) using sterile glass beads. Two μl of the PBS suspension were used to inoculate 2 mL LB/amp (100 μg/mL)/chl (20 μg/mL), incubated at 30 °C until mid-log, and used to make freezer stocks (10% DMSO (v/v)). Five Colonies from each of the plates containing LB/amp (100 μg/mL)/atc (100 ng/mL) were patched onto plates containing LB/amp (100 μg/mL) or LB/amp (100 μg/mL)/chl (20 μg/mL) and incubated at 30 °C overnight. Colonies that were able to grow on plates containing LB/amp (100 μg/mL), but not LB/amp (100 μg/mL)/chl (20 μg/mL) were picked from the LB/amp (100 μg/mL) patch plate and suspended in 50 μL sterile $H_2O$.

PCR and subsequent Sanger sequencing of the amplified fragment was done using primers that flank the area where the mutation cassette was introduced to screen colonies that could not survive on plates containing LB/amp/chl for a loss of the selection marker and the desired recombination. Two μL of the cell suspension (from the last section) were used in a 30 μL PCR reaction with 15 μL OneTaq HS 2x MM, 0.6 μL of each 10 μM forward and reverse primers, and 11.8 μL sterile $H_2O$ for both primer sets. Three μL of each PCR reaction were run on a 2% agarose gel in order to assess the size of the amplified fragment with a DNA ladder. The remainder of the PCR reactions were run on

a 2% agarose gel and extracted using the GeneJET Gel Extraction Kit or simply purified using the GeneJET PCR Purification kit (ThermoScientific) and eluted in 30 µL EB. Then, 100-200 ng of the eluted DNA was sent with each primer that was used to amplify it for Sanger sequencing (Macrogen). The remaining cell suspensions were stored at 4 °C until the sequencing results were returned. Once one colony was confirmed by Sanger sequencing to have the correct mutation inserted, the cell suspension was used to inoculate 2 mL LB/strep (50 µg/mL), incubated at 37 °C (to removed pSLTS plasmid) until mid-log, and used to make freezer stock (10% DMSO (v/v)).

Setting up the growth curve experiment and analyzing the data

The Growth curve for this experiment was run on a Varioskan (Thermo) plate reader using the SkanIt software version 2.2. Aliquots of freezer stocks were streaked onto M9/glucose plates and grown overnight at 37 °C. Single colonies were used to inoculate 2 mL LB and grown to early log-phase at 37 °C. Cultures were harvest via centrifugation at 4,500 xg for 8 min at room temperature, washed twice with 500 µL sterile PBS, and re-suspended in 500 µL sterile PBS. The cells were diluted to an $OD_{600}$ of 0.01 in M9/glucose and 10 µl aliquots were used to inoculate their corresponding wells containing 90 µL M9/glucose (2%) to give an initial $OD_{600}$ of 0.001. The plates were incubated in a Varioskan (Thermo) plate reader at 37 °C with shaking every 5 minutes. The absorbance at 600, 900 and 975 nm was measured every 20 minutes for up to 500 hours. The baseline absorbance for each well (the average over several smoothed data points before growth) was subtracted from each point of the smoothed growth curve. Growth parameters (maximum specific growth, $\mu_{max}$; lag time, $\lambda$; maximum growth, $A_{max}$) were estimated by non-linear regression using the modified Gompertz equation

(24). Non-linear least-squares regression was performed in Excel using the Solver feature.

In order to make the four growth conditions used during my growth curve experiment, proline or arginine was added to M9/glucose for a starting concentration of.444 mM or 5.777 mM respectively. Consequently, once 10 µL aliquots containing cells is added to wells containing 90 µL of the supplemented medium, the final concentrations of proline and arginine will be .4 mM and 5.2 mM respectively.

## Chapter 4

## Determination of *proA** copy number of population samples after long-term adaptation of the Δ*argC proA* S. enterica* parental strain in M9/glucose liquid medium

None of the 100 faster-growing colonies obtained after growth of the Δ*argC proA** strain of *S. enterica* on M9/glucose were determined to have amplification of the *proBA** operon. This is in contrast to a similar experiment involving a Δ*argC proA** strain of *E. coli* by Dr. Jamie Kershner in which faster-growing mutants with varying sizes and degrees of amplification of the *proBA** operon were observed (Jamie et al., 2015). The faster-growing colonies analyzed during both these experiments were only growing for ~30 generations, but perhaps more time is necessary to see amplification in *S. enterica*. This chapter describes a long term adaptation experiment meant to exhaust the question as to whether allowing more time to adapt would result in amplification of the *proBA** operon and its surrounding regions. To do this, four cell lineages of the Δ*argC proA** parental strain of *S. enterica* (JK328 and JK329) were grown in 5 mL M9/glucose until mid-log, at which point some of the culture was used to inoculate fresh 5 mL M9/glucose and grown to mid-log again. This was done for ~260 generations over the time period of a month. The constructed "wild type" version of *S. enterica* capable of independently producing histidine, JK411, was used as my control.

All of the faster-growing colonies obtained after growth of the Δ*argC proA** strain of *S. enterica* on plates containing M9/glucose were found to have M2, and I do not expect growth in M9/glucose liquid medium to be any different. However, amplification

does happen in the Δ*argC proA** strain of *E. coli* containing M2 when adapted further in M9/glucose medium (Jamie et al., 2015). Therefore, maybe the same thing can happen in a Δ*argC proA** strain of *S. enterica* containing M2 if allowed to adapt further in M9/glucose medium.

## Results

Long-term adaptation suggests three marked levels improvement in growth rate

The change in $OD_{600}$ was used to calculate generations passed during each culture and was divided by the time of incubation to determine generations passed per day. These values were plotted against the number of generations passed since the start of the long-term adaptation in order to get a qualitative assessment of their change in growth rate (Fig 4.1). Interestingly, all four genetic replicates seem to follow the same pattern of increasing growth rate with 3 marked levels of improvement. Furthermore, there appears to be one marked improvement in the growth rate of the control.

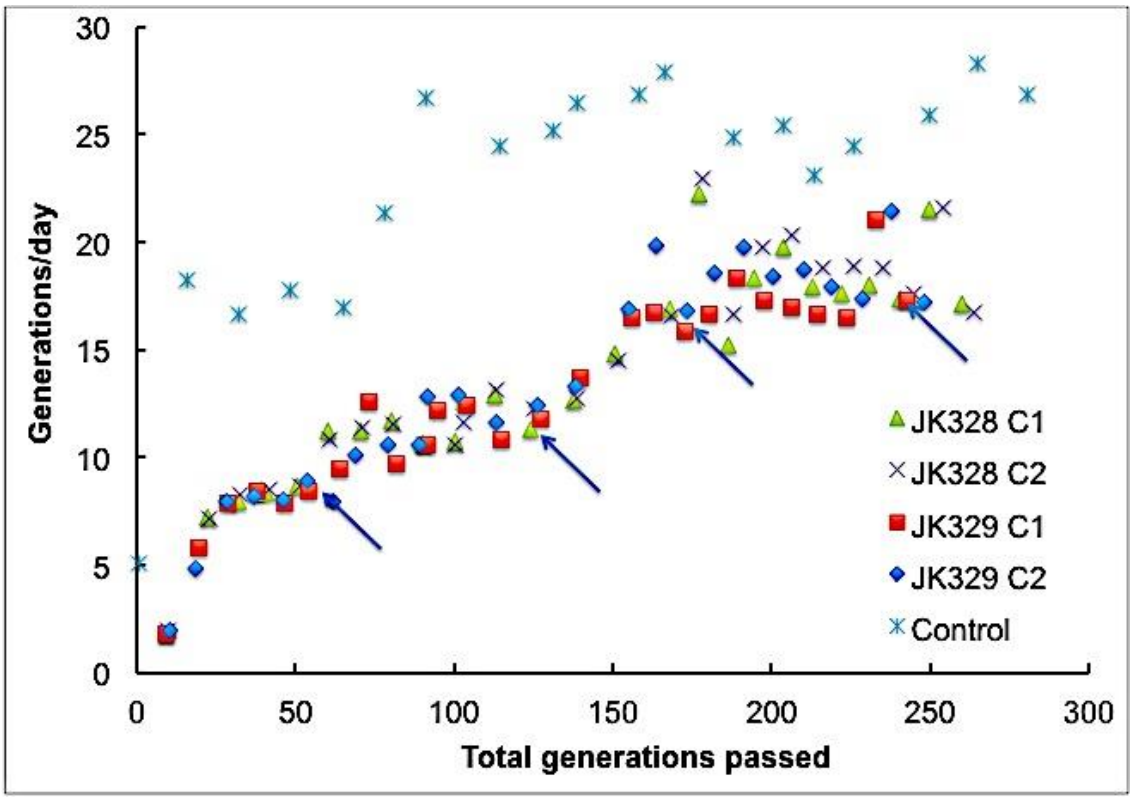



**Fig 4.1.** Shows changes in growth rate as a function of generations per day during adaptation of four *S. enterica* parental strain lineages (JK328 and JK329) and one WT control (JK411). Growth rate was measured as generations per day and plotted over the total generations passed since starting adaptation in M9/glucose. Arrows show time points when genomic DNA of the population was analyzed by qPCR to determine *proA** copy number.

Screening the genomic DNA of the population across several time points for increased *proA** copy number using qPCR reveals no amplification of the *proBA** operon

In order to check if amplification of the *proBA** operon has occurred during the course of my long-term adaptation experiment, individual samples from four different time point around ~60 generations apart were selected for qPCR analysis (Table 4.1). Frozen cell pellets were available for the latter 3 time points, but not for my earliest time point. For this time point, a cell pellet was prepared from an aliquot of the freezer stock. Samples were screened with the best primer set for each target gene on the same plate in order to eliminate the need for inter-run calibrators (Table 2.2). The first and last

dilutions of the parental strain JK328 genomic DNA standards were used as the reference for determining *proA** copy number. Samples were screened without replicates. Primer efficiencies and average Ct values for each reaction were exported to Excel for analysis with qBASE (11). The mean CNRQ of *proA** determined for all samples was 1.05 with a s.d. of 0.14, no CNRQ values fell above or below 2 standard deviations of the mean.

## Discussion

No increase in *proA** copy number was observed for any of my parental lineages during the time points selected for qPCR analysis. The fact that there was one marked improvement in the growth rate of the control is problematic, but can be partially explained by how the strain was constructed. The increase in growth rate suggests that there is adaptation occurring to M9/glucose which may interfere or confound the actual growth rate effects of genetic changes thought to be compensating solely for the inefficient bi-functional enzyme ProA*. However, this strain of *S. enterica* is normally a histidine auxotroph due to a nonsense mutation in the coding region of *hisG*, a gene required for the independent production of histidine. In order to allow growth of this strain on M9/glucose, Dr. Juhan Kim reversed the mutation using the pSLTS method (14). Consequently, it is not unlikely that adaptation would occur in order to re-optimize the pathway necessary for histidine production. More importantly, this emphasizes the use of genome editing techniques that can allow us to study the effect of genetic changes that are observed, like M2, in the absence of other genetic changes that could have occurred.

In terms of my parental strain, the three marked improvements in growth rate suggest that alternative ways to improve fitness have been selected for. In the future, I hope to investigate what genetic changes occurred and how these changes improve fitness of the Δ*argC proA** strain of *S. enterica* in M9/glucose. The absence of amplification during my long-term adaptation experiment may be due to difference in the genetic context between the two bacterial species. For example, it may be that the absence of an IS5 transposable element in *S. enterica* that is consistently observed at the 3' end of the amplified region in *E. coli* is making amplification of the region less frequent (Jamie et al., 2015). Alternatively, amplification of the *proBA* operon and surrounding regions may have serious fitness costs in *S. enterica* compared to similar amplifications in *E. coli* due to differences in the surrounding genes (8). The later was explored in preliminary experiments described in Chapter 5.

## Methods

Setting up and running the long-term adaptation experiment

Aliquots of freezer stocks of the JK328 and JK329 parental strains were streaked onto plates containing LB/strep (50 μg/mL) and incubated overnight at 37 °C. Two replicates from each plate (4 genetic replicates) were used to inoculate 2 mL LB/strep (50 μg/mL) and incubated at 37 °C. Cultures were harvested at mid-log by centrifugation at 10,000 xg for 1.5 min at room temperature, washed twice with 1 mL ice-cold sterile PBS, and suspended in 1 mL ice-cold sterile PBS. The suspensions were used to inoculate 5 mL M9/glucose (0.2%)/strep (50 μg/mL) to an initial $OD_{600}$ of 0.001 and incubated at 37 °C. At mid-log, the $OD_{600}$ was measured for each culture and the change in $OD_{600}$ was

used to calculate the generations passed during growth of the culture. Then, the amount of generations passed was divided by the amount of days passed to calculate generations per day. Some of the culture was used to inoculate fresh 5 mL M9/glucose (0.2%)/strep (50 μg/mL) medium to an initial $OD_{600}$ of 0.001 and incubated at 37 °C. 1 mL of each culture was used to make freezer stock (10% DMSO (v/v)) while another 1 mL was harvested via centrifugation at 10,000 xg for 1.5 min at room temperature, drained of supernatant, and stored at -20 °C.

Four cell lineages of the Δ*argC proA** parental strain of *S. enterica* and one cell lineage of my control (JK411, *+hisG*) were followed in this way for ~250 generations or 27 cultures over the time period of a month. When cultures started to grow faster, a lower initial $OD_{600}$ of 0.0001 or 0.00001 was used or more transfers were done per day (2 instead of 1). This is needed to prevent the growth of cultures into stationary phase, which interferes with accurate gen/day calculations. When a culture overgrows, a new culture can be started from the freezer stock of the previous culture, but requires the use of sterile DMSO or antibiotics to prevent contamination of cultures. For my long-term experiment, streptomycin was supplemented in all cultures to select against other bacterial species.

Screening genomic DNA of population samples for increased *proA** copy number

The same primers used during the qPCR in Chapter 2 were used for this experiment (Table 2.2). Cycling conditions and settings for the 7500 Fast Real-Time PCR System was also maintained. A re-run of the standard curve for all primer sets using 4 1:10 serial dilutions of JK328 genomic DNA was done in order to generate new and reliable primer efficiencies. However, thresholds were manually set to those previously

determined via auto-threshold during my adaptation experiment (Table 2.2). Genomic DNA was prepped from frozen cell pellets across 4 different time points (gen ~50, ~125, ~175, and ~250) for each parental cell lineage using the PureLink Genomic DNA Mini Kit (Invitrogen); eluted in 50 µL EB. When a cell pellet does not exist for a cell culture (gen ~50), an aliquot of the freezer stock corresponding to that time point can be used to inoculate 2 mL M9/glucose (0.2%)/strep (50 µg/mL), grown at 37 °C until mid-log, and pelleted by centrifugation at 10,000 xg for 1.5 min at room temperature.

Parental lineage samples were screened with the best primer set for each target gene on the same plate in order to eliminate the need of inter-run calibrators. Reaction mixtures contained 2 µL template, 0.4 µL each of the 10 µM forward and reverse primers, 7.2 µL sterile $H_2O$, and 10 µL 2X SYBR green MM (Applied Biosystems). The first and last dilutions of the parental strain JK328 genomic DNA standards were used as the reference for determining the CNRQ of *proA**. The average efficiencies for each primer set and the average Ct values for each reaction were exported into Excel for analysis with qBase in order to generate the copy number relative quantity (CNRQ) of *proA** in my population samples (11). The genomic DNA of population samples were initially screened without replicates, and those that showed a CNRQ value greater than 1.5 or less than 0.5 were streaked onto M9/glucose and incubated overnight at 37 °C in order to screen single colonies and identify amplification mutants within the population.

## Chapter 5

## Preliminary investigation of the effects amplification of the *proBA** operon has on surrounding gene expression in a Δ*argC proA** strain of *E. coli*

Overexpression of genes within an amplified region could have a detrimental effect if it interferes with highly constrained cellular systems and hence be selectively unfavorable (8). However, if cells are limited by the ability of an enzyme to sufficiently produce metabolites or clear toxins, amplification could increase fitness by increasing gene dosage (17). Consequently, selection could favor amplification when its benefit outweighs the cost. For example, various amplifications of the *proBA** operon were identified in faster-growing colonies after growth of a Δ*argC proA** strain of *Escherichia coli* BW25113 on M9/glucose. The segmental amplifications were determined to be very large (~30 kb) and to include genes upstream and downstream of *proBA** (Jamie et al., 2015). In contrast, no amplification of the *proBA** operon was found in faster-growing colonies after growth of the Δ*argC proA** *S. enterica* strain in M9/glucose for ~250 generations (Chapter 4). In addition, the genetic context of the *proBA** operon is different between the two bacterial species, especially the downstream region (Fig 2.4). It is unknown whether these large amplifications in *E. coli* are causing overexpression of genes within the amplified segment. If they are, then the genetic context becomes important for determining the cost of amplification and the potential of certain evolutionary trajectories when the growth of cells are limited by an inefficient bi-

functional enzyme. However, if some homeostatic mechanism keeps protein expression at a constant level, then this would suggest that the genetic context may not be important.

In this chapter, I describe preliminary experiments testing whether expression of genes within the amplified region increases as gene copy number increases or if homeostatic mechanisms keep protein expression at a constant level. To do this, I selected two amplification mutants identified after growth of a Δ*argC proA** strain of *E. coli* on M9/glucose, JKAD 1-17-1 T4 (A1) and JKAD 3-4-2 t4 (A3), with varying amounts of amplification determined previously by qPCR analysis (60-80 and 13-15 copies of *proA** respectively). Aliquots of cells from a freezer stock of A1 and A3 were used to inoculate M9/glucose medium and cultured until an $OD_{600}$ of 0.2, at which point they were harvested and used to make genomic DNA or prepare tryptic digests for mass spectroscopy analysis. *proA** copy number was determined by qPCR analysis of the genomic DNA and compared to fold-changes in relative protein levels determined by label-free mass spectrometry of tryptic digests. Label-free mass spectrometry has been gaining attention as a way to quantitatively compare protein levels as technological advances allow greater resolution of whole cell peptide spectra and more databases to search against. Viable methods for prepping samples for label-free mass spectrometry and analysis of spectral counts are documented in the literature (20, 21, 22). Alternatively, mass spectrometry using spiked standards is more accurate, but these methods are time-consuming and expensive.

## Results

Mass spectrometry analysis results in various coverage of the proteins encoded in genes within the amplified segments of JKAD 1-17-1 t4 and JKAD 3-4-2 t4

I examined the effect amplification of the *proBA** operon had on surrounding gene expression by label-free mass spectrometry analysis of tryptic digests prepared from cell extracts of JKAD 1-17-1 t4 and JKAD 3-4-2 t4. I prepared the trypic digests from cell extracts and Dr. Christopher Ebmeier analyzed them on an LTQ-Orbitrap mass spectrometer (Thermo Scientific). Dr. Ebmeier searched identified spectra against an *E. coli* database using Mascot software (Matrix Sciences) to identify peaks corresponding to peptides from all known proteins. The two amplification mutants described in this experiment, JKAD 1-17-1 t4 and JKAD 3-4-2 t4, have different sizes of amplification ending at the same downstream element. I was only interested in genes that were included in the amplified region of both amplification mutants (Fig 5.1). More than 22,000 unique peptides were detected that could be unambiguously assigned to various proteins, 25 unique peptides that could be unambiguously assigned to ProA (73.4% coverage), and 33 unique peptides that could be unambiguously assigned to ProB (87.5% coverage). There were no unique peptides identified that could be assigned to proteins expressed from genes downstream of the *proBA** operon. For proteins that are expressed from genes upstream of the *proBA** operon, 16 unique peptides could be unambiguously assigned to Crl (72.18% coverage), 35 unique peptides could be unambiguously assigned to FrsA (89.6% coverage), 9 unique peptides could be unambiguously assigned to Gpt (58.6% coverage), and 33 unique peptides could be unambiguously assigned to PepD (65.4% coverage).

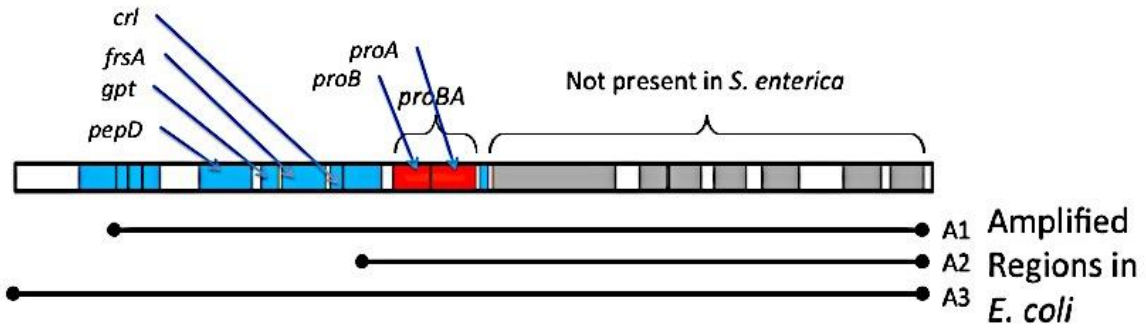



**Fig 5.1.** Genes (in the upstream region) chosen as good candidates for comparing mean normalized spectral counts between *E. coli* amplification mutants JKAD 1-17-1 t4 (A1 - amplified region) or JKAD 3-4-2 t4 (A3 – amplified region) and the wild type strain of *E. coli* BW25113.

Normalized spectral counts can be compared between the wild type strain of *E. coli* and JKAD 1-17-1 t4 or JKAD 3-4-2 t4 strains to determine the fold-change in relative protein levels

Spectral counts for each protein for each sample were normalized to the mean total spectral counts of all samples. The normalized spectral counts determined for a given protein were compared between the wild type strain of *Escherichia coli* BW25113 and JKAD 1-17-1 t4 or JKAD 3-4-2 t4 to determine fold-changes in relative protein levels. This was done for spectral counts of ProA, ProB, Crl, FrsA, Gpt, and PepD. These proteins are expressed from genes surrounding the *proBA* operon that are amplified in both strains (Fig 5.1). JKAD 1-17-1 t4 and JKAD 3-4-2 t4 show an increase in ProA and ProB relative protein levels compared to the WT. Interestingly, there was no increase in the protein levels of Crl, FrsA, Gpt, and PepD in tryptic digests of the JKAD 3-4-2 t4 amplification mutant, but varying increases in these protein levels in tryptic digests of the JKAD 1-17-1 t4 amplification mutant (Fig 5.2).

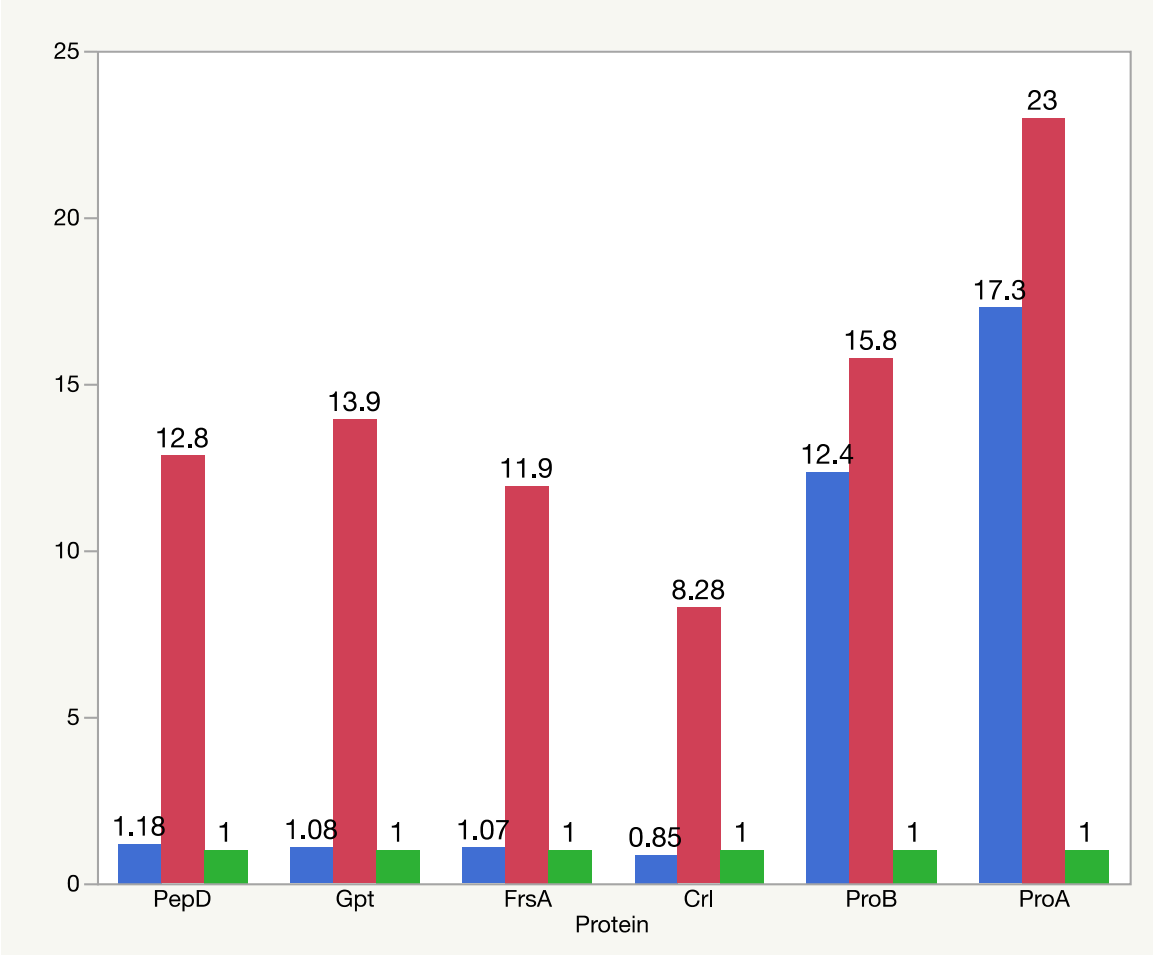



**Fig 5.2.** Fold-changes in relative protein levels by comparing normalized mean spectral counts of each protein of each strain to the wild type *E. coli*. Red, fold-changes in levels of the indicated proteins in JKAD 1-17-1 t4, which was determined by qPCR to have a *proA** copy number of 63. Blue, fold-changes in relative protein levels of JKAD 3-4-2 t4, which was determined by qPCR to have a *proA** copy number of 16. Green, fold-changes in relative protein levels of the wild type strain of *E. coli* BW25113, which has a *proA** copy number of 1.

Determination of *proA** copy number from genomic DNA of JKAD 1-17-1 t4 and JKAD 3-4-2 t4 and comparison to fold-changes in relative protein levels of ProA* and ProB

qPCR was used to determine the *proA** copy number from genomic DNA of Δ*argC proA** *E. coli* amplification mutants JKAD 1-17-1 t4 (A1) and JKAD 3-4-2 t4 (A3) using *icdA* and *gyrB* as internal reference genes. The primers used were obtained from Dr. Jamie Kershner and are listed in the Supplemental Information (Table S.2). The

genomic DNA from the Δ*argC proA* E. coli* parental strain (JK293) was used as the reference for determining the *proA** copy number of the amplification strains. Primer efficiencies and average Ct values for each reaction were exported to Excel for analysis with qBASE (11). The *proA** CNRQ values determined for JKAD 1-17-1 t4 and JKAD 3-4-2 t4 were 63 and 16 respectively. Previously reported *proA** copy numbers for JKAD 1-17-1 t4 and JKAD 3-4-2 t4 by Dr. Jamie Kershner were 60-80 and 13-15 respectively.

*proA** copy number of the Δ*argC proA* E. coli* amplification strains was compared to fold-changes in the relative protein levels of ProA and ProB (Fig 5.3). In order to make an equivalent comparison with the other genes in the amplified region, I would need to do qPCR analysis with primers that target them specifically. The change in *proA** copy number does not directly correlate with the fold-change in relative protein levels of ProA* and ProB in the JKAD 1-17-1 t4 amplification strain. The change in *proA** copy number seems to correlate to the fold-change in relative protein levels of ProA* and ProB in the JKAD 3-4-2 t4 amplification strain, but I need more biological replicates to be confident in my analysis.

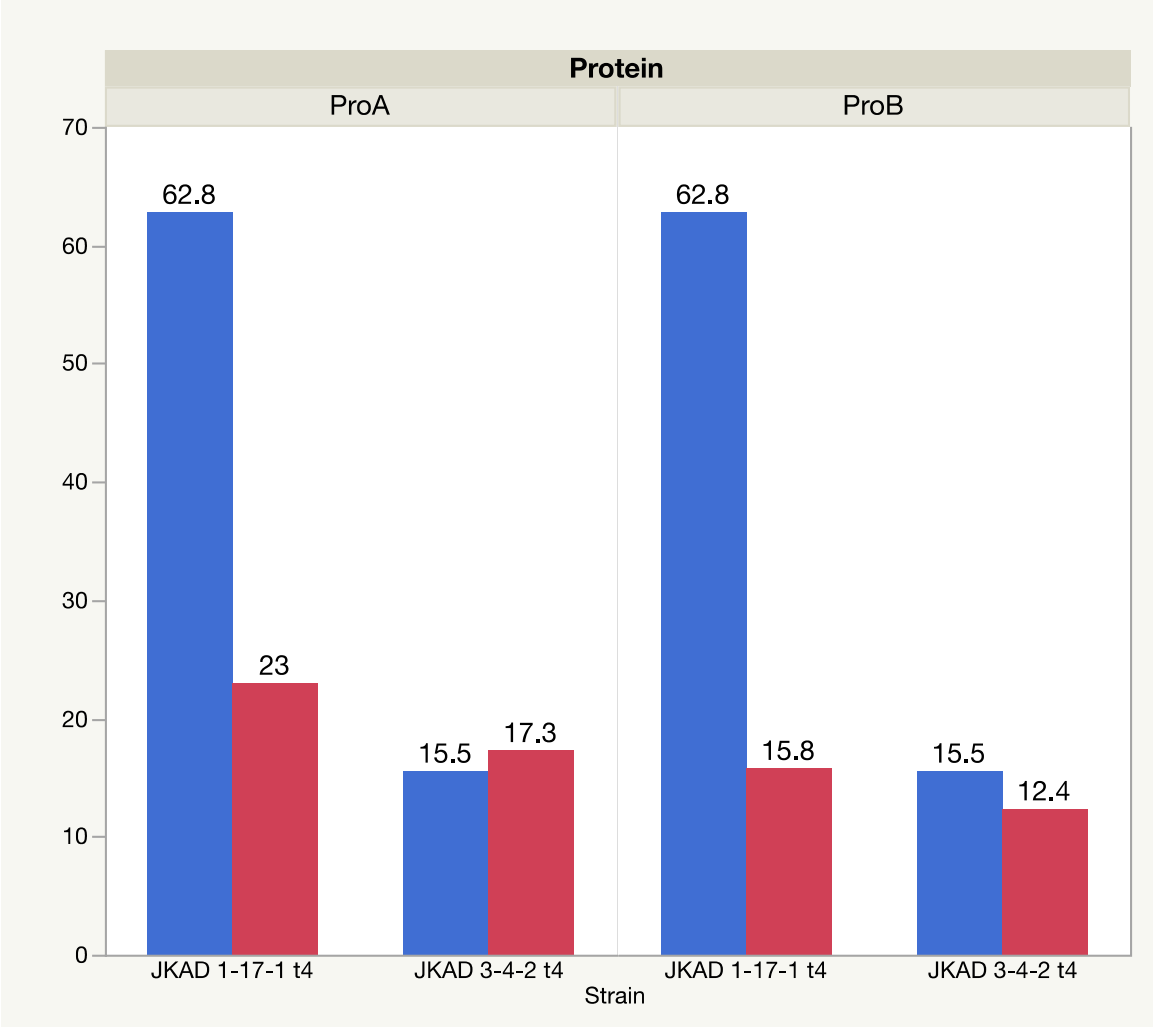



**Fig 5.3.** Comparison of *proA** copy number and fold-change in relative protein levels of ProA and ProB for the two amplification strains, JKAD 1-17-1 t4 and JKAD 3-4-2 t4. Blue, *proA** copy number; red, fold-changes in levels of the indicated proteins.

## Discussion

The experiments described in this chapter explored the effects amplification of the region surrounding the *proBA** operon has on the levels of proteins expressed from genes in the amplified region in a Δ*argC proA** strain of *E. coli*. I determined the fold-change in relative protein levels of my amplification strains by comparison to the wild type strain of *E. coli*, which has an unmodified *argC* and *proA* allele. This was chosen over the Δ*argC proA** parental strain of *E. coli* (JK293) because the *proBA* operon and

surrounding regions can amplify in this strain during growth in M9/glucose. In retrospect, I think such a comparison can be made if qPCR analysis of genomic DNA from JK293 cells cultured in M9/glucose is done prior mass spectrometry analysis on tryptic digests prepared from the same cells to show that no amplification of the *proBA** operon has occurred. The Δ*argC proA** parental strain of *E. coli* (JK293) would allow me to see the effect of amplification alone on ProA* and ProB expression. In addition, I would want to have more biological replicates to be more confident in my quantification of *proA** copy number and fold-changes in relative protein levels.

Interestingly, there seems to be little to no expression of genes between the *proBA** operon and the conserved downstream junction of the amplified regions. This is important considering that this region is different in S. enterica (Fig 5.1). Furthermore, there is a stark difference in the expression of genes upstream of the *proBA** operon between my two amplification mutants. It is unlikely that some homeostatic mechanism is affecting one amplification mutant and not the other. Furthermore, I am not sure that these genes are present in multiple copies; qPCR was only used to determined *proA** copy number. In order to see if these genes are also present in multiple copies, I plan to design primers that target them specifically. In addition, it would be important to repeat the experiment using multiple biological replicates showing the same pattern of expression to be confident that what I see is accurate.

## Methods

Preparing strains for qPCR and label-free mass spectrometry

An aliquot of a freezer stock of JK1 was streaked onto plates containing M9/glucose (0.2%) and incubated overnight at 37 °C. Aliquots of freezer stocks for two

amplification strains, JKAD 1-17-1 t4 (A1) and JKAD 3-4-2 t4 (A3), were streaked onto plates containing M9/glucose (0.2%)/kan (20 μg/mL) and incubated overnight at 37 °C.

A single colony from overnight plates of each strain was suspended in 50 μl sterile $H_2O$. Five μl was used to inoculate 2 mL M9/glucose (0.2%) and incubated at 37 °C until mid-log phase ($OD_{600}$ of 0.5). Kan 20 μg/mL was supplemented in medium used to culture the amplification strains. Two μL of the 2 mL culture was used to inoculate 50 mL M9/glucose (0.2%) and incubated at 37 °C with shaking to an $OD_{600}$ of 0.2, at which point, 4 x 10 mL aliquots of the 50 mL culture were pelleted via centrifugation at 4,500 xg for 10 min at 4 °C and stored at -80 °C prior to preparation for Mass Spectrometry. In addition, 2 ml of the same 50 mL culture were harvested for the WT and amplification strains via centrifugation at 10,000 xg for 2 min at room temperature for genomic DNA purification using the PureLink Genomic DNA Mini Kit (Invitrogen); the purified genomic DNA was eluted in 50 μL EB.

qPCR

The primers used by Dr. Jamie Kershner for screening mutants via qPCR were used for this experiment. Information on these primers can be found in the Supplemental Information (Table S.2). qPCR was performed on a 7500 Fast real-time PCR system. A re-run of the standard curve for all primer sets was done using four 1:10 serial dilutions of the JK1 genomic DNA in triplicate as my standards to generate new and reliable primer efficiencies. However, thresholds were manually set to those previously used by Dr. Kershner. Samples were screened with each primer set for each target gene on the same plate in order to eliminate the need for inter-run calibrators. Reaction mixtures contained 2 μL template, 0.4 μL each of the 10 μM forward and reverse primers, 7.2 μL

sterile $H_2O$, and 10 µL 2x SYBR green MM (Applied Biosystems). The first and last dilutions of the JK1 genomic DNA standards were used as a reference to determine the *proA** copy number of the amplification strains. Primer efficiencies and the average Ct values for each reaction were exported into Excel for analysis with qBase in order to generate the copy number relative quantity of *proA** for each sample (11).

Preparation of tryptic digests for Mass Spectrometry

Frozen cell pellets from 10 mL of cultures cells ($OD_{600}$: 0.2) were suspended in 200 µL 4% SDS, 0.1 M Tris-pH 8.5, 10 mM TCEP, 40 mM 2-chloroacetamide, and sonicated for 1 min using the Branson Sonifier 250 (50% duty, 2.5 output). The samples were boiled for 10 min at 95 °C and then kept at room temperature. A Millipore 30kD MWCO Amicon Ultra-0.5 centrifugal filter device was prepared by washing it with 300 µL of 8M urea in 0.1M Tris-pH 8.5 and spinning it, along with your sample, at 16,000 xg for 5 min at room temperature. Next, 500 µL of 8 M urea in 0.1 M Tris-pH 8.5 and 50 µl of the sample supernatant were added immediately and the device was centrifuged at 16,000 xg for 10 min at room temperature. The filter was then washed three times with 300 µL of 8 M urea in 0.1 M Tris-pH 8.5 and three times with 300 µL of 2 M urea in 0.1 M Tris-pH 8.5. The filter was transferred to a fresh collection tube for sample digestion. Next, 200 µl of 2 M urea in 0.1 M Tris-pH 8.5 was added along with 1 µg lysyl endopeptidase (Wako) at a ratio of approximately 1:50 (protein: protease) and incubated at room temperature for 3 hrs. Then, 1 µg of Trypsin (Promega) at a ratio of approximately 1:50 (protein: protease) was added and the filter was incubated overnight rocking at room temperature. In the morning, the column was centrifuged at 16,000 xg for 10 min at room termperature, 200 µl of 0.5 M NaCl was added to the column, and it

was centrifuged again at 16,000 xg for 10 min at room temperature to elute sample. Samples were cooled with liquid nitrogen and stored immediately at -80 °C.

A Pierce C18 spin column was used to desalt and wash the digested peptides for label-free mass spectrometry. First, the column was centrifuged at 300 xg for 1.5 min at room temperature in order to settle the resin to the bottom of the column. Then the column was equilibrated by washing it twice with 300 µL of 70% ACN and three times with 300 µL of 5% ACN using centrifugation at 300 xg for 1.5 min at room temperature prior to loading the sample. The entire sample was then loaded onto the column, centrifuged at 300 xg for 1.5 min at room temperature, and then re-applied and centrifuged again to ensure optimal binding of peptides to the resin. The column was washed 5 times with 300 µL of 5% ACN before moving the column into a clean, low-protein-binding tube. Fifty µL of 70% ACN was added to the column, centrifuged at 300 xg for 1.5 min at room temperature, and then repeated with another 50 µL of 70% ACN to obtain a total of 100 µl of eluted peptides. The samples were frozen with liquid nitrogen and stored immediately at -80 °C prior to mass spectrometry. The peptide mixture was evaporated to dryness in a Speed-Vac and re-suspended in 0.1% (v/v) formic acid in HPLC grade water. Dr. Christopher Ebmeier then injected aliquots containing 2 µg of peptides onto a C18 UPLC column (Waters nanoAcuity 1.7 µm BEH130, 75 µm x 250 mm). The column was eluted with a gradient from 3% to 33% acetonitrile at 0.3 µL per minute over a period of five hours. Eluted peptides were detected using an LTQ-Orbitrap (Thermo Scientific). Spectra were searched against an *E. coli* database using Mascot software (Matrix Sciences) to identify peaks corresponding to peptides from all known proteins

**Supplemental Information**

| Strain | Genotype | Reference | Notes |
|---|---|---|---|
| JK328 | *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str*. SL1344; *ΔargC*; E382A ProA; +*hisG* | This work | *S. enterica* "parent strain" |
| JK329 | *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str*. SL1344; *ΔargC*; E382A ProA; +*hisG* | This work | *S. enterica* "parent strain" |
| JK411 | *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str*. SL1344; +*hisG* | This work | *S. enterica* "wild type" (+*hisG)* |
| JC67 | *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str*. SL1344; *ΔargC*; E382A ProA; +*hisG; has M2* | This work | A M2 mutant from my adaptation experiment |
| JC518 | *E. coli* MACH1 w/ M1 mutation template plasmid | This work | M1 mutation template plasmid transformant made during my genome editing experiment |
| JC546 | *E. coli* MACH1 w/ M2 mutation template plasmid | This work | M2 mutation template plasmid transformant made during my genome editing experiment |
| JC519 | *E. coli MACH1 w/ M3 mutation template plasmid* | This work | M3 mutation template plasmid transformant made during my genome editing experiment |
| JC517 | *Salmonella enterica* subsp. *enterica serovar Typhimurium str. SL1344; ΔargC; E382A ProA; +hisG; +pSLTS* | This work | Contains the pSLTS plasmid necessary for λ-red recombinase expression. |
| JC520 | *Salmonella enterica* subsp. *enterica serovar Typhimurium str. SL1344; ΔargC; E382A ProA; +hisG; +pSLTS; +M1 mutation cassette* | This work | Has integrated the M1 mutation cassette and selection marker (CAT) |

| | | | |
|---|---|---|---|
| JC521 | *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str*. SL1344; *ΔargC*; E382A ProA; *+hisG; +pSLTS; +M1 mutation cassette* | This work | Has integrated the M1 mutation cassette and selection marker (CAT); genome editing experiment |
| JC522 | *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str*. SL1344; *ΔargC*; E382A ProA; *+hisG; +pSLTS; +M1 mutation cassette* | This work | Has integrated the M1 mutation cassette and selection marker (CAT); genome editing experiment |
| JC551 | *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str*. SL1344; *ΔargC*; E382A ProA; *+hisG; +pSLTS; +M2 mutation cassette* | This work | Has integrated the M1 mutation cassette and selection marker (CAT); genome editing experiment |
| JC554 | *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str*. SL1344; *ΔargC*; E382A ProA; *+hisG; +pSLTS; +M2 mutation cassette* | This work | Has integrated the M2 mutation cassette and selection marker (CAT); genome editing experiment |
| JC555 | *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str*. SL1344; *ΔargC*; E382A ProA; *+hisG; +pSLTS; +M2 mutation cassette* | This work | Has integrated the M2 mutation cassette and selection marker (CAT); genome editing experiment |
| JC556 | *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str*. SL1344; *ΔargC*; E382A ProA; *+hisG; +pSLTS; +M2 mutation cassette* | This work | Has integrated the M2 mutation cassette and selection marker (CAT); genome editing experiment |

| Strain | Genotype | Reference | Notes |
|---|---|---|---|
| JC523 | *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str*. SL1344; *ΔargC*; E382A ProA; *+hisG; +pSLTS; +M3 mutation cassette* | This work | Has integrated the M3 mutation cassette and selection marker (CAT); genome editing experiment |
| JC524 | *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str*. SL1344; *ΔargC*; E382A ProA; *+hisG; +pSLTS; +M3 mutation cassette* | This work | Has integrated the M3 mutation cassette and selection marker (CAT); genome editing experiment |
| JC525 | *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str*. SL1344; *ΔargC*; E382A ProA; *+hisG; +pSLTS; +M3 mutation cassette* | This work | Has integrated the M3 mutation cassette and selection marker (CAT); genome editing experiment |
| JC526 | *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str*. SL1344; *ΔargC*; E382A ProA; *+hisG; +pSLTS; +M3 mutation cassette* | This work | Has integrated the M3 mutation cassette and selection marker (CAT); genome editing experiment |
| JC543 | *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str*. SL1344; *ΔargC*; E382A ProA; *+hisG; M1* | This work | The pSLTS plasmid has been remove M1 has been inserted; genome editing experiment |
| JC559 | *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str*. SL1344; *ΔargC*; E382A ProA; *+hisG; M2* | This work | The pSLTS plasmid has been remove M2 has been inserted; genome editing experiment |
| JC560 | *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str*. SL1344; *ΔargC*; E382A ProA; *+hisG; M2* | This work | The pSLTS plasmid has been remove M2 has been inserted; genome editing experiment |
| JC544 | *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str*. SL1344; *ΔargC*; E382A ProA; *+hisG; M3* | This work | The pSLTS plasmid has been remove M3 has been inserted; genome editing experiment |

| | | | |
|---|---|---|---|
| JC545 | *Salmonella enterica* subsp. *enterica serovar* Typhimurium *str*. SL1344; *ΔargC*; E382A ProA; +*hisG; M3* | This work | The pSLTS plasmid has been remove M3 has been inserted; genome editing experiment |
| JKAD 1-17-1 t4 | *Escherichia coli* BW25113; *ΔargC*; E383A ProA; *proBA* amplification | Jamie et al., 2015 | *E. coli* amplification mutant (60-80 copies) |
| JKAD 3-4-2 t4 | *Escherichia coli* BW25113; *ΔargC*; E383A ProA; *proBA* amplification | Jamie et al., 2015 | *E. coli* amplification mutant (13-15) |
| JK1 | *Escherichia coli* BW25113 | Jamie et al., 2015 | *E. coli* WT |
| JK293 | *Escherichia coli* BW25113; *ΔargC*; E383A ProA | Jamie et al., 2015 | *E. coli* "parent strain" |

**Fig S.1.** List of all strains used in this thesis including their name, genotype, source, and notes

| Primer Name | Direction (Forward or Reverse) | Sequence (5'->3') | Reference | Notes |
|---|---|---|---|---|
| JC-P11 | Forward | GCGTCCGAATGTTACCGTTG | This work | *proA*set1; qPCR |
| JC-P12 | Reverse | TTCCCACCGCGTAGAATCAC | This work | *proA*set1; qPCR |
| JC-P13 | Forward | TGCGCCAGGTCTGTAATCTC | This work | *proA*set2; qPCR |
| JC-P14 | Reverse | CAACGGTAACATTCGGACGC | This work | *proA*set2; qPCR |
| JC-P5 | Forward | CAGGCGGAGCTTATCAAAGA | This work | *proA*set3; qPCR |
| JC-P6 | Reverse | CCTGAAGCTTAGTGCTCATACC | This work | *proA*set3; qPCR |
| JC-P15 | Forward | TCGTTACTATCAGGGCACGC | This work | *icdA*set1; qPCR |
| JC-P16 | Reverse | TCCGGCGTAGATGTCTTCAG | This work | *icdA*set1; qPCR |
| JC-P17 | Forward | CGCCTGTATGAACCTGAACG | This work | *icdA*set2; qPCR |
| JC-P18 | Reverse | GCAATACCGATACCGCCAAC | This work | *icdA*set2; qPCR |
| JC-P7 | Forward | GCCGTCGAGAAAGCCTATAA | This work | *icdA*set3; qPCR |
| JC-P8 | Reverse | CAGACATCCTGGCCGTAAA | This work | *icdA*set3; qPCR |

| | | | | |
|---|---|---|---|---|
| JC-P19 | Forward | TGCCCGTGAAGGTCTGATTG | This work | *gyrB*set1; qPCR |
| JC-P20 | Reverse | TCATCTGCTGTTCTACCGCC | This work | *gyrB*set1; qPCR |
| JC-P21 | Forward | TGCCCGTGAAGGTCTGATTG | This work | *gyrB*set2; qPCR |
| JC-P22 | Reverse | GCTGTTCTACCGCCGATTTC | This work | *gyrB*set2; qPCR |
| JC-P23 | Forward | CGCTGCTGTTGACCTTCTTC | This work | *gyrB*set3; qPCR |
| JC-P24 | Reverse | ACTGATCCATCGCTTCGTCG | This work | *gyrB*set3; qPCR |
| JC-P25 | Forward | CCACTATTGCCAGAGTGC | This work | PCR-amplify *proBA* operon and regulatory region for short-read sequencing (5' region) |
| JC-P28 | Reverse | CAGCAATCGCTTTCAGACG | This work | PCR-amplify *proBA* operon and regulatory region for short-read sequencing (5' region) |
| JC-P2 | Forward | ACCCGAGAGTGTACCCGAG | This work | PCR-amplify *proBA* operon and regulatory region for short-read sequencing (3' region) |
| JC-P3 | Reverse | AACGTGGCAGTTCACTTCTG | This work | PCR-amplify *proBA* operon and regulatory region for short-read sequencing (3' region) |
| JC-P37 | Forward | AGGCGTATCACGAGGCCCTTACGTTCGGTAATGAAAGAT | This work | PCR-amplify upstream region of M2 mutation template plasmid; Chapter 3 |

| Primer Name | Direction (Forward or Reverse) | Sequence (5'->3') | Reference | Notes |
|---|---|---|---|---|
| JC-P38 | Reverse | ACCGCTGCCACTCTTGAGATTAAAGACCAAATGAACATT | This work | PCR-amplify upstream region of M2 mutation template plasmid; Chapter 3 |
| JC-P39 | Forward | GTGGCAGGGCGGGGCGTAAGCCAGTCCTGCTAAAATGTTC | This work | PCR-amplify downstream region of M2 mutation template plasmid; Chapter 3 |
| JC-P40 | Reverse | CTCACATGTTCTTTCCTGCGATGTCCCGCGGCGTGTAACTGCG | This work | PCR-amplify downstream region of M2 mutation template plasmid; Chapter 3 |
| pHA.seq.F | Forward | TATCAGGGTTATTGTCTCATGAGCG | Dr. Juhan Kim (17) | Amplify segment of mutation template plasmid containing the mutation cassette; Chapter 3 |
| pHA.seq.R | Reverse | ACTTGAGCGTCGATTTTTGTGATGC | Dr. Juhan Kim (17) | Amplify segment of mutation template plasmid containing the mutation cassette; Chapter 3 |
| JC-P29 | Forward | TATGTTTAATTGTTGCGGTAA | This work | Amplify M1 mutation cassette; Chapter 3 |
| JC-P30 | Reverse | CCGAGTTTTACGACCA | This work | Amplify M1 mutation cassette; Chapter 3 |
| JC-P41 | Forward | GTTGCGGTAATAAAATGGTCAAATTCCGC | This work | Amplify M2 mutation cassette; Chapter 3 |
| JC-P42 | Reverse | CGATCCGCCTGTTAGCACGC | This work | Amplify M2 mutation cassette; Chapter 3 |
| JC-P31 | Forward | GAGGTCATCAGAATTTG | This work | Amplify M3 mutation cassette; Chapter 3 |
| JC-P32 | Reverse | CTTTAGCAGCAATGC | This work | Amplify M3 mutation cassette; Chapter 3 |
| JC-P33 | Forward | CACTTTCGGCTTTATTACTGGC | This work | Verify M1 and M2 insertion; Chapter 3 |

| | | | | |
|---|---|---|---|---|
| JC-P34 | Reverse | GATTGCTACGCGGATCGG | This work | Verify M1 and M2 insertion; Chapter 3 |
| JC-P35 | Forward | GTGTATGGCGTTGACGATG | This work | Verify M3 insertion; Chapter 3 |
| JC-P36 | Reverse | GTAGAATCACCGCATTACCG | This work | Verify M3 insertion; Chapter 3 |
| 1357 | Forward | CGTGCAGGCGATTGATAA | Jamie et al., 2015 | *E. coli proA* qPCR; Chapter 5 |
| 1359 | Reverse | CTGTTCACGGCACAGTTT | Jamie et al., 2015 | *E. coli proA* qPCR; Chapter 5 |
| 1361 | Forward | CCCGTGGCTGAAAGTTAAA | Jamie et al., 2015 | *E. coli icdA* qPCR; Chapter 5 |
| 1362 | Reverse | CAGGTTCATACAGGCGATAAC | Jamie et al., 2015 | *E. coli icdA* qPCR; Chapter 5 |
| 1359 | Forward | CGTAGATCTGACGGTGAATTT | Jamie et al., 2015) | *E. coli gyrB* qPCR; Chapter 5 |
| 1360 | Reverse | CGTTGGTGTTTCGGTAGTA | Jamie et al., 2015 | *E. coli gyrB* qPCR; Chapter 5 |

**Fig S.2.** List of all primers used in this thesis including their name, direction, sequence, source, and description

## References


1. Arentson BW, Sanyal N, Becker DF. Substrate channeling in proline metabolism. Front Biosci. 2012;17:375-88.

2. Berg OG, Silva PJ. 1997. Codon bias in Escherichia coli: the influence of codon context on mutation and selection. Nucleic Acids Res. 25:1397-1404.

3. Bergthorsson, U., Andersson, D. I., & Roth, J. R. (2007). Ohno's dilemma: Evolution of new genes under continuous selection. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(43), 17004–17009. http://doi.org/10.1073/pnas.0707158104

4. Brown J. E., Clarke P. H. (1970) J. Gen. Microbiol. 64, 329–342 [PubMed]

5. Carlini DB, Stephan W. 2003. In Vivo Introduction of Unpreferred Synonymous Codons Into the Drosophila Adh Gene Results in Reduced Levels of ADH Protein. Genetics 163:239-243

6. Chamary J V, Hurst LD. 2005. Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. Genome Biol. 6:R75

7. Chursov A, Frishman D, Shneider A. 2013. Conservation of mRNA secondary structures may filter out mutations in Escherichia coli evolution. Nucleic Acids Res. 41:7854-7860.

8. Conant GC, Wolfe KH. Turning a hobby into a job: How duplicated genes

find new functions. Nat Rev Genet. 2008;9:938–50. doi:10.1038/nrg2482. [PubMed] [Cross Ref]

9. Copley SD. Enzymes with extra talents: Moonlighting functions and catalytic promiscuity. Curr Opin Chem Biol. 2003;7:265–272. [PubMed]

10. Copley S.D. Toward a systems biology perspective on enzyme evolution. J. Biol. Chem. 2012;287:3–10. doi: 10.1074/jbc.R111.254714. [PMC free article] [PubMed] [Cross Ref]

11. Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol. 8: 2 [PMC free article] [PubMed]

12. Hughes AL. The evolution of functionally novel proteins after gene duplication. Proc R Soc Lond Ser B. 1994;256:119–124. doi: 10.1098/rspb.1994.0058. [PubMed] [Cross Ref]

13. Khanal A, Yu McLoughlin S, Kershner JP, Copley SD. Differential effects of a mutation on the normal and promiscuous activities of orthologs: implications for natural and directed evolution. Mol Biol Evol. 2015;32(1):100-8.

14. Kim, J., Webb, A., Kershner, J., Blaskowski, S., & Copley, S. (2014). A versatile and highly efficient method for scarless genome editing in Escherichia coli and Salmonella enterica. *BMC Biotechnol.,* 84-84. doi:10.1186/1472-6750-14-84

15. Kitagawa M., Ara T., Arifuzzaman M., Ioka-Nakamichi T., Inamoto E., Toyonaga H., Mori H. (2005) DNA Res. 12, 291–299 [PubMed]

16. McClelland, M., Sanderson, K., Spieth, J., Clifton, S., & Latreille, P. (2001). Complete genome sequence of: Salmonella enterica: Serovar Typhimurium LT2.

17. McLoughlin SY, Copley SD. 2008 A compromise required by gene sharing enables survival: implications for evolution of new enzyme activities. *Proc. Natl Acad. Sci.USA* 105**,** 13497 – 13502. (doi:10.1073/pnas.0804804105)

18. Näsvall, J., Sun, L., Roth, J. R., & Andersson, D. I. (2012). Real-time evolution of new genes by innovation, amplification, and divergence. *Science, 338*(6105), 384-387. doi:http://0-dx.doi.org.libraries.colorado.edu/10.1126/science.1226521

19. Ohno S. Evolution by Gene Duplication. New York: Springer; 1970.

20. Wang W, Zhou H, Lin H, Roy S, Shaler TA, Hill LR, et al. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. Anal Chem. 2003;75(18):4818-26.

21. Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nat Methods. 2009;6(5):359-62.

22. Wong JW, Cagney G. An overview of label-free quantitation methods in proteomics by mass spectrometry. Methods Mol Biol. 2010;604:273-83

23. Wu T. T., Lin E. C., Tanaka S. (1968) J. Bacteriol. 96, 447–456 [PMC free article]

24. Zwietering, M. H., Jongenburger, I., Rombouts, F. M., & van 't Riet, K. (1990). Modeling of the Bacterial Growth Curve. *Applied and Environmental Microbiology, 56*(6), 1875–1881.