# UC Berkeley

**UC Berkeley Electronic Theses and Dissertations**

**Title**

Understanding the Contribution of Cellular Factors to Heterologous Type III Secretion in Salmonella enterica

**Permalink**

https://escholarship.org/uc/item/3480w9hv

**Author**

Wong, Han Teng

**Publication Date**

2019

Peer reviewed|Thesis/dissertation

Understanding the Contribution of Cellular Factors to Heterologous Type III Secretion in
*Salmonella enterica*


By

Han Teng Wong


A dissertation submitted in partial satisfaction of the

requirements of the degree of

Doctor of Philosophy

in

Microbiology

in the

Graduate Division

of the

University of California, Berkeley


Committee in charge:

Professor Danielle Tullman-Ercek, Co-Chair
Professor Kathleen Ryan, Co-Chair
Professor Michiko Taga
Professor Nicholas Ingolia


Fall 2019

Abstract

Understanding the Contribution of Cellular Factors to Heterologous Type III Secretion in
*Salmonella Enterica*

by

Han Teng Wong

Doctor of Philosophy in Microbiology

University of California, Berkeley

Professor Danielle Tullman-Ercek, Co-Chair

Professor Kathleen Ryan, Co-Chair

Type III secretion systems (T3SS) are complex, membrane-embedded macromolecular machines found in Gram-negative bacteria. These bacteria utilize T3SSs to facilitate microbe-host interactions by translocating specific proteins out of their cytosol in a single step. This provides an interesting system for engineering the selective translocation of proteins out of the bacterial cell. Secretion strategies have been useful in exporting proteins of biotechnological interest in systems such as yeast, insect cells and mammalian cells. Bacteria fermentations are faster and cheaper than other systems.

In our lab, we use the *Salmonella* Pathogenicity Island 1 (SPI-1) T3SS in *Salmonella enterica* serovar Typhimurium for heterologous protein export. As the *Salmonella* SPI-1 T3SS is not an essential metabolic pathway, we can repurpose it for heterologous protein secretion without affecting cell viability under laboratory conditions. The availability of numerous genetic tools for *S. enterica* makes it an attractive platform for engineering purposes. Different heterologous proteins of interest can be directed to the T3SS using known native secretion signals. We were able to achieve high secretion titer greater than 400mg/mL through extensive engineering efforts.

My work focuses on the contribution of cellular factors to heterologous secretion by the SPI-1 T3SS through the use of comparative genomics and transcriptomics. Despite a deep understanding of the regulation of SPI-1 and its regulatory inputs, this complex circuit has yet to be successfully expressed heterologously in another host such as *E. coli*. This suggests that SPI-1 relies on cellular factors specific to *S. enterica,* despite the high genetic conservation of homologs between *S. enterica* and *E. coli*. The difference in the molecular makeup of the cellular environment could be a result of differential regulation of genes or genes unique to each species. Through my work presented here, I identified numerous cellular factors important for SPI-1 activation and engineered strains capable of high secretion of heterologous proteins. I also showed how insights from systems biology can be used to guide precise strain engineering and process engineering.

Dedicated to my mum and my long-suffering friends everywhere.
Fewer texts at weird hours of the day.

## Table of Contents

*List of Figures*

## List of Tables

## List of Abbreviations

aTc         Anhydrotetracycline
DH          DH domain of the human intersectin protein
GFP         Green fluorescent protein
IPTG        Isopropyl β- d-1-thiogalactopyranoside
LacZ        Beta-galactocidase
MBP         Maltose binding protein
mCherry     Monomeric red fluorescent protein
PhoA        Alkaline phosphotase
POI         Protein of interest
SPI-1       *Salmonella* pathogenicity island 1
SptP        Protein tyrosine phosphatase, a SPI-1 secreted effector protein
T3SS        Type 3 Secretion System

## Acknowledgements

Being away from home for a long time has never an easy feat, even if this was the second time away from family and friends. These past few years have definitely been a roller coaster ride for me. I am grateful to the continual support of many people.

To my PI, Danielle, who is always appreciative, nurturing and supportive. She allowed me to pursue research directions that were not traditionally done in the lab and had the confidence in me to pull through even at times when I doubted myself. Finding a mentor who I was willing to move from bright and sunny Berkeley to seasonal Chicago with frigid winters of wind chills where it didn't matter whether it was in Celsius or Fahrenheit anymore. I have learned how to be a better scientist and a better science communicator. I would still spell things with u's.

To my mum who has never failed to be there for me. You no longer have to check when I was last online on WhatsApp to guess my sleeping patterns and guess whether I am stressed out or not. Most importantly, thanks for understanding my time away from home.

To Anum who first showed me how to purify a protein. You always had time for me even after graduation. While you may not see yourself as the "perfect" mentor, but you will always be a great mentor to me because you cared. Most importantly, thanks for being a friend and for making navigating academia easier. One cannot really ask for a better mentor and a better friend.

To Rocio who was never more than an email away. You are the pillar of support any graduate student can have in a GSAO. You helped me navigate the various administrative loops needed to get here. I would not have made it past the initial years of graduate school without your guidance.

To my thesis committee members who provided guidance through my PhD journey. You have all provided invaluable advice in shaping my project during every thesis committee meeting. You were always willing to accommodate to my time restrictions as well as extension requests.

To my friends in lab who made working in lab a joy and my second home. There are many people past and present who made going to work a joy. Watching the lab shrink, move and expand again is an unforgettable experience. All of you in one way or another have made this a memorable journey. I have learnt to better tolerate the things I cannot change and to better accept people for all the good and bad in them.

To undergraduates and graduate students, I have mentored – Kim, Mandy, Vivian, Jason, Amber, Tanner and Julie, hopefully you have learnt something from me.

To my fellow A*STAR scholars. Going back to the Bay is always a treat and meeting up makes life a whole lot easier. You all made being part of the A*STAR family a lot more comfortable. We will all be back home soon and seeing each other a lot more. We will all get there.

To my fellow graduate students in Berkeley who made moving to the US easier. The warmth and camaraderie of PMB was uniquely PMB and not something I found anywhere else.

To all my friends everywhere, thanks for never losing faith in me. Even when you are far away, be it the UK, Australia or Singapore, you were but a text away. Thank you for enduring long conversations and being the support that I needed. Sorry for not being there at times and in the moments when you needed me. You guys have kept me sane throughout my entire PhD and have given me the strength to finish this, especially Ranika who painstakingly read through this entire tome. We may not be able to grab dinner whenever we wanted to but you all always gave me your previous time when needed. The emotional support sorely needed was never more than a text or call away.

# Chapter 1 Microbial Cell Factories for Recombinant Protein Production

The use of bacteria for heterologous protein production heralded the emergence of the bioeconomy, which now spans across numerous industries and is worth billions of dollars. The first recombinant protein that was commercially produced in a bacterium was insulin. Made in the Gram-negative bacteria, *Escherichia coli*, it was approved for sale as Humulin in 1982 and remain an important drug for the management of diabetes even to this day[1–3]. By producing insulin recombinantly in *E. coli*, there was no longer a need to rely on insulin harvested from pigs. This was followed by numerous blockbuster protein-based drugs, or biologics, produced in *E. coli* such as Protropin, Roferon A, IntronA and Humatrope in the 1980s[4].

From its humble beginnings, biologics have now drastically transformed the modern-day pharmaceutical industry[5]. In 2018, 13 of the top 20 best selling drugs are biologics, representing a huge market demand for these products[6]. The predominant choice of production host is Chinese hamster ovary (CHO) cells with only two drugs, insulin glargine and pegfilgrastim, produced in *E. coli* (Table 1.1).

| RANK | DRUG | SALES IN 2018 | BIOLOGIC | HOST CELL |
|---|---|---|---|---|
| 1 | Humira® (adalimumab) | $19.936 billion | Yes | CHO |
| 2 | Eliquis® (apixaban) | $9.872 billion | No | |
| 3 | Revlimid (lenalidomide) | $9.685 billion | No | |
| 4 | Opdivo® (nivolumed) | $7.570 billion | Yes | CHO |
| 5 | Keytruda® (pembrolizumab) | $7.171 billion | Yes | CHO |
| 6 | Enbrel® (etanercept) | $7.126 billion | Yes | CHO |
| 7 | Herceptin® (trastuzumab) | $6.981 billion | Yes | CHO |
| 8 | Avastin® (bevacizumab) | $6.847 billion | Yes | CHO |
| 9 | Rituxan® (also sold as MabThera; rituximab) | $6.750 billion | Yes | CHO |
| 10 | Xarelto® (rivaroxaban) | $6.589 billion | No | |
| 11 | Eylea® (aflibercept) | $6.551 billion | Yes | CHO |
| 12 | Remicade® (infliximab) | $5.908 billion | Yes | Hybridoma |
| 13 | Prevnar 13® / Prevenar 13® (Pneumococcal 13-valent Conjugate Vaccine [Diphtheria CRM197 Protein] | $5.802 billion | Yes | *Streptococcus pneumoniae* & *Corynebacterium diphtheriae* |
| 14 | Stelara (ustekinumab) | $5.156 billion | Yes | Recombinant Sp2/0 host cell line |
| 15 | Lyrica® (pregabaliln) | $4.970 billion | No | |
| 16 | Lantus® (insulin glargine) | | Yes | *E. coli* |
| - | Ibrance® (palbociclib) | | No | |
| 20 | Tecfidera® (dimethyl fumarate) | | No | |

| | | |
|---|---|---|
| Genvoya® (elvitegravir, cobicistat, emtricitabine, tenofovir alafenamide) | No | |
| Neulasta® / Peglasta® (pegfilgrastim) | Yes | *E. coli* |

**Table 1.1 Top 20 grossing drugs in 2018[6]**

There are currently numerous platforms for recombinant protein production in addition to *E. coli* which includes Gram-positive bacteria like *Bacillus subtilis*, yeast like *Saccharomyces cerevisiae* and *Pichia pastoris*, insect cells like Baculovirus expression system and mammalian cells like CHO[7–13]. As a biologic's manufacturing process can define its chemical and therapeutic properties, the choice of host can be limited. The dominance of CHO-based biologics is in part due to the power of antibodies-based therapies. Yet, approximately 30% of recombinant protein approved by the FDA are still made in *E. coli*, demonstrating the continued reliance on bacterial-based production[3,4].

In contrast, *E. coli* still reigns as the workhorse of biomedical research; for example, more than 85% of the protein structures deposited on the Protein Data Bank (PDB) were expressed in *E. coli* (Figure 1.1)[14]. This speaks to the ease of growth and manipulation of *E. coli* compared to the other production systems mentioned earlier.
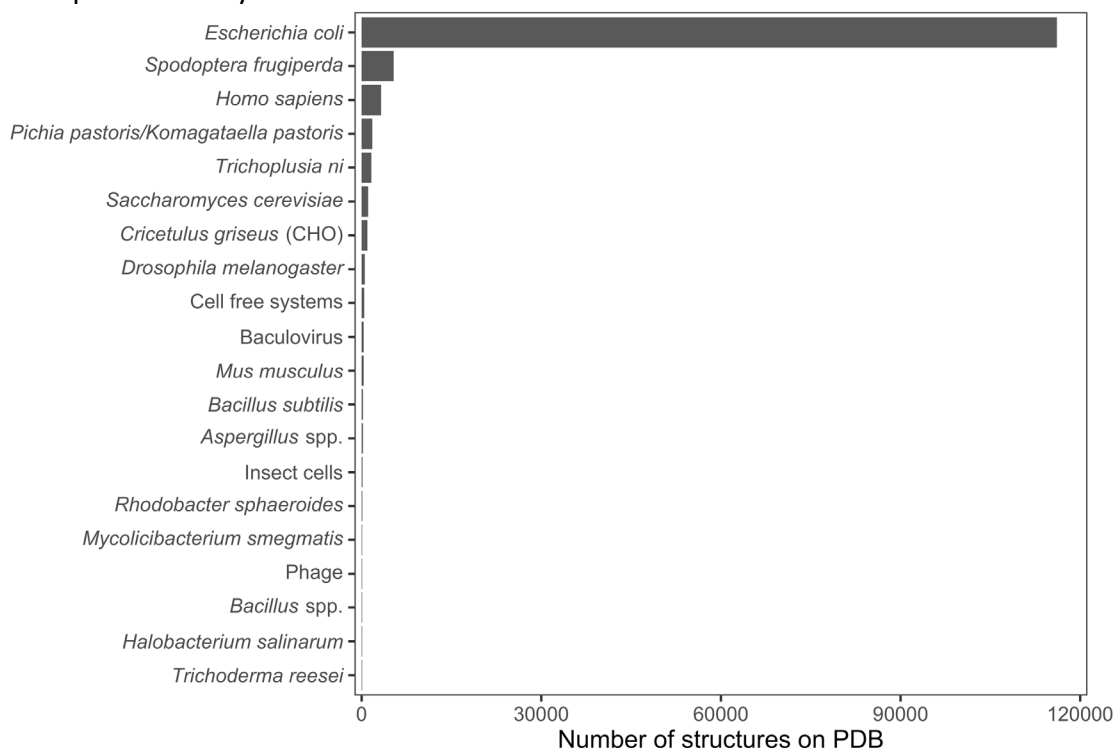


**Figure 1.1 The top 20 host for protein expression for structures on PDB shown here. Different strains for each host type were combined into a single group.[14]**

## *Limitations of recombinant protein production in Gram-negative bacteria*

A key difference between *E. coli* and the other available platforms is namely where the product of interest is located after growth and expression. In *E. coli*, the heterologous protein

2

remains intracellular whereas in the other production systems the heterologous protein is often found extracellularly. This is achieved by employing cellular secretion to target the protein of interest (POI) out of the cell.

Intracellular protein production is *E. coli* results in the need for numerous purification steps. This is due to the complex intracellular milieu being filled with various proteins that may interfere with the purification of the POI. Given typical overexpression conditions, the POI only makes up approximately 10% of the biomass. As such, costly purification processes are often needed to isolate the recombinant protein from the cell lysate[15,16]. In addition, the intracellular heterologous protein expression often leads to formation of inclusion bodies, resulting in the need for another costly process of refolding the protein. In a study on the heterologous production of the HIV-1 protein Nef, the primary purification represented 26% of the manufacturing cost[17]. This effectively reduces the advantages arising from cheaper development and cheaper production. One solution to reduce the need for extensive purification is to selectively secrete the target protein extracellularly[18,19].

In contrast to the intracellular milieu of bacteria, the extracellular space has fewer complex molecules per volume, approximately 10-100 times less molecules per volume. This means that the POI is likely to be greater than 10% of the dry mass. For this reason, a secretion strategy has been successfully employed in many other systems listed above. Adopting a secretion strategy in microbial cell factories would combine the benefits of cheap bacteria fermentations and efficient downstream processing.

In addition, coupling a bacterial fermentation with secretion strategy could enable a shift away from the currently predominant batch or fed-batch process fermentation used for *E. coli* protein production to continuous process fermentation. This improvement has technical and economic benefits that would in turn drive the development of more robust microbial systems for heterologous protein secretion[20].

## *Selecting a protein secretion system for heterologous protein secretion*

Developing a bacterial protein secretion production platform requires the consideration of numerous factors such as the number of steps needed to achieve secretion, the need for tag cleavage post secretion, the availability of beneficial cellular factors such as folding chaperones and the essentiality of the system. As there is a huge diversity in biochemical and biophysical properties in heterologous proteins, each secretion system provides different benefits in the production process[20].

In a Gram-negative bacterium such as *E. coli*, the protein of interest will need to cross both the inner and outer membrane of the bacterium. There are several major secretory pathways with the potential for creating a protein secretion platform in a Gram-negative bacterium – the general secretory pathway (sec), the twin-arginine translocation pathway (tat), Type 1 secretion system (T1SS), Type 2 secretion system (T2SS), Type 3 secretion system (T3SS), Type 5 secretion system (T5SS) and Type 8 secretion system (T8SS)[21]. Out of these systems, only T1SS and T3SS are single-step systems[20].

The dependence on a second step for complete secretion into the extracellular milieu may create the need to engineer not just one but two systems, thereby increasing the complexity of designing a bacterial secretion platform. In certain situations, separating the secretion into two

steps can be beneficial. One of the major drawbacks of single-step systems is the lack of folding chaperones. In contrast, two-step systems can employ folding chaperones and protein quality control systems in the periplasm that can be beneficial to the protein production. The availability of folding chaperones can drastically improve yield of functional protein products[22,23]. However, protein purity is the highest priority for creating a bacterial protein production system that could offer significant improvements over current options, and of these, we chose to employ a T3SS-based system because it enables robust secretion of a wide range of target proteins.

Utilizing *Salmonella enterica* to secrete proteins for biotechnological purposes combines the advantages of using a bacterial host and using a secretion strategy. Two of the major benefits of repurposing the *Salmonella* pathogenicity island 1 (SPI-1) T3SS for heterologous protein secretion are the ability to translocate protein out of the cytosol in a single step and the non-essentiality of SPI-1 T3SS for growth under conditions used in protein production[24–27]. Thus, we are able to repurpose the SPI-1 T3SS to secrete heterologous proteins of our choice. This allows the development of a new protein production platform in the age of biology.

## Developing SPI-1 T3SS as a secretion platform

*Salmonella enterica* serovar Typhimurium is a Gram-negative enteric bacterium that has been well-studied for decades. Pioneering work in *Salmonella* biology includes studies in bacterial transformation, bacterial transduction, the mutagenic nature of chemicals, and host-pathogen interactions[28–31]. The long history of *Salmonella* provides numerous tools for the manipulation of *S. enterica*. The ease of genetic manipulation of *S. enterica* is comparable to *E. coli*, allowing any strain development cost to be kept at similar levels to *E. coli*. Furthermore, the growth rate of *S. enterica* is similar to *E. coli* and the nutritional requirements of both bacteria are largely similar. These characteristics of *S. enterica* mean that the many advantages of *E. coli* as a heterologous production host are transferable to *S. enterica*. The vast number of tools available to manipulate and study *S. enterica* makes it a choice organism to engineer. Moreover, *S. enterica* has protein secretion apparatus that well characterized for their ability to translocate proteins.

The SPI-1 T3SS had been studied in great detail for its role and contribution to *Salmonella* pathogenicity, which is used by the bacterium to interact with and modulate host cells for successful colonization. This complex nanomachine is embedded in the inner and outer bacterial membrane of *S. enterica* spanning across the periplasm. It translocates proteins in an unfolded state from the bacterial cytosol into the host cytosol across three membranes – the bacterial inner membrane, the bacterial outer membrane and the host cell membrane[24,25,32,33], as well as the peptidoglycan layer. The proteins that make up the needle complex (Figure 1.2) are fully encoded on the island while some of the effector proteins, the native substrates for the system, can be encoded at different loci[34–37]. This amazing translocation is carried out in a single step and believed to be powered by the basal body of the injectosome complex mainly using ATP[38,39]. The contribution of each component to the function of the complex T3SS nanomachine is another area of active investigation[27,40]. This will allow us to tweak the functionality of the T3SS nanomachine using traditional protein engineering approaches and directed evolution strategies. In Chapter 2, I explore the role of SipD, the needle tip protein, on SPI-1 secretion.
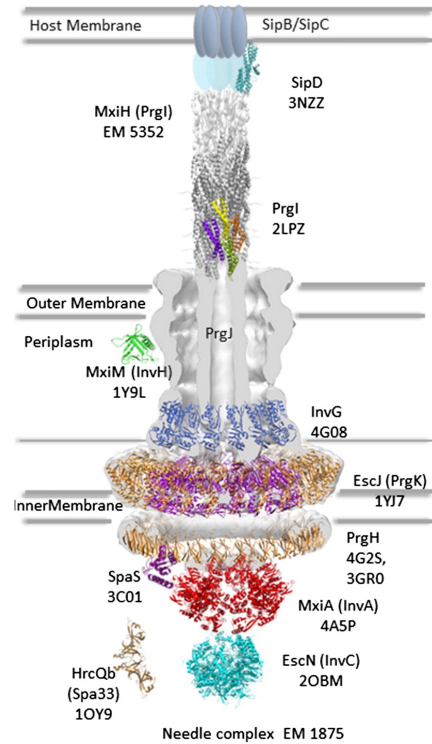
**Figure 1.2 Cartoon of the SPI-1 needle complex. Taken from Burkinshaw and Strynadka, 2014[41].**

Native regulation of SPI-1 T3SS is complicated and depends on the coordination of numerous transcription factors within the SPI-1 island (Figure 1.3)[42–44]. At the top of the regulatory network is a triad of transcription factors, HilD, HilC and RtsA[35,45–48]. Under the right environmental conditions, these transcription factors will trigger the expression of another transcription factor, HilAs which in turn drives the expression of two major SPI-1 operons. These operons, *prg* and *inv*, encode the structural components of the T3SS injectosome[49,50]. Another transcription factor, InvF, in turn activates the *sic/sip* operon and effectors, the native substrates of the system[51,52]. This activity of InvF is facilitated by the multifunctional chaperone, SicA[53]. This highly ordered network results in the timely expression of genes for the correct construction of the complex nanomachine[41,54,55].

The deployment of SPI-1 to invade host cells is a bet-hedging strategy of *S. enterica*. Only 30% of the population expresses the system under the right environmental conditions of high salt and low oxygen[56–58]. The commitment to the system activation is postulated to lie in the speed of HilD expression as the regulation of HilD is a feed-forward loop[46,59–61]. Once the level of HilD level crosses the threshold needed, the bacterium is committed to the invasion cellular state.

In order to take advantage of the existing network architecture, key transcription factors near the top of the network can be synthetically expressed in trans. This strategy was successfully employed in a previous study where HilA was placed under the control of the commonly used pLac-UV5 promoter. Metcalf *et al.* were able to synthetically induce the T3SS system using the inducer IPTG[43]. In addition, the ordered network architecture was further exploited through the use of the native promoter pSicA to drive the heterologous protein expression, eliminating the need for a second inducer[43,62]. The two-plasmid system allowed for both expression of the SPI-1 injectisome and heterologous protein with a single inducer.

Inducing SPI-1 activation by overexpressing HilA is but one of the many available strategies. In theory the deregulation of the *hil* genes can result in the activation of the system. The hyper invasive locus (*hil*) was discovered by studying *S. enterica* mutants capable of hyper-invasion resulting from constitutive expression of these genes. Additionally, the overexpression of HilD is also commonly employed in the pathogenicity studies. Other strategies include changing the external environment to favor the accumulation of HilD or other *hil* genes. In Chapter 4, I explore a different approach for activating activate SPI-1 expression by modulating HilD levels.
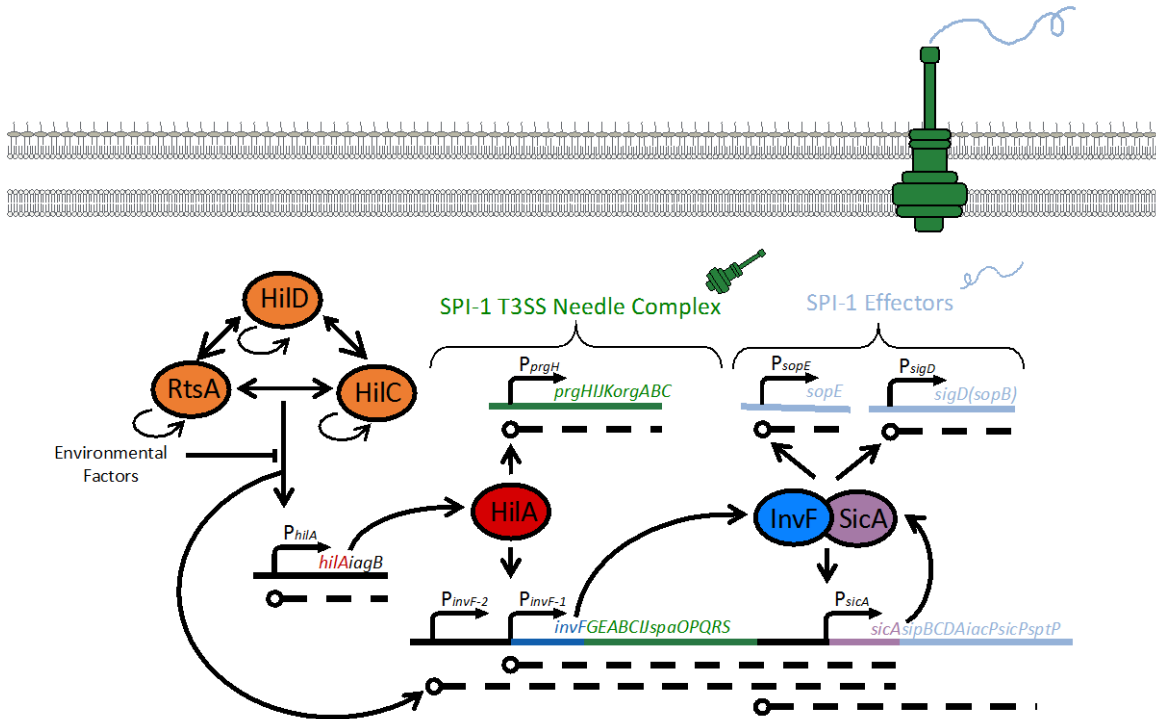


**Figure 1.3 The genetic regulation of SPI-1 is well studied. Shown here are the different transcription factors and their respective cognate promoters. Note that there is some readthrough in transcription from the *inv* operon into *sicA*. This provides some SicA proteins needed for InvF functionality. Adapted from Lostroh and Lee, 2001[34].**

Most secretion systems are dependent on a peptide signaling tag to direct proteins towards the secretion machinery[20]. In order to effectively target a foreign protein to the T3SS, secretion tags derived from the N terminus of native effector proteins are appended to the N terminus of the heterologous protein (Table 1.2)[62]. Of these tags, the first 167 amino acid of SptP is currently the most robust secretion tag and has been used successfully to target numerous proteins for heterologous secretion. The large 19kDa tag can alter the expression levels of the overall protein. The use of different secretion tags can result in very different secretion efficiencies. On the other hand, the biochemical nature of the protein can also drastically affect the likelihood of its translocation by the system. Our model target protein for SPI-1 mediated heterologous secretion is the DH domain of the human intersectin protein. When appended to the SptP tag, up to 30% of the expressed protein is exported by the T3SS apparatus[42,63]. In contrast, only 5-20% of the expressed protein may be translocated in other use cases. Furthermore, each protein may have an ideal tag. The spider silk protein, ADF-1, secretes at the highest titers when appended to the sigE secretion tag[62]. The development of *de novo* secretion tags and design rules are areas of active investigation.

Adding tags can affect the functionality of the final product, so tags must be removed in a downstream process. The signal tag is cleaved in the sec pathway, tat pathway, and partially in T5SS, but not in T3SS. Using specific proteases that cleave after a specific peptide sequence, such as the TEV protease, can be ameliorate the lack of native cleavage machinery. Of note, this adds an additional processing step post-secretion, adding to the production cost.

| Sec Tags | | |
|---|---|---|
| Effector | Chaperone | Tag length (AA) |
| SipA | InvB | 169 |
| SipC | SicA | 167 |
| SopA | InvB | 96 |
| SopB | SigE | 168 |
| SopD | None | 40 |
| SopE2 | InvB | 105 |
| SptP | SicP | 167 |

Table 1.2 List of chaperone and tag pairs.

## Regulation of SPI-1 island by factors outside the island

The horizontal acquisition of SPI-1 is an ancient evolutionary event compared to the appearance of the other *Salmonella* pathogenicity islands[33]. This meant that SPI-1 could have co-evolved with the rest of the genome and the bacterium's overall physiological regulation. The tight regulation of SPI-1 is achieved in tandem with the cell's physiological state by interfacing with numerous global regulators including the stringent response, osmoregulation pathways, and nucleoid silencing (Figure 1.4)[64–79]. This can be triggered by different growth conditions such as temperature, osmolarity, growth phase and pH. These complex inputs ensure that *S. enterica* only express the T3SS under conditions conducive to invasion of a mammalian host cell[80].
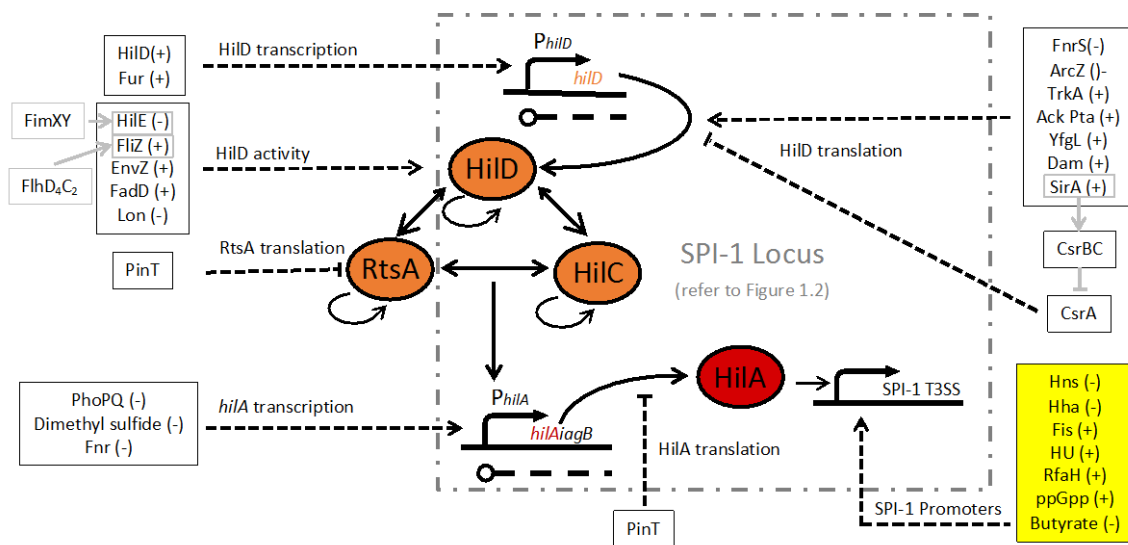


Figure 1.4 There are many cellular factors that are known to interact with SPI-1. The factors in the yellow box are of high interest as they are known to affect numerous promoters in the system. Adapted from Golubeva et al, 2012[64].

Global regulators also shape the cellular physiology and state. The right cellular state may be needed to accommodate the synthesis of SPI-1 T3SS numerous components and its insertion across both membranes. For example, the peptidoglycan layer would need to be remodeled and this may partially be achieved by *iagB*[81]. The Cpx envelope stress response and quality control of the periplasmic disulfide system are also known to contribute to successful needle complex assembly[82–84].

Furthermore, the expression of SPI-1 and the heterologous protein of interest are also dependent on cellular resources. This can range from sigma factor competition, ribosomal occupancy and charged tRNA availability[85]. Additionally, spatial distribution of SPI-1 T3SS transcripts may also affect translational efficiency. The unfolding and translocation of SPI-1 substrates also requires energetic resources such as ATP and PMF, thus placing strain on cellular metabolism. Thus, remodeling the cellular physiology and metabolism could be critical for robust continuous heterologous secretion[86].

By overexpressing *hilA*, we were able to bypass many layers of control and be less reliant on the environmental conditions that SPI-1 needed for activation. Other than the environmental factors, there are also various cellular factors (shown in yellow box of Figure 1.4) that can modulate the activity of SPI-1 promoters. In Chapter 5, I examine the changes in global regulation during heterologous secretion and explore how tweaking these cellular factors may help maintain cells in a secretion-active state.

## *Evolution provides clues for essential cellular factors*

The large number of Typhimurium serovars as well as Typhi serovars can have slightly different phenotypes, different levels of pathogenicity and different levels of secretion capacity[87,88]. Thus, the choice of strain can have a huge impact on heterologous secretion. By comparing the genomes of the different strains, we can identify cellular factors that distinguish the different strains and determine which ones can influence heterologous secretion. With the power of a high-throughput screening, it is possible to harness the power of adaptive lab evolution to uncover more influential cellular factors. Both of these evolution-based approaches are investigated in Chapter 3. I approached developing a high-throughput assay for heterologous secretion in Chapter 6.

One interesting evolutionary mark is the lack of the lac repressor, *lacI*, in *Salmonella* species with SPI-1 but is present in other species without SPI-1[89]. Ectopic expression of LacI can affect SPI-1 expression[90]. In fact, the presence of LacI can inhibit heterologous secretion. Adding 100μM IPTG to *S. enterica* bearing an empty pLac plasmid (the same backbone is used for overexpressing *hilA*) results in slightly higher expression and secretion of our model protein (Figure 1.5a). In addition, the use of other common promoters can also an unexpected positively impact on SPI-1 expression. In contrast, use of arabinose dampens SPI-1, making the use of pBAD plasmids tricky.
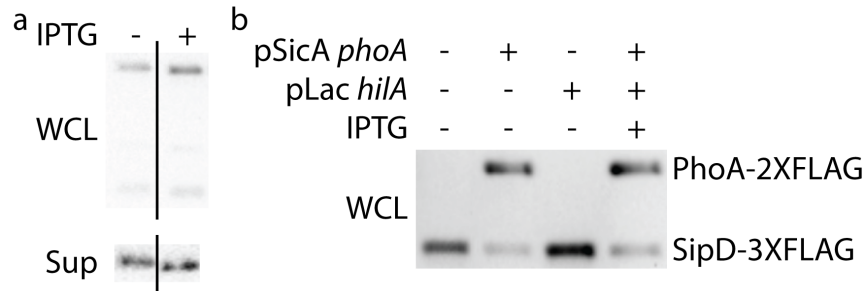
**Figure 1.5(a)** Western blot of strains bearing pLacUV5 empty vector with and without 100μM IPTG. **(b)** Western blots of hilD-3XFLAG with different plasmids and conditions.

## Need for plasmid-free system

The use of plasmid can lead to strain instability and metabolic burden from plasmid maintenance. As mentioned earlier, the use of pLac or pBAD can be tricky due to the impact of either the transcription factor or inducer. In addition, the presence of the export plasmid using pSicA can also reduce the expression of SPI-1 components such as SipD (Figure 1.5b). This could be due to transcription factor competition since *sipD* is part of the sic operon or resource competition. Thus, there is a strong need to develop plasmid-free systems as well as to re-evaluate the impact of different plasmids and promoters on SPI-1 expression. The first attempts to address these challenges can be found in Chapter 4.

## Conclusion

In this chapter, I made a case for developing a secretion platform based on the *S. enterica* SPI-1 T3SS for protein production but there remain challenges to implementing this vision. With this dissertation, I describe my work to address the issue as follows. In Chapter 2, the role of the needle tip complex, SipD, in SPI-1 is explored. As SipD acts as the gatekeeper to ensure targeted secretion into a host cell, altering its properties or expression can have an impact on secretion. In Chapter 3, I explore the genetic and phenotypic differences between the common *S. enterica* strains available to uncover cellular factors that may influence heterologous secretion. In Chapter 4, alternative strategies to induce SPI-1 expression based on the cellular concentration of another SPI-1 regulator, HilD. Both protein levels and HilD activity were modulated to achieve inducer-free heterologous secretion. In Chapter 5, I detail initial efforts in using transcriptomics data and system biology to guide engineering efforts. In Chapter 6, I discuss the development of different tools to better assay heterologous secretion in *Salmonella* biology. This includes developing selection and screening strategies for isolating strains that secrete heterologous proteins at higher titers. Taken together, the results described in these chapters moved us significantly towards the vision of exploit T3SS for heterologous protein production.

## Chapter 2 Role of SipD in Heterologous Secretion

The SPI-1 T3SS injectisome needle is capped by three different structural proteins – SipB, SipC and SipD[91,92]. Collectively known as the translocon complex, these proteins allow for directed translocation of proteins into the host cytosol by contacting the host cell membrane[49,93–95]. The needle complex is initially capped by SipD and then upon contact, SipB and SipC are secreted to form a pore in the host membrane[96]. This creates a channel through which proteins can be translocated into the host cytosol[26].

Secretion through the SPI-1 T3SS can be characterized by leaky and burst secretion. Leaky secretion occurs when cells are grown in the media. Up to 40% of the expressed heterologous protein can be released into the supernatant by this mechanism (Figure 2.1 A)[63]. Native substrates can also be detected in the supernatant upon synthetic induction of the SPI-1 T3SS. Burst secretion is when an unknown trigger, possibly cholesterol on the host membrane, causes the release of the effectors into the host cells[97,98]. Whilst the trigger for burst secretion is better studied in other pathogens such as *Shigella flexneri* and *Yersinia pestis*, the specific trigger for engaging the SPI-1 T3SS is an area of active research[99–101].

As SipD and the associated translocon caps the needle complex, SipD is believed to be the gatekeeper that controls translocation of proteins through the SPI-1 T3SS complex. Knocking out the translocon resulted in a less controlled release of proteins into the supernatant by the SPI-1 T3SS complex. This also led to lower pathogenecity. When *sipD* was knocked out, we observed increased heterologous secretion with more of the expressed intracellular protein pool being secreted through the needle complex (Figure 2.1A). The ability for constitutive ungated secretion would be ideal for robust heterologous secretion. In addition, the translocon is not essential to the protein translocation function of the SPI-1 apparatus.

In addition to the role described above, we uncovered another unique role of SipD. The addition of purified SipD exogenously unexpectedly resulted in increased secretion titer instead of restoring the wild type phenotype (Figure 2.1A). This phenomenon is dependent on the concentration of SipD protein added to the culture. The increase in secretion was mainly driven by greater mobilization of the expressed intracellular pool of proteins into the supernatant. Up to 90% of the expressed protein could be detected in the supernatant upon addition of high concentrations of SipD (Figure 2.1A).

### Hypersecretion occurs even in absence of de novo transcription and translation

We reasoned that the exogenous addition of SipD to T3SS-induced cells increases their secretion titer either (a) post-transcriptionally, (b) post-translationally or (c) both post-transcriptionally and post-translationally. To explore these options, we inhibited transcription or translation of *S. enterica* DW01 cells secreting heterologous model protein DH from an export plasmid and in which *hilA* was overexpressed, using sublethal concentrations of rifampicin (100 μg/ml) or tetracycline (8 μg/ml), respectively[102]. These cells were positive for the hyper-secretion phenotype in both cases (Figure 2.1B), suggesting that proteins that are targeted for secretion

build up in the cell before they can be secreted, and that the secretion is activated by a post-translational signal. Therefore, it is likely that (a) secretion of effector proteins can occur post-translationally as well as co-translationally, and (b) the natural extracellular effect of SipD on secretion levels is likely to be post-translational. Secretion of native effectors was also be observed when exogenous SipD was present.

Additionally, this observation leaves open the possibility that effector secretion activation occurs as a result of a conformational or steric change in needle structure to permit more rapid protein trafficking through already-established T3SS in the cell.
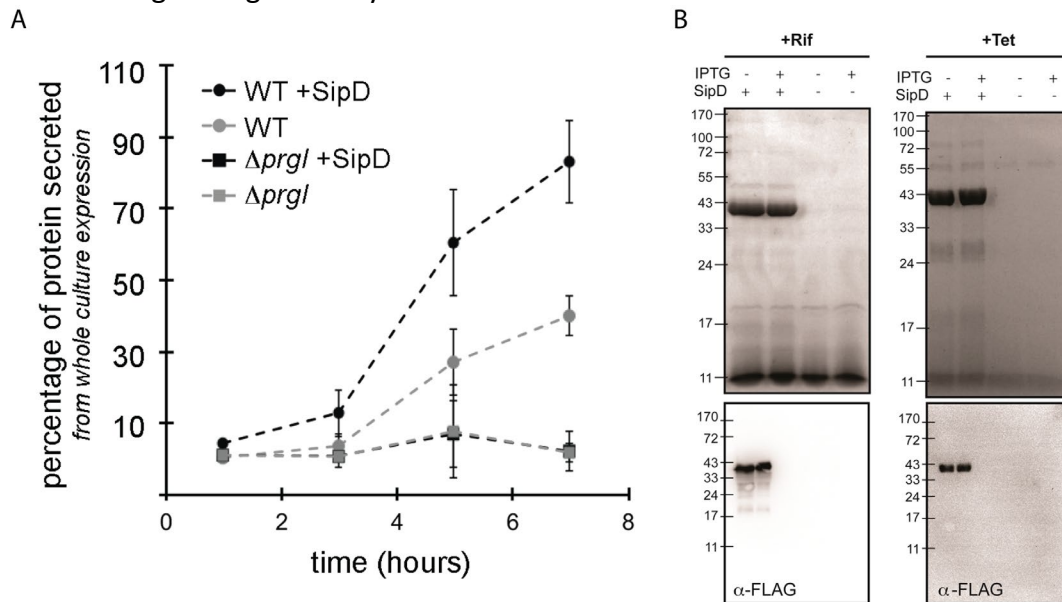


**Figure 2.1 Whole culture expression and secretion titer were measured by western blotting. (A) Adding SipD to Δ*prgI* did not result in increased secretion, showing that changes in secretion is due to the SPI-1 apparatus. The increase in secretion observed is mainly driven by higher levels of protein translocation. (B) *De novo* transcription and translation were stopped using rifampicin and tetracycline respectively. Hypersecretion is still observed when SipD was added to the culture. Taken from Azam et al, 2017[63].**

## *Determining the SipD domain responsible for increased secretion*

We wanted to determine whether a specific structural domain is responsible for the secretion phenotype we observed with exogenous SipD addition. The first 133 amino acids, SipD$^{\Delta134-343}$, include a flexible region of the protein interspersed with three short alpha helices for which no function has previously been described. Structures determined for SipD eliminate the extreme N-terminal section of this domain, residues 1-32, entirely; this is due to the difficulty of protein crystallography with flexible peptide sequences. However, there is evidence to suggest that this domain may serve as an intramolecular chaperone that prevents premature oligomerization of the residues in the C-terminal coiled-coil region (amino acids 295-322), which are involved in interactions with the PrgI needle and/or other SipD molecules based on studies on its homolog, IpaD in *S. flexneri*[95,103,104]. The C-terminal residues may form oligomeric structures at the T3SS needle tip that shift the conformations of the N-terminal domains to regulate the secretion of other effectors upon interaction with PrgI. Corroborating this idea, another study shows that the N-terminal domain of SipD unfolds independently from the rest of the molecule prior to binding to PrgI. In alignment with these observations, in this study, we experienced difficulties in purifying

recombinant SipD$^{\Delta1-96}$ due to its propensity to aggregate without the self-chaperoning N-terminal domain, leading to low yields as compared with wild-type SipD.

To determine which structural domain of SipD confers the secretion activation, we added truncations of SipD to T3SS-induced cultures. The truncations were chosen to test the functions of the distinct domains of SipD individually and are known to fold independently. The exogenous addition of truncated SipD domains to cultures revealed that the N-terminal domain SipD$^{\Delta134-343}$ is necessary and sufficient to achieve the increased secretion titer phenotype (Figure 2.2B). Adding 1.2 µM of SipD$^{\Delta134-343}$ and native SipD achieved an equal increase in secretion titer in T3SS-induced cultures, but the addition of 1.2 µM of an N-terminal truncation of SipD$^{\Delta1-96}$ did not result in the hypersecretion phenotype but is instead similar to wild-type secretion levels (Figure 2.2B). Based on previous models for SipD and PrgI interactions, the N-terminus of SipD is unlikely to interact with PrgI. SipD$^{\Delta134-343}$ also lacks the mixed α/β region that interacts with SipB. These new data, combined with what is already known, contribute to a clearer picture of the different structure/function characteristics of SipD (Figure 2.2C), which is valuable particularly in the absence of a crystal structure of the complete protein.
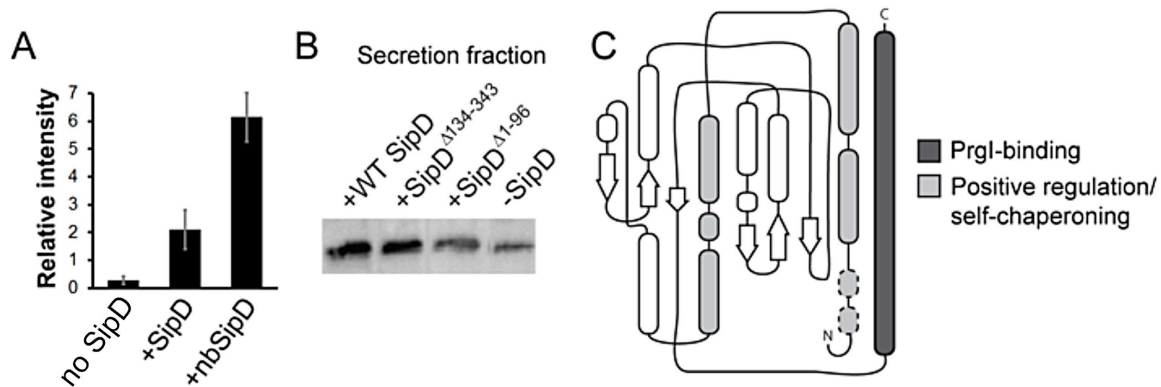


**Figure 2.2 The nbSipD is a triple mutant, SipDD$^{320R/V323K/S327R}$, that no longer bind to PrgI. This SipD mutant was able to further increase secretion as seen in (A). (B) The N-terminus of SipD is needed for the hypersecretion phenotype but not the C-terminus. (C) shows the different functional domains of SipD. Taken from Azam et al, 2017[63].**

The effects of exogenous addition of the N-terminal domain of protein to T3SS-induced cultures support a secretion-independent positive regulation role for extracellular SipD. As such, the localization of SipD could serve as a signal for the secretion competency of *S. enterica's* T3SS. Interestingly, the effect is enhanced with increasing concentrations of SipD well beyond what we expected to be physiologically relevant.

## Changes in transcriptional programming of SPI-1 due to SipD

Previous data on SipD showed that it can alter the behavior of promoters in SPI-1. Other homologs of SipD had been known to alter mRNA levels of their own T3SS system[105]. To better study the native context of these promoters, I carried out reverse quantitative transcriptase PCR (RT qPCR) on seven genes that spanned across the different operons of SPI-1 – *hilD*, *hilA*, *invF*, *prgH*, *prgK*, *sicA* and *sipA* (see Chapter 5 for development of RT qPCR). To determine the impact of SipD intracellularly and extracellularly, I carried out RT qPCR across three timepoints (2 hours,

4 hours and 6 hours) in three different strains – ∆*sipD*, wild type + exogenous SipD and wild type overexpressing *hilD* on a pBAD plasmid.

The biggest positive fold change for wild type + exogenous SipD was observed at 2 hours in *invF* and *prgH* (Figure 2.3). Knocking out *sipD* resulted in increased expression of *hilD*, *invF*, *prgH*, *prgK*, *sicA* and *sipA* at 2 hours as well as *hilA* at 6 hours. In contrast, wild type overexpressing *sipD* resulted in the decrease in the transcription of numerous genes at different timepoints. Unexpectedly, there was an increase in *hilA* transcript at 6 hours in this strain. Genes on the same operon did not have similar patterns. The *prg* operon had different patterns for *prgH* and *prgK*. The *sic/sip* operon had largely similar patterns for *sicA* and *sipA*. This could be due to early termination in transcription. Taken together with the promoter fusion data reported in Azam et al, 2017[63], this suggests that manipulating SipD level intracellularly and extracellularly can shift the temporal dynamics of the SPI-1 regulation.
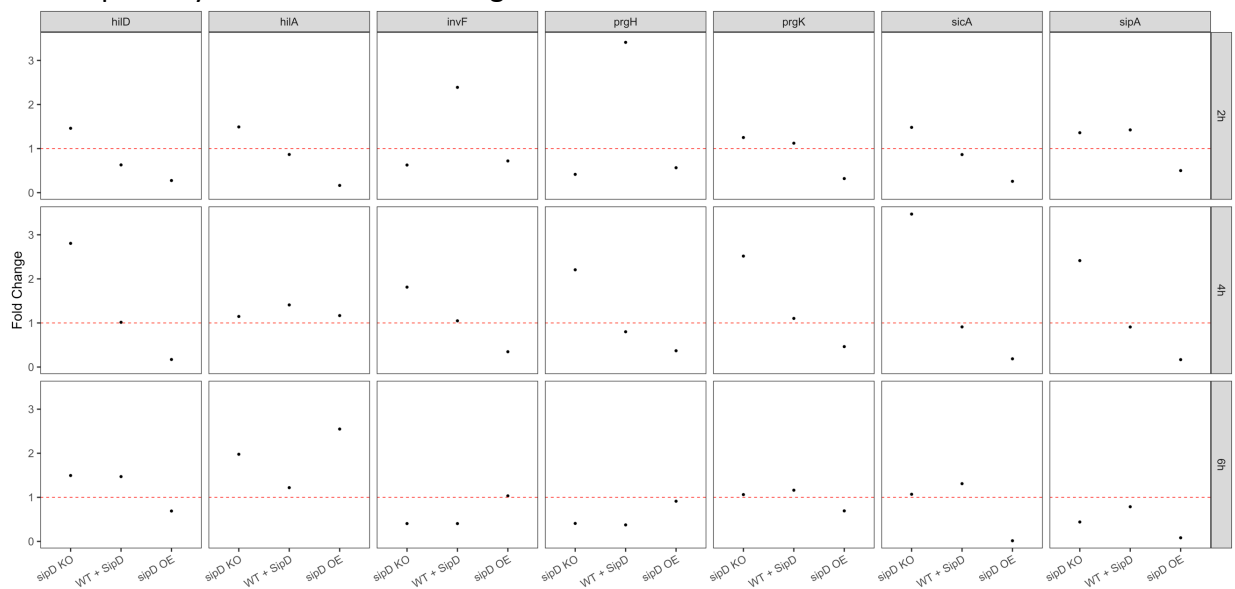


**Figure 2.3 The fold change was calculated by the ∆∆C$_t$ method. The expression of each gene was normalized to the wild type expression level at the same time point.**

## Oligomeric state of SipD

The oligomeric state of SipD can be altered based on the presence or absence of the N-terminus[92,103]. Prior work on SipD showed that dimer formation can be detected. The dimerization was shown to be dependent on the cysteine 244 on SipD[93]. Thus, I wanted to confirm the oligomeric state of SipD purified using our purification scheme and determine if the oligomeric state is important to SipD extracellular function. The oligomeric state of SipD was assessed on size exclusion chromatography using the HPLC and compared against a set of protein standards with known sizes. The purification scheme in our lab resulted in a protein solution of mostly monomer and small proportion of dimer (Figure 2.4).

I wanted to determine the nature of the interactions driving the dimerization/oligomerization of SipD. To test this hypothesis, I treated SipD by boiling it, reducing it with DDT, and denaturing

it with urea. The resulting protein solution was then run on the HPLC to determine its oligomeric state as above.

The dimer formation was not driven by disulfide bonds as the dimer peak did not disappear when the protein was treated with a strong reducing agent, 0.1M DTT (Figure 2.4A). 10mM iodoacetamide was used to alkylate any reduced cysteine residues to prevent the reformation of disulfide bonds prior to running on the HPLC. Neither did boiling SipD for 10 minutes resulted in changes to its oligomeric state (Figure 2.4A). The dimer peak disappeared upon treatment with increasing concentration of urea (Figure 2.4B). This further confirmed that dimer formation is not due to covalent bonding as urea denatures the protein but would not break covalent linkages such as disulfide bridges.

To determine if the purification scheme used could result in a different oligomeric state of Sip D, I used a chemical lysis approach, BugBuster, and then ran the purified protein solution on the HPLC. The method of protein preparation only had monomers (Figure 2.4 C) which could be due to the unfolding of the protein by the detergents present in BugBuster. Although the resulting protein had been buffer exchanged into PBS, there was no reconstitution of the dimer. Of note, this preparation of SipD did not elicit the hypersecretion phenotype. Preliminary evidence showed that the higher order of oligomerization in SipD might be essential to elicit the hypersecretion phenotype. As I was unable to isolate enough SipD that were of a higher oligomeric state, it was not possible to determine the exact form of SipD that is needed for its exogenous role.
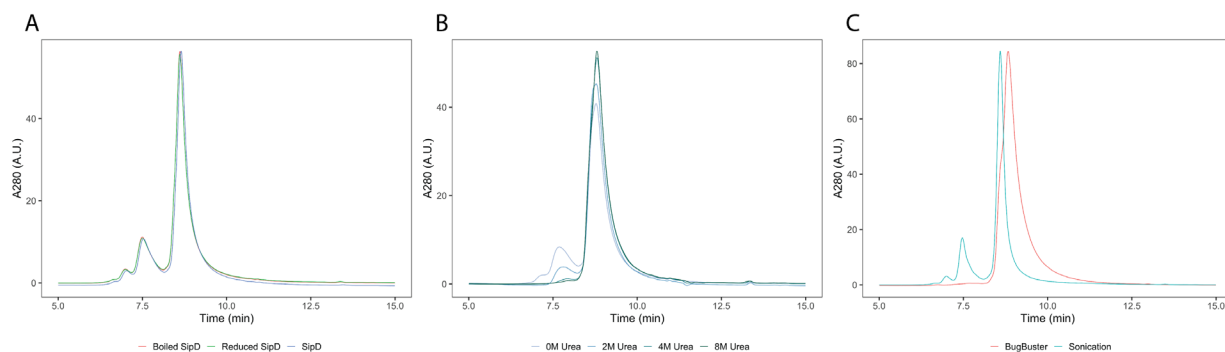


**Figure 2.4 The A280 traces shown here are taken from the HPLC. (A) SipD was boiled and reduced but neither treatment changed the traces shown here. (B) Increasing levels of urea leads to the loss of higher order states of SipD. (C) Using BugBuster for lysis resulted in a very different trace. The detergents in BugBuster might have unfolded SipD and any higher order structures did not reform when the detergents were removed by buffer exchange.**

## Discussion

Apart from demonstrating a facile method for increased protein secretion in a bacterium, our findings shed light on the complexity of the role of SipD in pathogenicity. Interestingly, in other T3SS-expressing organisms, including the well-studied *Yersinia* spp. and *Pseudomonas* spp., small chaperone proteins LcrG and PcrG negatively regulate effector protein secretion upon intracellular binding to their respective SipD-like translocon proteins, LcrV and PcrV, and block secretion[106–108]. However, LcrV and LcrG are positive regulators for secretion in addition to serving as structural proteins. In secretion-permissive conditions in *Yersinia* spp., LcrV

concentration is increased in the cell relative to LcrG concentration, and formation of an LcrV/LcrG complex removes the block on effector secretion via LcrG sequestration[109]. The Δ*lcrG* strain exhibits decreased LcrV secretion, and a mutant LcrG that cannot bind LcrV constitutively blocks secretion in all conditions. The homologous PcrV/PcrG system in *Pseudomonas* behaves similarly[110]. These results suggest that the interaction between a negatively regulating chaperone protein and a positively regulating translocon protein–and the expression levels of these proteins with respect to the time scale of secretion–are necessary for controlling the secretion of effector proteins.

However, no LcrG or PcrG homologue has been described for *Salmonella* or *Shigella*, and an absence of an intramolecular chaperoning domain in LcrV and PcrV suggests a divergent assembly process at the needle tip between the *Salmonella/Shigella* and *Yersinia/Pseudomonas* species. Our data supports a model in which the N-terminal domain of SipD has a positive regulation function as well as a chaperoning function, and the C-terminal domain independently performs the negative regulation function attributed to the LcrG/LcrV complex in *Yersinia*. SipD and *Shigella* SipD homologue IpaD share >90% sequence similarity at the C-terminus but very little sequence similarity with LcrV and PcrV, so it is possible that the absence of a small IpaD chaperone in *Shigella* is also compensated by the activity of the IpaD N-terminal region. Indeed, through random mutagenesis, IpaD has been shown to act as both a regulator and as part of a signal transduction pathway in type III secretion, though the mechanism for these activities remains uncertain. Nonetheless, the regulatory activity of SipD that we have described here does not preclude the involvement of other T3SS elements. Notably, the multifunctional nature of LcrV in three distinct locations (i.e., inside the bacterial cytoplasm, at the tip of the secretion needle, and inside the host cell), depending on its binding interaction with LcrG, lends support to a complex role for SipD in *Salmonella* secretion and virulence.

This study describes our construction and characterization of a hyper-secreting strain for the high-titer production of proteins of biotechnological interest, as well as regulatory and structural information pertaining to its mechanism of action. The ability to short-circuit the tightly regulated SPI-1 T3SS for constitutive secretion allows for direct application in repurposing the T3SS for biotechnological purposes. In previous studies, most of the expressed heterologous proteins that were targeted for secretion were trapped in the cytosol, and this is largely alleviated with exogenous SipD addition. Although high-titer production of a protein as a prerequisite to high-titer production of another protein is nonideal in an industrial setting, the secretion titer of the T3SS and other microbial secretion systems can be increased by dissecting and rewiring the roles of all of the proteins in the system. This work demonstrates that a better-understood and refactored T3SS has the capacity to serve at industrial scales. Future work will focus on detailing the molecular mechanism governing this effect and how to replicate it without protein addition. Furthermore, better characterization of T3SS proteins will reveal new targets for next-generation technologies with applications in energy, biomaterials and electronics, as well as pharmaceutical development to reduce illness and death resulting from *Salmonella* infections[42,99,111].

## Materials and methods

### Cloning of various constructs

The various SipD truncations were cloned using the Golden Gate type II endonuclease strategy[112]. Gene inserts were prepared using PCR with the BsaI restriction site flanking the 5' and 3' ends using primers found in Table 2.1. Inserts were then thermocycled with Golden Gate-compatible vectors, T4 DNA ligase (NEB), T4 DNA ligase buffer and BsaI (NEB) using standard procedures as described previously. 5 µl of the Golden Gate reaction was then used to transform chemically competent *E. coli* DH10B cells by heat shock. The chemically competent cells were subjected to 20 minutes on ice, followed by 60 seconds at 42°C, 2 minutes on ice and then recovery for 1 hour at 37°C with 350 µl of SOC media. 50 µl of the transformed cells were plated on LB agar plates with kanamycin. All genes were sequenced-verified by Quintara Inc. (Richmond, CA) using their inhouse pET vector primers.

### Strains and growth conditions used

DW01 cells were transformed with the required plasmids by electroporation and plated accordingly on LB agar plates with the respective antibiotics. All overnight cultures were grown aerobically in lysogeny broth (LB) - Lennox (VWR) supplemented with the required antibiotics, kanamycin (50 µg/ml) or chloramphenicol (34 µg/ml), with shaking at 225 rpm and 37°C (MaxQ 8000, 443, Thermo Scientific).

For secretion assays, the cells were cultured overnight for 12 to 16 hours and subcultured from the overnight culture at approximately 1:100 dilution to give a starting OD of 0.05 in 5 ml media. 100 µM IPTG (Fisher Scientific) was used to synthetically induce the overexpression of HilA as needed. After 8 hours of growth at 37°C with shaking at 225 rpm, cells were pelleted by spinning at 4000 g for 10 minutes in a tabletop plate centrifuge (Beckman Coulter). The supernatant was collected for analysis and filtered using .22 µM syringe filters (Thermo Scientific) and stored at 4°C. Whole culture lysate and secretion fraction samples for western blotting and protein gel electrophoresis were prepared immediately and stored at 4 °C until used.

### Purification of recombinant SipD

pET28b plasmids with different SipD mutants were introduced into *E. coli* BL21 DE(3) with pLysS. The strains were grown for 12-16 hours at 37°C and 225 rpm. Saturated overnight cultures were added at a 1:100 dilution into 1 L Terrific Broth and then grown at 37°C and 225 rpm. After 2 hours of growth, protein expression was induced with 1 mM IPTG. The cells were harvested by centrifugation at 6000 g for 20 minutes after another 6 hours of growth. The resulting cell pellets were re-suspended in binding buffer (20 mM sodium phosphate (Fisher Scientific), 500 mM NaCl (Fisher Scientific), and 25 mM imidazole (Acros Organics), pH 7.3). The cell solution was then sonicated at 50% amplitude using 10 second on and 20 seconds off cycles for an accumulated sonication time of 30 minutes. The insoluble fractions were removed by centrifugation at 12,000 g for 10 minutes twice. The soluble fraction of the cell lysate was passed through a $Ni^{2+}$-NTA affinity column (GE Healthcare) at room temperature under native conditions. The column was washed with 60 mL of binding buffer. Proteins were eluted with binding buffer supplemented with 250 mM imidazole, pH 7.4. Samples were desalted using PD-10 columns (GE Healthcare) and

then buffer exchanged into phosphate buffered saline (PBS). SipD purity was assessed by confirming a single band at 35 kDa with SDS-PAGE, and purified SipD concentrations were determined by absorbance at 280 nm using the extinction coefficient for SipD (as calculated using the ExPASy ProtParam tool). The identity of the protein was confirmed multiple times with tandem mass spectrometry using the UC Berkeley QB3 Mass Spectometry Core Facility with 98% protein coverage.

Lysis by BugBuster was carried out using the manufacturer's protocol by diluting the 10X buffer to 1X. Benzonase was added to degrade the nucleic acid for easier purification. $Ni^{2+}$-NTA affinity column purification and buffer exchange were carried out as above.

### SDS-PAGE & western blotting

Whole culture lysate and secretion fractions were analyzed for protein content using protein gel electrophoresis followed by western blotting. For whole culture lysates, 10 μl of culture was collected before pelleting cells. For secretion fractions, 10 μl of sample was collected after cells were pelleted without further concentration unless noted. Protein samples were prepared for sodium dodecyl sulfate polyacrylamide electrophoresis (SDS-PAGE) by boiling for 10 minutes in 1X Laemmli buffer with 2% SDS (Fisher Scientific)[113]. The samples were then loaded onto 15% polyacrylamide gels and subjected to 130 V for 70 minutes. For samples analyzed by Coomassie staining, the gels were stained according to the method of Studier. For samples analyzed by western blotting, the samples were then blotted from the gels to polyvinylidene fluoride membranes (Millipore) following standard procedures[114], at 4 °C using 100 V for one hour in a wet transfer apparatus (Mini Trans-Blot Cell, Bio-Rad) in transfer buffer (25 mM Tris, 190 mM glycine, 20% methanol, pH 8). Western blots were probed with either anti-FLAG primary monoclonal antibodies produced in mice (Sigma, product numbers F3165) against the incorporated C-terminal FLAG tags on DH. After washing, the blots were then probed using anti-mouse secondary antibodies conjugated to horseradish peroxidase (HRP) (Sigma). To control for lysis, the blots were incubated with anti-GroEL primary antibodies produced in rabbit (Sigma) and then anti-rabbit secondary antibodies conjugated with HRP (Sigma). Detection was performed with SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific) for whole culture lysate samples over a total exposure time of 10 minutes on a Bio-Rad Chemidoc Imager. Either the last image, or the last image before the Imager detected pixel saturation, was used. Relative quantification was performed using DW01/pLac-*hilA* pSicA-DH without addition of SipD as the standard. For secretion fraction samples, detection was performed with SuperSignal West Femto Chemiluminescent Substrate (ThermoFisher Scientific). Quantification was carried out using the Bio-Rad Image Lab software. The resulting densitometry was then normalized to the final $OD_{600}$.

### Transcriptional and translational arrest

Overnight cultures of *S. enterica* DW01/pLac-*hilA* pSicA-DH were grown aerobically as described above. These cultures were then subcultured at an $OD_{600}$ of 0.4 for 1 hour at 225 rpm and 37°C. The cells were centrifuged at 4000 g for 10 minutes and washed twice in LB-Lennox to remove previously secreted proteins. The washed cells were next resuspended in LB-Lennox to an $OD_{600}$ of 0.6 and supplemented with 1 mM of SipD, 1 mM of IPTG (Fisher Scientific), and either 100 μg/ml rifampicin (Fisher Scientific) or 8 μg/ml tetracycline (Fisher Scientific) to inhibit

transcription and translation, respectively. The resuspended cultures were grown for 4 hours at 225 rpm and 37°C. The cells were then pelleted at 4000 g for 10 minutes. The supernatant was collected for analysis and concentrated 25-fold by spin concentration using a 30 kDa MWCO membrane (Sartorius). The concentrated supernatant was analyzed by SDS-PAGE and western blotting as described above.

## RNA and cDNA preparation

2 $OD_{600}$ of cells were spun down and resuspended in 500 µL of Trizol. The samples were then shaken vigorously after the addition of 100 µL chlorofoam and left at room temperature for 2 minutes. The samples were then centrifuged at 12000g for 15 minutes at 4°C to phase separate. The resulting aqueous phase was isolated and placed into a new tube. 250 µL of isopropanol was added to the aqueous phase and left at room temperature for 10 minutes. The mixture was then spun again at 12000 g for 10 minutes at 4°C to phase separate. The supernatant was decanted, leaving behind a faint RNA pellet. The pellet was then washed with 500 µL of 70% ethanol and spun at 7500 g got 5 minutes at 4°C. The wash was decanted, and the pellet allowed to dry to remove any residual ethanol. The pellet was then resuspended in 30 µL of RNAse-free water.

Genomic DNA was further removed by treating 3 µg of RNA with RQ1 DNAse at 37°C for 1 hour. 1 µL of RiboLock was added. The reaction was stopped by using the provided Stop solution. The DNAse-treated total RNA was then reverse transcribed with iScript cDNA synthesis kit to make cDNA.

## Quantitative reverse transcription PCR

qPCR was carried out using the SsoFast Evagreen supermix on the BioRad CFX96 system using the primers found in Table 2.1. A melt curve at the end of the cycle was used to determine the specificity of the amplification. Contamination of gDNA was determined using a no RT control. Fold change was calculated using the extended $\Delta\Delta C_t$ method[115]. The $C_t$ values of *glnS* and *eno* were used to calculate the mean $C_t$ for housekeeping gene.

## Size exclusion chromatography

SEC was carried out on an Agilent HPLC using Yarra column. 5 µL of the purified SipD or SEC standards were injected for each run. Each sample was run on the HPLC for 30 minutes at 1 mL per minute flow rate. The mobile phase used was 0.22 µm filtered 1XPBS.

## Primers used in Chapter 2

| Primer for cloning SipD truncations | |
|---|---|
| sipD Δ1-96 F | AGTATCGGTCTCCCATGCACCACCACCACCACCAC |
| sipD Δ1-96 R | AGTATCGGTCTCACTTA TCCTTGCAGGAAGCTTTTG |
| sipD Δ134-343 F | AGTATCGGTCTCCCATGCACCACCACCACCACCACCTTAATATTCAAAATTATTCCGCTTTTATCCGCAGAACAAAATGAGAAC |
| sipD Δ134-343 R | AGTATCGGTCTCACTTAAACAACCATTTCTGATGCTGAG |
| qPCR Primers | |
| eno F | ATCGAGAAAGGCATCGCTAA |
| eno R | CGTCTTTCGCCATCTTGATT |

| | |
|---|---|
| glnS F | TGCTGGACAACATCACCATT |
| hilD F | AGGAGCGCGTTTACAACATT |
| hilD R | AAGTTTCCGTTTGAGCGTTG |
| hilA F | ATGCGATTAAGGCGACAGAG |
| hilA R | GCAAACTCCCGACGATGTAT |
| invF F | CCGATAAATGGGTTTTGCTG |
| invF R | GCCGGAGAAGGCGTAATAAT |
| invA F | TGTCACCGTGGTCCAGTTTA |
| invA R | CTGTTTACCGGGCATACCAT |
| prgH F | TTATCCGCAGCTGGCTTATT |
| prgH R | TTCTTGCTCATCGTGTTTCG |
| prgK F | GGATTCGCTGGTATCGTCTC |
| prgK R | GCCCTCCATCGTCTGTAATG |
| sicA F | AATGCGTAAGGCAGCAAAAG |
| sicA R | TAGCGCCTCCAGATAGACCA |
| sipA F | GCGTAACCAGCAAGAGCATT |
| sipA R | TTCACAATCTCTGCCGTCTG |

**Table 2.1 Primers used in Chapter 2**

## Chapter 3 Harnessing Genetic Diversity for Heterologous Secretion

The most popular strains of *S. enterica* used for biotechnological purposes have been derived from *S. enterica* serovar Typhimurium LT2. However, there is a huge amount of natural genetic diversity in *S. enterica* with over 2000 serovars of *S. enterica*[116–118]. Even within the serovar of Typhimurium, there are at least 80 different strains[119]. Different strains of *S. enterica* give rise to varying levels of invasiveness and severity of salmonellosis[87,88,117,120]. This genetic and phenotypic diversity can be harnessed to identify key cellular factors for heterologous secretion and thus improve secretion titer.

Specific cellular factors or specific transcriptional program, such as those that led to the diverse phenotypes of the different serovars, may be key for heterologous secretion by SPI-1 T3SS. Currently, the complete set of cellular factors that is necessary and sufficient for SPI-1 mediated heterologous secretion is not well defined. The SPI-1 operon was recently refactored and the regulatory control of the entire machinery was placed under synthetic control[44]. However, the refactored SPI-1 operon still required the use of *S. enterica* serovar Typhimurium DW01, a LT2 derivative, as a host for heterologous secretion. This suggests that heterologous secretion required a specific cellular environment that may not be found in *E. coli*.

The regulation of SPI-1 has been extensively studied as SPI-1 is essential for pathogenicity, with numerous cellular factors outside the island identified as inputs to SPI-1 activation and repression[64,67,71,121–124]. However, as discussed in Chapter 2, pathogenicity does not necessarily correlate directly to secretion titers – the deletion of the tip protein SipD resulted in reduced pathogenicity but increased heterologous secretion[63]. Whilst heterologous protein secretion titer is a more continuous trait (i.e. we have observed protein titers can range from 1 μg/mL to 400 mg/mL), these studies focus on fairly discrete phenotypic traits such as whether *S. enterica* can infect host cells or not, and on the presence or absence of native effectors in the supernatant.

A full understanding of cellular factors that impact T3SS protein secretion will lead to improvements in the robustness of the platform, decreasing costs associated with its use in industry. Here we describe the differences in secretion capability between the different strains of *S. enterica* as well as other phenotypic differences. We also report the sequences of several representative strains, which will give insight into the genes potentially responsible for such phenotypic variations. Finally, we report efforts to increase the robustness of heterologous secretion through the removal of selfish genetic elements by strain engineering.

## Genetic diversity of the different S. enterica strains

Whole-genome resequencing allows for better determination of the evolutionary relationship of different strains of *S. enterica* used for biotechnological purposes. The *S. enterica* serovar Typhimurium strains available in the lab, DW01 and LT2, were sequenced by Illumina sequencing and compared to the publicly available sequence of *S enterica* LT2 from NCBI[119]. We also chose SL1344 for this study as it is commonly used in studies on SPI-1 T3SS. LT2 and SL1344 differ in their plasmid content – LT2 has the pSLT plasmid while SL1344 has 3 plasmids, pCol1B9, pRSF1010 and pSLT. *S. enterica* serovar Typhi Ty21a was chosen as an outgroup that is evolutionarily distinct from LT2 and SL1344 (Figure 3.1a) and harbors no known plasmid[44,65,118,125]. Both SL1344 and Ty21a were obtained from ATCC and thus not sequenced.

Surprisingly, we found that our derivative of DW01 differed from the published sequence of DW01, as well as from the published sequenced of its parent strain LT2. We therefore proposed to rename the derivative of DW01 in our lab as AS12 (ActivelySecreting12mutations). Numerous structural variations that were presented in the published DW01 genome were not observed in AS12[44]. The differences to the LT2 reference genome are discussed below.

The genomes of DW01 and AS12 are evolutionarily similar to the publicly available NCBI LT2 reference genome as seen in Figure 3.1a. The evolutionary distance between the LT2-derived strains is much smaller (approximately 10X) than the distance between LT2 and SL1344, as can be seen in the branch lengths (Figure 3.1a). The divergence observed here could be due to random genetic drift.
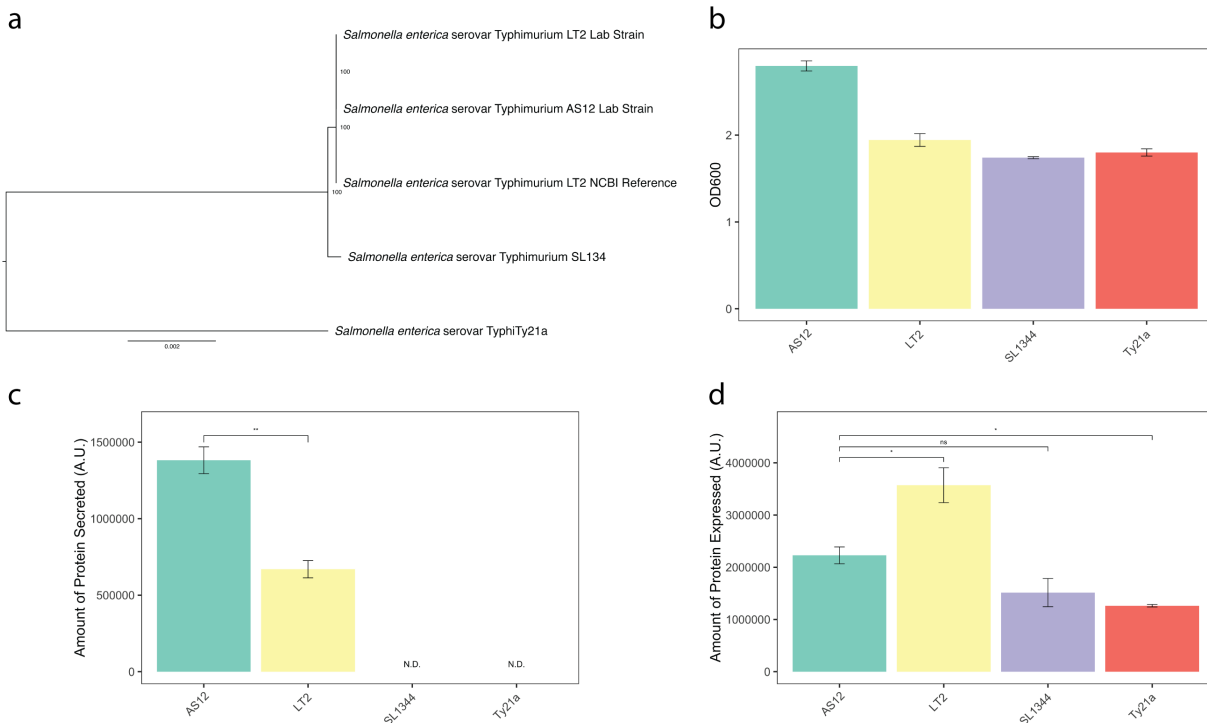


**Figure 3.1 The evolutionary distance between the different strains of *S. enterica* used in this study is shown here (a). (b) AS12 grew to a much higher OD$_{600}$ compared to the other strains which may suggest that it had become more lab adapted. Western blotting was used to determine the amount of protein secreted (c) and the total protein expressed (d), and the protein levels were normalized to the respective OD$_{600}$.**

## Phenotypic differences among S. enterica strains

We first set out to establish the secretion titer phenotype for the four selected strains. To do so, all four of our lab strains – AS12, LT2, SL1344 and Ty21a – were transformed with both a hilA overexpression plasmid and an export plasmid as previously used in Metcalf et al. 2014[43]. The hilA overexpression plasmid upregulates the expression of the T3SS transcription factor HilA and activates the expression of SPI-1. This leads to improved secretion titer of a protein of interest (POI) that is encoded on the export plasmid. The POI used in this study is the DH domain of the human intersectin, a model protein for heterologous SPI-1 T3SS secretion. The POI was fused to the first 167 amino acids of the *S. enterica* effector protein, SptP, to allow for secretion through the SPI-1 T3SS. The resulting strains harboring these two plasmids were then grown and induced

for protein secretion, and samples were taken after 6 hours to assess protein expression levels as well as secretion titer in the media via western blotting of the FLAG tag on the POI. To assess lysis, we also monitored the concentration of cytosolic protein GroEL in the supernatant.

AS12 reached the highest cellular density (as measured by absorption at 600 nm, or $OD_{600}$) as compared to the other strains (Figure 3.1b). The higher cell mass will allow for potentially higher secretion titer and is a beneficial trait for biotechnological purposes. The growth rate of AS12 is comparable to LT2 (Figure 3.4b).

AS12 secreted more POI than LT2 (Welch t-test $p<0.01$) as seen in Figure 3.1c after normalizing for difference in OD600. There was no detectable secretion of the POI from SL1344 and Ty21a, even when blotted against 1/50 of the AS12 secretion fraction. There was significantly less POI detected in the whole culture lysate of AS12 compared to LT2 (Welch t-test $p<0.05$) and more POI in AS12 compared to Ty21a (Welch t-test $p<0.05$) after normalizing for differences in OD600 (Figure 3.1d). The contribution by cell lysis in releasing the POI into the supernatant was negligible, as the concentration of the cytoplasmic protein GroEL in the supernatant was much lower than the total concentration of GroEL in the whole culture lysate. Therefore, differences in secretion detected here cannot be solely attributed to cellular lysis.

## Cellular factors causing differences in heterologous secretion between LT2 and AS12

We focused on the strain with the highest heterologous secretion titer, AS12. There were 12 mutations detected in AS12 when compared against the NCBI LT2 reference genome (Figure 3.2a). Most of the mutations were single nucleotide polymorphisms (SNPs). These differences can be seen in Figure 3.2a. A single deletion resulted in the partial deletion of *gppA* and *rhlB*. We used RNA sequencing (see Chapter 5) to confirm that transcripts of the fusion are present, which would support the genome sequence annotated here. Indeed, the expression of fusion gene, *gppA:rhlB*, was detected in the transcriptomics data at 3 hours and 6 hours. Most of the other mutant alleles transcripts were observed in the transcriptomics data as well (Figure 3.3).

<table>
<tr><th colspan="3">Mutations unique to AS12</th></tr>
<tr><th>Gene</th><th>Amino Acid Change</th><th>Gene details</th></tr>
<tr><td>dcuC</td><td>Gly29Asp</td><td>anaerobic C4-dicarboxylate transporter</td></tr>
<tr><td>yeaG</td><td>Lys116fs</td><td>PrkA family serine protein kinase</td></tr>
<tr><td>rnb</td><td>Trp234*</td><td>exoribonuclease II for mRNA degradation</td></tr>
<tr><td>hnr</td><td>Val102Gly</td><td>two-component system response regulator of RssB</td></tr>
<tr><td>motA</td><td>Pro271Leu</td><td>flagellar motor stator protein</td></tr>
<tr><td>mglA</td><td>Val93Glu</td><td>galactose/methyl galactoside ABC transporter ATP binding protein</td></tr>
<tr><td>malQ</td><td>Leu96Arg</td><td>4-alpha-glucanotransferase</td></tr>
<tr><td>malP</td><td>Gln460fs, Leu468Ser</td><td>maltodextrin phosphorylase</td></tr>
<tr><td>gppA</td><td>deletion</td><td>guanosine-5'-triphosphate,3'-diphosphate pyrophosphatase</td></tr>
<tr><td>rhlB</td><td>deletion</td><td>ATP-dependent RNA helicase</td></tr>
<tr><th colspan="3">Mutations shared with LT2 Lab Strain</th></tr>
<tr><th>Gene</th><th>Amino Acid Change</th><th>Gene details</th></tr>
<tr><td>fimH</td><td>Gly61Ala</td><td>fimbrial adhesin</td></tr>
<tr><td>STM1747</td><td>Arg33Gln</td><td>hypothetical protein</td></tr>
</table>

**Figure 3.2 The mutations found in AS12 are shown here in (a). The single bp deletion in yeaG led to a frameshift (fs) mutation. Western blotting was used to determine the amount of protein secreted (b), normalized to the respective $OD_{600}$. Pairwise Welch t-tests were used to determine if changes in secretion are significant. An asterisk indicated $p < 0.05$, and two asterisks indicated $p < 0.01$.**
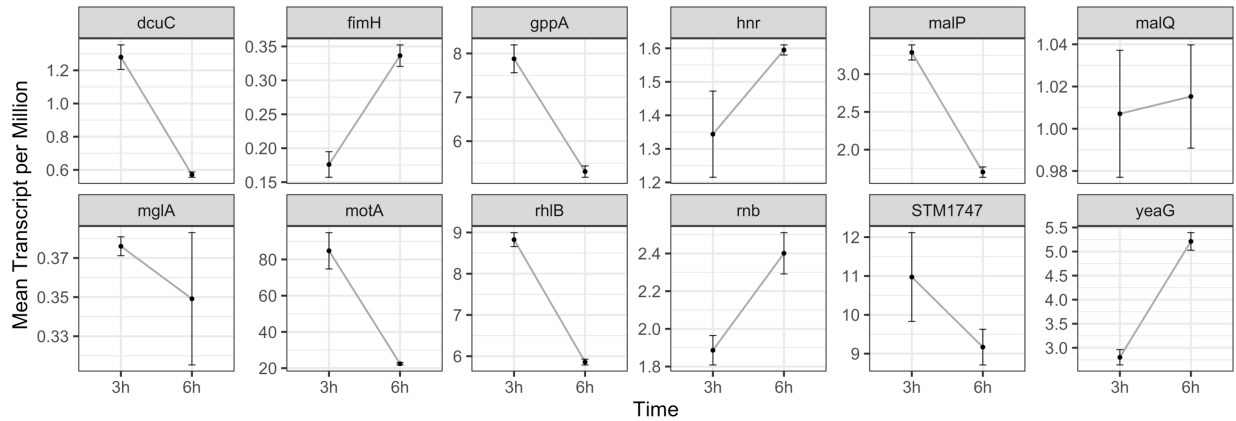
**Figure 3.3 Transcript levels of the different mutated genes are shown here. High throughput RNA sequencing was used to determine transcript levels.**

To determine the genetic basis of the observed differences in expression, secretion, and growth between AS12 and LT2, we replaced each variant allele of AS12 that was unique to AS12 with the reference LT2 allele and assessed the effect on each trait. The *mal* operon (with 3 SNPs) and the *gppA:rhlB* deletion were treated as single alleles.

We observed significant differences in secretion between AS12 *rnb*^LT2 and AS12 *yeaG*^LT2, AS12 *rnb*^LT2 and AS12 *mal* operon^LT2, AS12 *motA*^LT2 and AS12 *yeaG*^LT2, and AS12 *motA*^LT2 and AS12 *mal* operon^LT2 (Figure 3.2b). However, none of the AS12-based strains with replaced LT2 alleles secreted significantly less of the POI when compared to AS12, and POI expression remained largely similar among all the strains. As heterologous protein secretion is a continuous trait and can be potentially affected by multiple genes, it is possible that there are epistatic effects among different variant alleles that allowed for higher heterologous secretion in AS12.

None of the allelic swaps resulted in a final $OD_{600}$ at 6 hours that was similar to LT2 (Figure 3.4a). Most of the single mutations in AS12 did not affect growth rate, but *motA*^LT2 reached a lower cell density at 6 hours (Figure 3.4a).
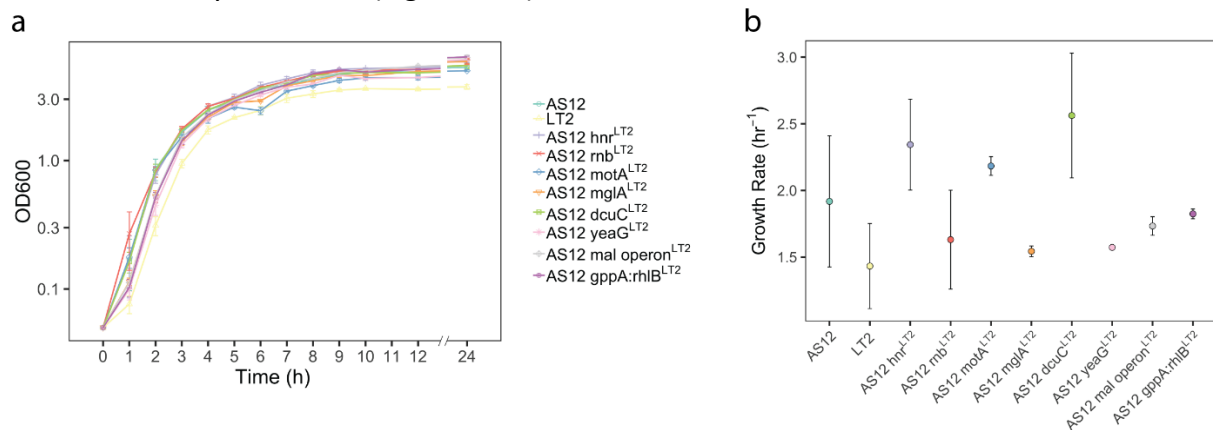


**Figure 3.4 The growth profile of all the strains were largely similar (a) except for AS12 *motA*^LT2 which had an unexpected but reproducible dip at 6 hours. The fitted growth rates (b) are similar between the different strains. The lag time for LT2 was slightly longer than AS12 and hits a lower maximum $OD_{600}$.**

## Contribution of pSLT to phenotypic differences between LT2 and AS12

Transcriptomics data on AS12 showed that numerous genes encoded by the pSLT plasmid were expressed at 3 and 6 hours (Figure 3.5). Therefore, to determine if the changes in secretion were solely due to differences in the genome, both strains were cured of pSLT and tested for secretion. The deletion of pSLT did not result in differences in the secretion and expression of the POI, or in growth (Figure 3.6). The phenotypic differences observed between AS12 and LT2 are most likely the result of differences in the genome. The plasmid copy number of pSLT in both strains was similar (Figure 3.6d). In addition, the removal of pSLT did not significantly alter the transformation efficiency.
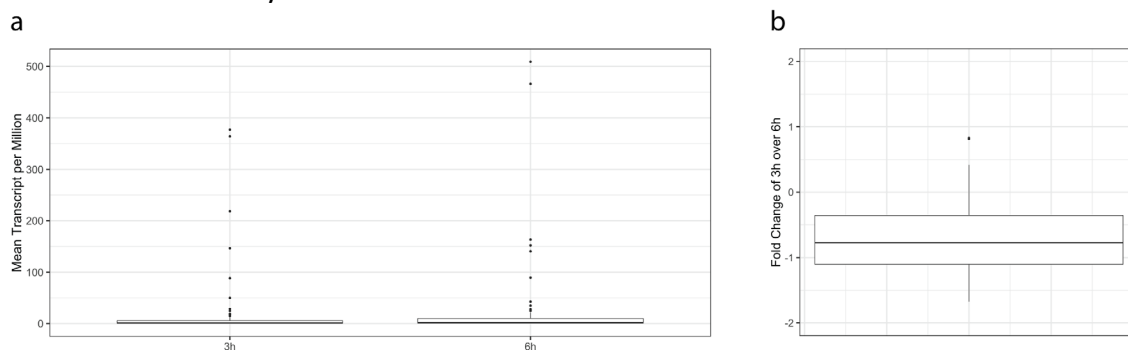


**Figure 3.5 Box and whisker plot of all the genes located on the pSLT plasmid (a) and the fold change of each gene in (b).**
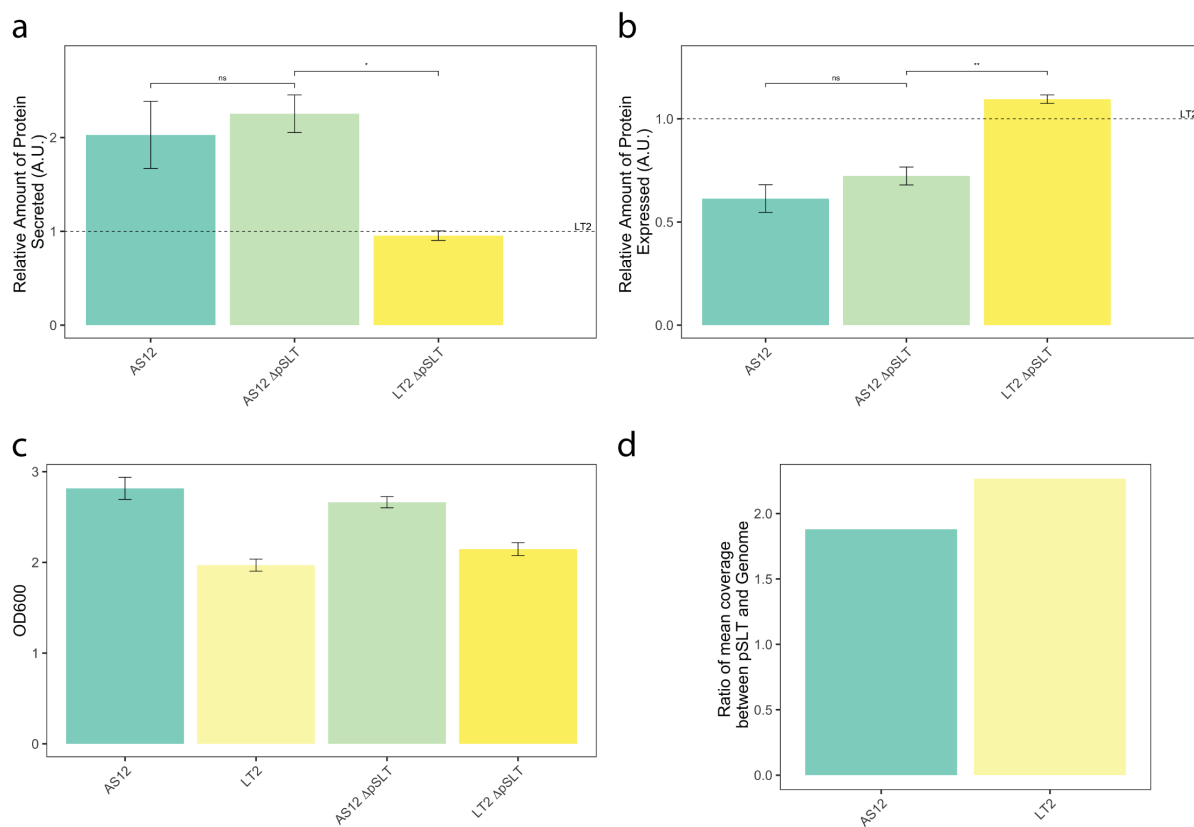


**Figure 3.6 The deletion of pSLT did not result in differences in secretion (a) or total expressed protein (b) as determined by western blotting. (c) The final OD$_{600}$ after 8 hours of induction was also not affected by the plasmid. (d) Copy number of the plasmid was similar between the two strains. An asterisk indicated p < 0.05, and two asterisks indicated p < 0.01.**

24

## Adaptive laboratory evolution of AS12 and LT2

As one of the major phenotypic differences between AS12 and LT2 is the final cell density, we performed an adaptive laboratory evolution (ALE) on LT2 and AS12 as described in the Materials and methods section. The continual passaging of the cells leads to a population bottleneck that creates a selection pressure for clones with better growth. Individuals with better growth will dominate the population, leading to a higher chance of being represented in the subculture. By continually passaging the cells to achieve higher growth, we hoped to replicate the phenotypic divergence observed between the two strains, especially the difference in final $OD_{600}$. An increase in final $OD_{600}$ was observed by day 30; for the evolved LT2 strains, this was comparable to that of DW01.

Using Illumina sequencing, we were able to determine the genomic differences between the different clones from the ALE. Each clone was denoted by the parental strain, the lineage, the day of ALE and then clone. Due to the limitations of the sequencing facility available, structural variants (SVs) that are changes in the genome larger than 1kb, such as large inversions, large indels and translocations were not conclusively resolved. SV calls made by BreakDancer were not supported by many reads[126]. The only deletion that was resolved and confirmed by Sanger sequencing was the deletion in the LT2 lineage 3, which included numerous flagellar genes listed in Table 3.1.

| Gene | Gene details |
|---|---|
| uvrC | excinuclease ABC subunit C |
| uvrY | two-component system response regulator UvrY |
| STM1948 | hypothetical protein |
| yecF | hypothetical protein |
| sdiA | transcriptional regulator SdiA |
| yecC | L-cystine ABC transporter ATP-binding protein |
| yecS | amino acid ABC transporter permease |
| yedO | D-cysteine desulfhydrase |
| fliY | cystine ABC transporter substrate-binding |
| fliZ | flagellar regulatory protein FliZ |
| fliA | RNA polymerase sigma factor FliA |
| tnpA_2 | IS200 family transposase |
| fliB | lysine-N-methylase |
| fliC | flagellin |
| fliD | flagellar filament-capping protein FliD |
| fliS | flagellar export chaperone FliS |
| fliT | flagellar biosynthesis protein FliT |
| amyA | alpha-amylase |
| yedD | hypothetical protein |
| yedE | putative membrane component of transport system |
| yedF | hypothetical protein |
| STM1967 | putative 50S ribosomal protein |
| fliE | flagellar hook-basal body complex protein FliE |

| | |
|---|---|
| *fliF* | flagellar M-ring protein FliF |
| *fliG* | flagellar motor switch protein FliG |
| *fliH* | flagellar assembly protein FliH |
| *fliI* | flagellum-specific ATP synthase FliI |
| *fliJ* | flagellar biosynthesis chaperone FliJ |
| *fliK* | flagellar hook-length control protein FliK |
| *fliL* | flagellar basal body-associated protein FliL |
| *fliM* | flagellar motor switch protein FliM |
| *fliN* | flagellar motor switch protein FliN |
| *fliO* | flagellar biosynthesis protein |
| *fliP* | flagellar biosynthetic protein FliP |
| *fliQ* | flagellar export apparatus protein FliQ |
| *fliR* | flagellar biosynthetic protein FliR |
| *rcsA* | transcriptional regulator for colanic acid capsular biosynthesis activation |

**Table 3.1 The large deletion in LT2 lineage 3 contains many genes shown here. Most of the genes are involved in flagellar biogenesis.**
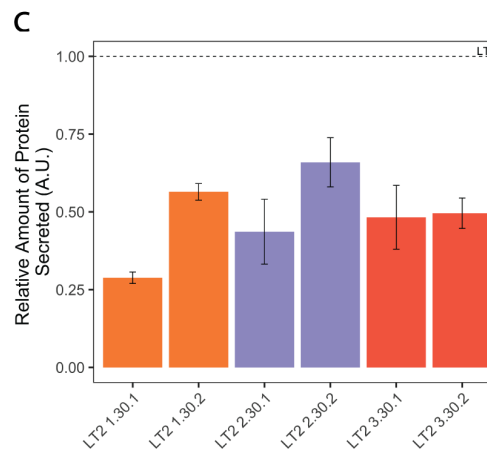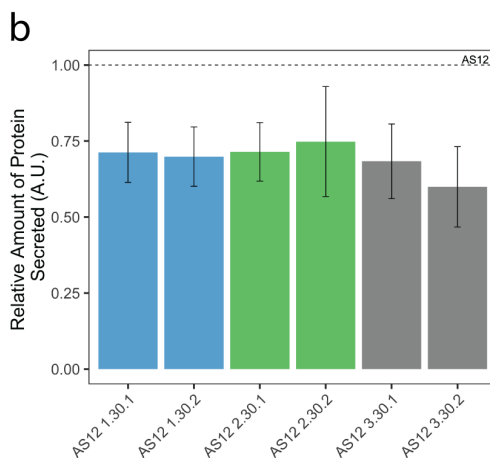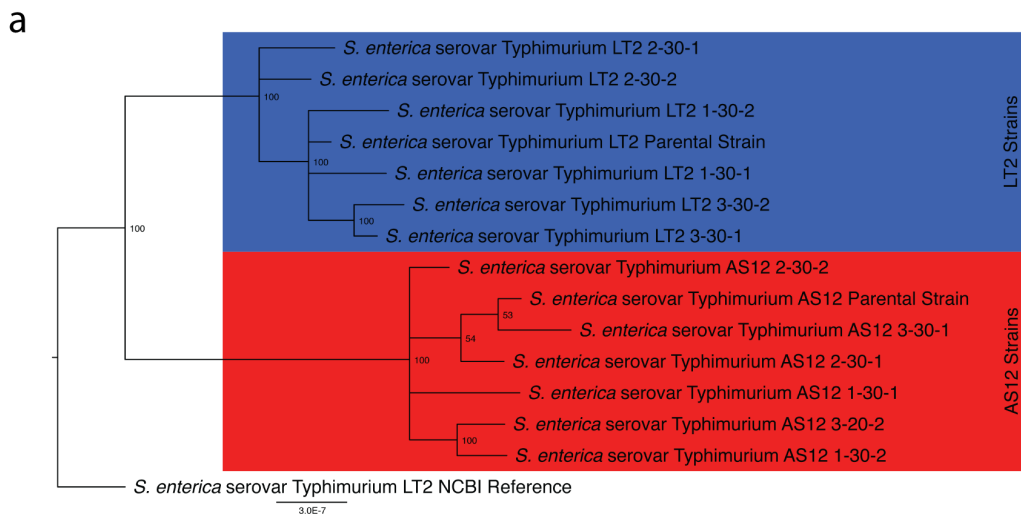


**Figure 3.7 Adaptive lab evolution over 30 days only resulted in small separation in evolutionary distance (a). The secretion titers of the evolved strains were all significantly lower than the parental strain for AS12 (b) and LT2 (c).**

26

The evolutionary distance between the clones and the parental strain is smaller than the evolutionary distance between AS12 and LT2 (Figure 3.7a) but could be underestimated due to the omission of potential genomic rearrangements. The branch lengths here are an order of magnitude smaller than the branch length between the different reference strains in Figure 3.1a. Lineage 1 and 3 of AS12 surprisingly did not form a monophyletic clade. This could be due to lack of initial diversity, convergent evolution or purifying selection.

The mutations found in each clone can be found in Table 3.2 and Table 3.3 for intragenic regions and intergenic regions respectively. Mutations to many genes in the flagellar system were observed in numerous strains in addition to the large deletion of LT2 lineage 3. This included *flhC* (3 clones), *flhD* (1 clone) and *fliH* (1 clone). These mutations leading to higher $OD_{600}$ were consistent with the literature. The adaptive evolution experiment also yielded mutations in two two-component systems: oxygen sensing, *arcA* and *arcB* (5 clones); and osmolarity sensing, *barA*, *envZ* and *ompR* (5 clones). The osmolarity sensing system is known to regulate SPI-1 by modulating the levels of HilD. The most common mutation observed is insertions upstream of *nrdD* observed in 9 clones. The back mutation to the wild-type sequence in the region upstream of *yaeB* in the LT2 lineage 2 suggests that this mutation is not under purifying selection.

| Gene | Amino Acid Change | Strain with mutation | Location | Gene Details |
|---|---|---|---|---|
| STM1829 | Arg51Gly | AS12 1.30.1 | Chromosome | putative cytoplasmic protein |
| *flhC* | Trp69* | AS12 1.30.1 | Chromosome | regulator of flagellar biosynthesis |
| | Cys91fs | AS12 2.30.1, AS12 2.30.2 | | |
| | Gln107* | LT2 1.30.2 | | |
| *flhD* | Gln30* | LT2 2.30.1 | Chromosome | regulator of flagellar biosynthesis |
| *fliH* | Thr11Pro | AS12 3.30.1 | Chromosome | flagellar assembly protein FliH |
| *arcB* | Pro165Ser | AS12 3.30.2 | Chromosome | aerobic respiration sensor-response protein |
| *arcA* | Asn175Thr | AS12 1.30.2 | Chromosome | negative response regulator of genes in aerobic pathways |
| | Asp11Asn | AS12 2.30.1, AS12 2.30.2 | | |
| | Leu80Met | LT2 3.30.2 | | |
| *barA* | Thr268Pro | AS12 1.30.2, AS12 3.30.2 | Chromosome | two-component sensor histidine kinase that activates OmpR by phophorylation |
| | Gly455Cys | AS12 3.30.1 | Chromosome | |
| *envZ* | Ile270Asn | LT2 1.30.1 | Chromosome | osmolarity sensor protein |
| *ompR* | Val82Gly | LT2 1.30.2 | Chromosome | transcriptional regulatory protein for ompC and ompF |

| | | | | |
|---|---|---|---|---|
| *glpR* | Gln46Lys | LT2 3.30.1, LT2 3.30.2 | Chromosome | repressor of the glp operon |
| *cytR* | Tyr98fs | LT2 3.30.1 | Chromosome | regulator for deo operon, udp, cdd, tsx, nupC, and nupG |

**Table 3.2 Mutations were found in these genes in the different strains derived from adaptive lab evolution.**

| Gene downstream of mutation | Mutation | Strain with mutation | Location | Gene Details |
|---|---|---|---|---|
| *gdhA* | 1382158 T>C | AS12 1.30.1 | Chromosome | glutamate dehydrogenase |
| *repA2* | 19885 C>A | AS12 1.30.1 | pSLT | replication regulatory protein |
| *nrdD* | 4697694 insA | AS12 1.30.1, AS12 1.30.2, AS12 2.30.1, AS12 2.30.2, AS12 3.30.1, AS12 3.30.2, LT2 3.30.1, LT2 3.30.2 | Chromosome | anaerobic ribonucleoside-triphosphate reductase |
| | 4697693 insT | LT2 2.30.1 | | |
| *proS* | 289100 G>T | LT2 2.30.2 | Chromosome | proline-tRNA ligase |
| *yaeB* | 289293 G>T (reversion) | LT2 2.30.1, LT2 2.30.2 | Chromosome | tRNA (N6-threonylcarbamoyladenosine(37)-N6)-methyltransferase TrmO |

**Table 3.3 There were numerous mutations in the intergenic regions. The genes shown here could have their expression altered due to these mutations.**

## Heterologous secretion of evolved clones

We next set out to compare the secretion titers of the ALE strains as compared to their respective parental strains using the secretion assay described earlier with a FLAG-tagged POI and western blotting. Following the ALE of AS12, none of the clones secreted significantly less proteins compared to the parental strain (Figure 3.7b). In contrast, following the ALE of LT2, all of the clones except LT2 2.30.2 secreted significantly less than the parental strain (Figure 3.7c). An ANOVA model with lineage and lineage:clone as explanatory variables was evaluated using a TukeyHSD test. Through this analysis, the differences in secretion titer could be observed between the different lineages of AS12 and LT2 were statistically significant.

## Deletion of prophage improves secretion

The genotypes of the ALE strains demonstrate that active selfish genetic elements (SGEs) in *S. enterica* can drastically remodel the genome. SGEs enhance their own transmission often at the cost of the host genome through various means such as horizontal gene transfer or transposition[127,128]. Although SGEs can be a significant source of driving mutations for evolutionary change and innovation, their persistence can also lead to increased clonal variations[127]. Even benign SGEs can have an associated cost to them[129,130]. The large deletion found in the LT2 lineage 3 is likely mediated by the transposase *tnpA_2* found in the region of the deletion. The other potential SVs supported by low numbers of reads could represent a subpopulation of cells with altered genome architecture.

In addition to the plasmid pSLT, prophages are SGEs that can spread easily in a population. The activation of prophages into actively virulent phages can cause instability to the population and the genome of the production strain[120,131]. The prophages fels-1 and fels-2 had previously been shown to be active during exponential growth[28]. This is supported by the transcriptomics data (Figure 3.8) where genes in fels-1 and fels-2 were being transcribed.
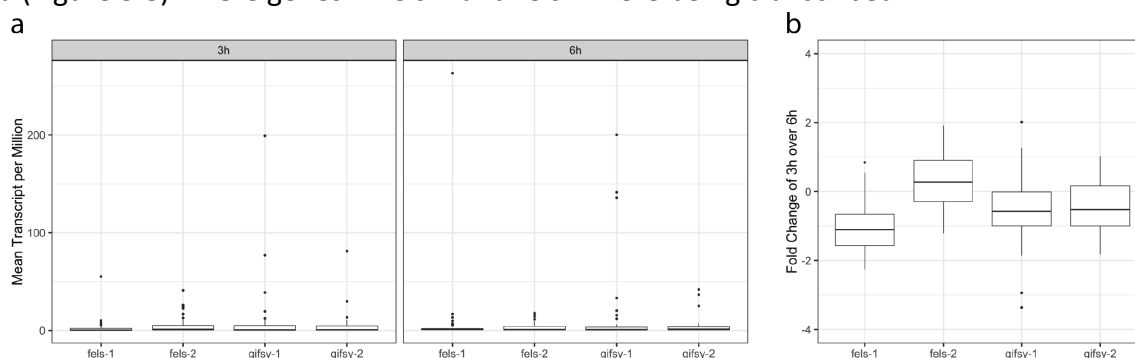


**Figure 3.8 RNA transcript levels of the different prophages are shown here (a). Most of the phages are being expressed during growth. The fold change for all the genes is calculated in (b)**

We were interested in examining the impact of removing these prophages on secretion, so we deleted these prophages using recombineering techniques and then compared secretion titer using the aforementioned assay. Removing fels-1, but not fels-2, improved POI secretion by AS12 (Figure 3.9a) without impacting its cellular expression (Figure 3.9b). The $OD_{600}$ at 6h were similar, suggesting that these prophages may not have a large metabolic cost in culture.
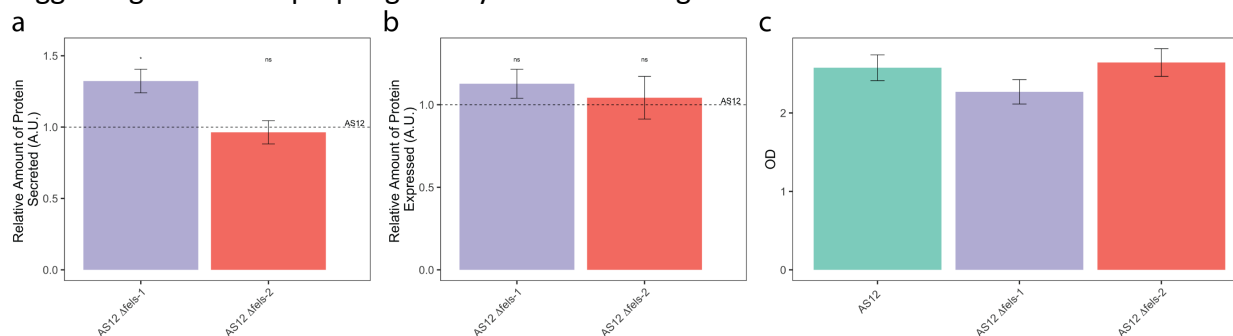


**Figure 3.9 The deletion of fels-1 increase secretion (a) but not fels-2. The total protein expression did not change (b) nor did the final $OD_{600}$ differ significantly (c).**

## Discussion

This study showed that there is a significant difference in heterologous secretion even in highly related strains of *S. enterica* serovar Typhimurium. Yet no differences were found secretion-related genes between the *S. enterica* serovar Typhimurium strains AS12 and LT2 used in this study. Moreover, the lack of detectable secreted POI in SL1344 and Ty21a was surprising, as previous studies show that natural effectors can be detected in the supernatant under various conditions[32,49,132]. Despite fusion to a native secretion tag, the POI is biochemically different from the effector proteins, which could lead to decreased secretion in SL1344, a highly infective strain. In addition, the POI was expressed at a higher level as compared to the native effectors as it is ectopically expressed from a high-copy plasmid.

The potential for epistatic interactions between the different genes is expected for complex traits such as heterologous secretion and growth. It is possible that undetected SVs in the AS12 genome could be causal for higher secretion titer. Systematically exploring the impact of individual SVs on secretion, expression and growth is challenging with the current genome editing methods available. The ability to genetically cross different strains through phage transduction using negative selection could allow us to determine the role of genome structure on these complex traits in a high-throughput manner[133]. Furthermore, larger swaps of DNA can be achieved by combining the stringent negative selections[133] with Hfr-mediated conjugation or transduction to create definite genome blends and allow us to investigate the role of genome structure on these complex traits.

Genes that encode factors known to affect SPI-1 expression were found to be mutated during the course of the adaptive lab evolution, demonstrating that these factors are not important for laboratory-based culture. For example, changes to the osmolarity sensing system (observed in LT2 1.30.1 and LT2 1.30.2) can affect the levels of HilD in the cell but the synthetic induction of the downstream regulator, HilA, likely circumvented any regulatory inputs at the level of HilD. Moreover, SPI-1 is natively expressed in environments with low oxygen[134]. Mutations found in the *arc* operon could have altered the expression of SPI-1 genes were observed in numerous ALE strains - AS12 1.30.2, AS12 2.30.1, AS12 2.30.2, AS12 3.30.2 and LT2 3.30.2. As all of these ALE strains had lower secretion titers compared to their respective parental strain, we hypothesized that these systems still play an important role in altering heterologous secretion titers even when *hilA* was overexpressed.

Although the strains used here represent a fair amount of genetic diversity, they constitute merely a small fraction of the *S. enterica* strains available. By comprehensively screening more strains for their secretion capability and comparing their genomes, we can potentially uncover new novel inputs to the regulation of SPI-1 for heterologous production and pathogenicity. Furthermore, the diversity in genome architecture could provide a clue to how arrangement of the genome could impact the expression of regulons or gene clusters such as SPI-1.

The fair stability of AS12 genome with few SNPs and small indels during ALE is encouraging for the development of a stable production strain. Successful continuous fermentation of AS12 for industrial protein production will need to have low clonal variation. Efforts to increase the stability by removing SGEs in this work will help increase the robustness of the production strain and reduce the impact of environmental fluctuations that might inactivate synthetic circuits used for heterologous protein secretion[135,136]. The increased heterologous secretion observed by deleting fels-1 is an encouraging sign that creating a stable production strain of *S. enterica* for

heterologous secretion is possible. Removing these SGEs can have fitness benefits as cellular resources and energy are freed up in the cell[129,136].

## Materials and methods

### Strains, media, growth and harvest of bacteria

*Salmonella enterica* strains were grown overnight in LB Lennox (Dot Scientific Catalog no. DSL24066-500) with the appropriate antibiotics as per Metcalf 2014 at 37°C and 225 rpm in 24-well blocks. The overnights were sub-cultured to an OD of 0.05 in 24-well blocks for all experiments (Axygen Catalog no. PDW10ML24C). Cells were electrotransformed except for SL1344 and Ty21a, which were transformed by heat shock using the CaCl method. All the strains harbored the pLac *hilA* plasmid and an export plasmid with DH as the POI.

For secretion assay, secretion was induced by the addition of 0.1 mM IPTG (Dot Scientific Catalog no. DSI56000-5) at the point of subculture for 6 hours at 37°C and 225rpm. Antibiotics were added as needed. At the point of harvest, 20 µL of the culture was added to 40 µL of 4X Laemilli buffer for the whole culture lysate (WCL) samples. The culture blocks were then spun at 4000 g for 10 minutes to collect the supernatant (Sup) fractions. 40 µL of the supernatant was added to 16 µL of 4X Laemilli buffer.

The growth curves were generated using strains with no plasmids. Samples were taken out every hour. At each sampling <100 µL of culture was taken out to be measured for $OD_{600}$ on the Nanodrop 2000c using the appropriate cuvettes. Growth curves were non-parametrically fitted to splines and the growth rate determined in R using the growthrates package. All graphs were plotted using the ggplot2 package[137].

### Adaptive lab evolution

3 separate lineages of AS12 and LT2 were passaged every 24 hours for 30 days. For each subculture, the overnight culture was diluted to an $OD_{600}$ 0.05. The strains were grown in LB-L in glass tubes at conditions similar to above. 2 clones from the 30th generation of each lineage was picked for further characterization for growth and secretion. The strains were then sequenced.

### Whole genome resequencing of lab strains

Genomic DNA of the lab strains were isolated using the GenElute Bacterial Genomic DNA kit (Sigma Catalog no. NA2110-1KT). The genomic DNA was then sent to UC Davis Genome Center for library generation. Barcode-indexed libraries were generated from genomic DNA sheared on an E220 Focused Ultrasonicator (Covaris, Woburn, MA). For each sample, ~25 ng sheared DNA were converted to sequencing libraries using a Kapa Hyper Library Preparation Kit (Kapa Biosystems-Roche, Basel, Switzerland). The libraries were amplified with 8 PCR cycles and analyzed with a Bioanalyzer 2100 instrument (Agilent, Santa Clara, CA), quantified by fluorometry on a Qubit instrument (LifeTechnologies, Carlsbad, CA), and combined in one pool at equimolar ratios. The pool was quantified by qPCR with a Kapa Library Quant kit (Kapa Biosystems-Roche). The resulting libraries were then run on a single lane of Illumina MiSeq PE300 run (Illumina, San Diego, CA). The average read depth came was about 50X. The evolved strains genomic DNA was library prepped using NexteraXT kit and then sequenced by the NUSeq Core on a single lane of

Illumina HiSeq 4000 PE150 run (Illumina, San Diego, CA). The reads were then analyzed with a pipeline modified from the GATK best practices.

## Data processing for genome resequencing

The raw reads were trimmed to remove adaptor sequences using Trimmomatic v0.36 for a minimum quality of 14 in a sliding window of 4bp[138]. Reads less than 36bp were also filtered out. The trimmed reads were then mapped to the *S. enterica* LT2 genome (obtained from NCBI) using the Burrows Wheel Aligner v0.7.12 MEM algorithm[139]. Picard v2.9.2 was then used to sort, mark duplicates and de-duplicated the mapped reads[140]. Variant calling was done by FreeBayes v.1.1.046 with the ploidy set to 1[141]. The variants are filtered by bcftools v1.3.1[142] to have a minimum quality of 30 and then annotated using SnpEff v4.3[143]. Structural variations (SV) were called using BreakDancer v1.3.6 BreakDancerMax algorithm[126]. The SVs were then manually annotated. The deletions were then manually checked in the IGV browser v2.3.90[144] and verified by Sanger sequencing with Quintara BioSciences (Table 3.4).

## Generation of strains used

Genomic modifications made in this study were done using the Court lab recombineering method[145,146]. Briefly, strains were transformed with pSIM6 and then grown overnight at 30°C, 225 rpm. The cells were then sub-cultured at a 1:100 dilution and grown for ~2hrs to an OD of 0.4 – 0.8. The lambda red system was then induced at 42°C for 15 minutes and then the cultures were cooled in an ice water bath for 10 minutes. The cells were spun down at 4600 g for 3 minutes at 4°C, washed thrice in ddH$_2$O and finally resuspended in 200 μL of ddH$_2$O. 200ng of DNA products or 5 μL of 10mM primers were electroporated into 50 μL of cells as needed. The DNA products used in round one of recombineering were generated by Phusion PCR using the TUC01 genomic DNA as a template using primer found in Table 3.4. The PCR reaction was then cleaned up with the Wizard SV Gel and PCR Clean-up kit (Promega Catalog no. A9282) prior to electroporation. Round 2 primers can be found also in Table 3.4. Round 2 PCR products for the *mal* operon and *gppA:rhlB* were generated with the LT2 gDNA and the respective sequencing primers.

For round one of recombineering, the cells were recovered in 350 μL of SOC at 30°C for an hour and then plated on LB agar plates with 10 μg/L of chloramphenicol. For round two, the cells were recovered in 10 mL of LB Lennox at 37°C for 4 hours. Serial dilutions of the recovery are then plated on 6% sucrose plates. Patch plating on LB agar plates with 10 μg/mL of chloramphenicol or 30 μg/mL carbenicillin to determine successful recombination and loss of pSIM6 plasmid respectively. Colony PCR was then carried out using GoTaq (Promega Catalog no. M3008) to isolate the gene of interest to be Sanger sequence verified by Quintara BioSciences using the primers in Table 3.4. The verified clones were then electroporated with the relevant plasmids as needed.

To knock out the pSLT plasmid, the round one recombinant was selected on LB agar plates with 34 μg/L of chloramphenicol. Overnights of the success recombinants were grown with increasing concentration of chloramphenicol up till 544 μg/L of chloramphenicol to ensure homozygosity of the modified pSLT plasmid. Primers against the oriC were then used to delete

the plasmid (Table 3.4). pSLT specific primers were used to determine if pSLT was knocked out (Table 3.4).

## Western blotting

Samples were run on a 12.5% SDS PAGE gel at 150V for 60 minutes. 4 µL and 8 µL of samples of WCL and supernatant were loaded respectively. The gels were then equilibrated in Towbin buffer with 20% methanol for 15 minutes. The samples were then transferred to a PVDF membrane (Millipore Catalog no. IPVH00010) using the Owl HEP-1 blotter at 0.3A for 35 minutes. After which, the membrane was blocked in 5% milk in TBST for 1 hour. The membrane was then decorated with anti-FLAG M2 antibodies from mice (Sigma Catalog no. F3165-1MG) in 1% milk (1:6666) overnight, followed by three 5-minutes washes in TBST (0.1% Tween-20). The membrane was then decorated with anti-mouse secondary antibody from goat conjugated with HRP (Thermo Fisher Catalog no. 32430) in TBST (1:1000) for 1.5 hours, followed by three 5-minutes washes in TBST. Blots were then imaged using SuperSignal West Pico (Thermo Scientific Catalog no. 34080) on the Bio-Rad Chemidoc. The secretion titer was then quantified by densitometry using the Image Lab software v5.2.1 provided by Bio-Rad.

The blots were then striped using the mild stripping protocol modified from Abcam to reprobe for GroEL. The membranes were incubated in mild stripping buffer for 8 minutes twice, washed twice in ddH$_2$O for 10 minutes and then twice in TBST for 5 minutes. The stripped blot was blocked in 5% milk in TBST as above. The membrane was processed as above but using anti-GroEL antibodies from rabbit (Sigma Catalog no. G6532-.5ML) in 1% milk (1:10000) and anti-rabbit secondary from goat conjugated with HRP (Thermo Fisher catalog no. 32460) instead.

## Construction of phylogeny with whole genome

Multiple sequence alignment of the genomes was carried out using mugsy v1.2.3[147]. The output was then converted to phylip with a custom script. The phylogenetic relationship was then determined using MrBayes v3.2.7a[148]. The Bayesian model was built using the GTR model with the options nst=6 and rates=invgamma. The standard priors of MrBayes was used and the simulation ran until the standard deviation of the split frequencies was below 0.02. The consensus tree was then viewed using FigTree v1.4.4.

## Transcriptomics analysis

RNA was extracted from the culture of AS12 with pLac *hilA* at 3 hours and 6 hours by resuspending the cell pellet in Trizol (Thermo Fisher catalog no. 15596026). The Direct-zol RNA Miniprep kit (Zymogen catalog no. R2051) was then used to purify the RNA from the Trizol mixture. The isolated RNA was then sent to RTSF Genomics Core for library preparation using the TruSeq Stranded Total RNA Library preparation (Illumina, San Diego) with Ribo-Zero rRNA depletion. The resulting library was then run over two lanes of Illumina HiSeq 4000 SE50 (Illumina, San Diego) with other libraries.

The raw reads were concatenated and then trimmed to remove adaptor sequences using Trimmomatic v0.36 for a minimum quality of 14 in a sliding window of 4bp[138]. Reads less than 36bp were also filtered out. Quantification of the trimmed reads were done using Salmon v0.10.1 with 100 bootstraps with the reference LT2 transcriptome[149]. DESeq2 v1.22.2 was then used to

calculate the log2 fold change using time as a contrast (between 3 hours and 6 hours). Between library normalization was done with the size factors from DESeq2 with a custom script[150]. The normalized transcripts per million were plotted using the ggplot2 package in R[137].

*Primers used in Chapter 3*

| Recombineering Round 1 Primers | |
|---|---|
| dcuC catsacB F | ATCGGAGTCGTCGTTATTGTGGGTGTAGCGCGCTA CATCATTAAGGGATATTCCGCCACTTGTGACGGAAGATCACTTCG |
| dcuC catsacB R | TACCTTATGCCCCATCAGCGCGCTGATAATAAGCAGAGCCAGACCGCCGAC AAATAAAACATCAAAGGGAAAACTGTCCATAT |
| yeaG catGsacB F | GGGCTGGAAGAGAAGAAGCAAATTCTCTATTTGCTGGGACCTGTGACGGAA GATCACTTCG |
| yeaG catGsacB R | CACGCGCTGCATTAGCGATTTCAGCCGTTCAGCGAGCATCAAAGGGAAAAC TGTCCATAT |
| rnb catGsacB F | CAGACTCCAGCTTACGGTGGCTATCGCCGATCCTACCGCCTGTGACGGAAG ATCACTTCG |
| rnb catGsacB R | GAATTTTCGCTGTGTTATCCAGCTTACTGCCTTCGGCAATATCAAAGGGAAA ACTGTCCATAT |
| hnr catGsacB | AGAAAATATGGCGGATATCGCCAAAGCGTTGCGCCTTGGGTGTGACGGAA GATCACTTCG |
| hnr catGsacB R | AGCGGTTGAGATCTTTTACCGGCTTCAGCAAGACATCTTCATCAAAGGGAAA ACTGTCCATAT |
| motA catsacB F | CGTGGAATTTGGTCGTAAAACGCTTTATTCCAGTGAGCGTTGTGACGGAAG ATCACTTCG |
| motA catsacB R | TCACTGCGCGAACGTGTTCTTCCAACTCAATAAACGAATCAAAGGGAAAACT GTCCATAT |
| mglA catsacB F | CCATTCGGCGAAAGAAGCGCTGGAGAATGGGATTTCGATGTGTGACGGAA GATCACTTCG |
| mglA catsacB R | TGACCGAACGTTGTAATACCAGGTTTAACTCCTGGTGATCAAAGGGAAAAC TGTCCATAT |
| mal catsacB F | GGTAGCCCTCCGGCAATGTAGCCGGAAGGTTGTGACGGAAGATCACTTCG |
| mal catsacB R | ACGGAATCAACATAGCAATCACTTCACTGTGGGAAATATCAAAGGGAAAAC TGTCCATAT |
| gppA:rhlB:catGsacB F | GAGCGTAATGCGCTCGTCCATCCCCTGCGCCATCATGATTTGTGACGGAAGA TCACTTCG |
| gppA:rhlB:catGsacB R | CGATTTGGGCTTTATTAAAGATATCCGCTGGCTGTTCATCAAAGGGAAAACT GTCCATAT |
| ccdB catsacB F | CATTGACACCCCCGGACGGCGGATGGCTGTCCCGCTGGTCTGTGACGGAAG ATCACTTCG |
| ccdB catsacB R | GCATCACCGGGTAAAGATCGCGGGGAACCTTTTCCGAATCAAAGGGAAAAC TGTCCATAT |
| fels-1 catGsacB F | AGCAAAAACCGATCTACGATAATCAATTATATCCTTTCAGTGTGACGGAAGA TCACTTCG |

| | |
|---|---|
| fels-1 catGsacB R | ACGATCTCAAACGCATTATCAAGCTGATACGCTATCAATCAAAGGGAAAACTGTCCATAT |
| fels-2 catGsacB F | CTGGCGGCAACAAAAAACCCGCCATTAAAGCGGGTTCTGTTGTGACGGAAGATCACTTCG |
| fels-2 catsacB R | ACATACCATTTTTAGTATAGTAAAAACAGTGTATTGCGTTATCAAAGGGAAAACTGTCCATAT |
| | Recombineering Round 2 Primers |
| dcuC rec F | GCGCGCTACATCATTAAGGGATATTCCGCCACTGGCGTTTTATTTGTCGGCGGTCTGGCTCTGCTTATTA |
| yeaG rec F | AAGCAAATTCTCTATTTGCTGGGACCTGTGGGGGGGGGGCAAATCATCGCTCGCTGAACGGCTGAAATCGC |
| rnb rec F | CAGCTTACGGTGGCTATCGCCGATCCTACCGCCTgGATTGCCGAAGGCAGTAAGCTGGATAACACAGCGA |
| hnr rec R | GAGATCTTTTACCGGCTTCAGCAAGACATCTTCAGCCCCAAGGCGCAACGCTTTGGCGATATCCGCCATA |
| motA rec R | CTGCGCGAACGTGTTCTTCCAACTCAATAAACGATGGACGCTCACTGGAATAAAGCGTTTTACGACCAAA |
| mglA rec R | CCGAACGTTGTAATACCAGGTTTAACTCCTGGTGTACCATCGAAATCCCATTCTCCAGCGCTTCTTTCGC |
| pSLT oriC F | TTGAGATAAATGGAGCAAGTAATGCTGTGTTCATCATGAAGCAGCGAATTACAGTGACAGTGGACAGCGA |
| fels-1 KO F | GATCTACGATAATCAATTATATCCTTTCAGTGATAGCGTATCAGCTTGATAATGCGTTTG |
| fels-2 KO F | GGCAACAAAAAACCCGCCATTAAAGCGGGTTCTGTAGTCCTGGAATCTGTGCTGGCACACCAGAAAGAAG |
| | Primers to amplify PCR product for Sanger Sequencing |
| dcuC seq F | CAGGGGAAACAATGCTAACAG |
| dcuC seq R | GGTTTCGCTGGCAGGTAATA |
| yeaG seq F | GTTTTACGGCATGGAAGACG |
| yeaG seq R | CTTCCTGCGGATTAAACAGG |
| rnb seq F | GACGGCAGACTCCAGCTTAC |
| rnb seq R | AGCGAACACAGATCGTCAGA |
| hnr seq F | CTTACGTAATCGCGGAGACC |
| hnr seq R | CGAAACAGGCGTTCTTCTTC |
| motA seq F | GGTAATGGGGGTGGTTCAC |
| motA seq R | GCGAACGTGTTCTTCCAACT |
| mglA seq F | TATCATTATGGGCAGCACGA |
| mglA seq R | TCCATGACCGAACGTTGTAA |
| mal seq F | GTTCATCCTGGGTCAACGTC |
| mal seq R | CGGAAAGTTTAGCCAACAGC |
| gppA:rhlB seq F | TGTATCGCGCGTTGTTTTAG |

| | |
|---|---|
| gppA:rhlB seq R | CTGATCGACTACGCCAAACA |
| fels-1 seq F | TCTTTGTAGGCGGGAGTGAC |
| fels-1 seq R | TACCACACCGCAATACTCCA |
| fels-2 seq F | CCGAGGATGTAGGAATTTCG |
| fels-2 seq R | ACGTAGTTTCCCCACCACTG |
| Sanger Sequencing Primers | |
| dcuC int seq F | TCGGAGTCGTCGTTATTGTG |
| yeaG int seq F | AGGGCTGGAAGAGAAGAAGC |
| rnb int seq F | TGACGCGCTATATGCTGAGG |
| hnr int seq F | TGGAGTTGATGGGGCGTTTT |
| motA int seq F | TGGATTCATTTCACCGTTAGC |
| mglA int seq F | GATTCCGGCAGCATTGTATT |
| mal int seq 1R | GCTGAGAAGGTGGGTGAAGA |
| mal int seq 2R | ACGCTGGTCGATAACACCT |
| mal int seq 1F | ATCGACTTCCTCTTCCAGCA |
| mal int seq 2F | CAGTTTTGCCCATACCTGCT |
| mal int seq 3R | TGCAGCAATACTTCCAGTGC |
| mal int seq 3F | AAGGCATGTTTATCGCCATC |
| gppA:rhlB int seq F | CAGTTTGGCGAGCGTAATG |
| gppA:rhlB int seq R | GCATGTACGATTTGGGCTTT |
| fels-1 int seq | CGGGATGAGCAAAGAGAGTT |
| fels-2 int seq | TTACGTGAAAACGTAGCCCC |
| pSLT Primers | |
| pSLT parA F | CTTGCTCTGACCAGGGAGTC |
| pSLT parA R | CTGGGATAACCCCTCCAGTT |
| pSLT parB F | CCATCGTAAAGTGCCCAAGT |
| pSLT parB R | AAGCGGGAATAACATCAACG |
| pSLT traI F | GGTTTTGAGTCGGCGTATGT |
| pSLT traI R | GCCTCTTCACTGCCTTTCAC |
| Sequencing Primers for Evolved Strains | |
| uvrC:yod seq F | CACCTTACCCTGGCGAATAA |
| uvrC:yod seq R | CTACATCGACGCTCACTTCG |
| uvrC:yod seq int F | AAACAGCACATGCACACAGG |

**Table 3.4 Primers used in Chapter 3**

## Chapter 4 Modulating HilD levels in Salmonella enterica for heterologous secretion

Regulation of the T3SS SPI-1 system is complex, with a cascade of transcription factors involved, as detailed in Chapter 1. Controlling this regulation is critical for enabling the T3SS as part of a protein production process. Illustrating this point, the production strain with the highest titer currently reported in the literature employs the synthetic induction of the SPI-1 regulator, HilA[43,63]. Upon overexpression of *hilA*, various SPI-1 operons are activated; and functional needle complexes are assembled. However, *hilA* is not the master node of the complex regulation cascade in SPI-1. The expression of HilA is also not the commitment step of the cell to the SPI-1 program. Instead, there are several lines of inquiry that implicate HilD in these roles. The speed at which *hilD* is expressed is key to the bet-hedging lifestyle strategy in *S. enterica*[46,61]. In addition, HilD binds to promoter regions of various genes outside the SPI-1 locus and can drastically remodel the cellular physiology while HilA's main targets are found within the SPI-1 locus[59,151]. Although *hilD* and *hilA* overexpression are common strategies employed in the field to synthetically induce SPI-1 for pathogenicity studies, *hilD* overexpression is still a promising, as-yet untested target for heterologous secretion.

The commitment of a cell to the SPI-1 pathogenicity program is conditional on HilD levels crossing a threshold and turning on various operons[45,46]. As such both HilD protein levels and HilD activity are tightly regulated by various factors at the level of transcription and translation, and post-translation modificaitons[75]. The transcription factors, HilC, RtsA, and HilD form a complex feed forward loop in which HilC and RtsA act as signal amplifiers for HilD[46,61]. Other regulators such as Fur can also trigger the autoinduction of HilD[73]. Once the RNA transcript is made, *hilD* is not always successfully translated. The availability of the *hilD* transcript for translation is modulated by factors such as Hfq and sRNAs such as FnrS and ArcZ[64,65,134]. The translational regulation is mediated by sequences found in the 5'UTR and 3'UTR[152,153]. Even when *hilD* is successfully translated, the HilD protein can be bound by proteins such as HilE and FliZ and prevented from binding to its cognate promoter[154,155].
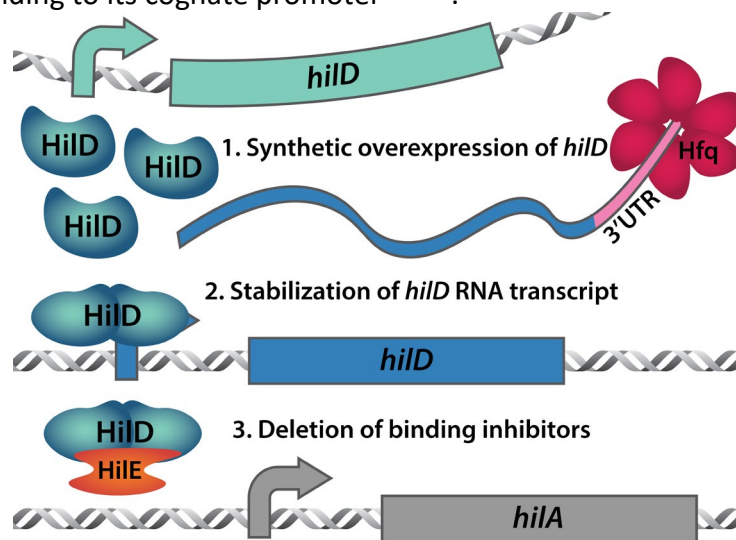


**Figure 4.1 HilD is controlled at many levels shown here. The corresponding strategies that are employed to modulate cellular HilD levels in cells in this work are also depicted.**

In this chapter, we demonstrate the three strategies that we employed to modulate HilD levels or activity, triggering the heterologous secretion of the protein of interest as shown in Figure 4.1. Three different strategies were employed: 1) the synthetic induction of *hilD* in a fashion similar to previously reported work on *hilA*, 2) stabilizing the *hilD* mRNA transcript to allow for increased translation levels, and 3) deleting binding partners that can negatively regulated HilD activity. We examine the impact of each strategy on the transcriptional activation of SPI-1 operons, as well as secretion titer. With these manipulations, we were able to achieve titers higher than or similar to the previously reported method of *hilA* overexpression.

## *Synthetic overexpression of hilD conferred an increase in secreted protein titer*

Controlling regulation using the transcriptional regulator, HilA, led to increased heterologous protein secretion titer, and we reasoned that other transcriptional factors might be used to the same or greater effect[43]. First, we examined the impact on heterologous secretion on synthetically overexpressing HilD. To do so, the sequence encoding *hilD* as well as the region 20bp upstream of the coding sequence was placed under a placUV5 promoter on a p15a-based plasmid. The plasmid structure is identical to the *hilA* construct from Metcalf et al 2014, which was used to achieve the highest-reported secretion titer[43]. This plasmid was co-transformed with the export plasmid. Briefly the export plasmid contained the *sicA* promoter, pSicA, controlling the operon consisting of the SicP chaperone and the fusion gene encoding the SptP secretion tag, a protein of interest (POI), and a C-terminal 2XFLAG tag (for detection by western blotting). The POI secreted here was DH, a model protein compatible with secretion via the T3SS and previously showed to achieve high titers. The presence of the flag-tagged DH in the culture supernatant is dependent on a functional T3SS, so for a negative control we ran the same experiment in a Δ*invA* strain, which does not have a functional T3SS. Finally, contributions to the supernatant protein concentration due to lysis were determined by western blotting against a cytoplasmic protein, GroEL, in the supernatant.

To determine the optimal *hilD* expression level, we added the inducer IPTG at several concentrations: 0, 1, 5, 10, 50, 100, 500 and 1000µM. There was a steady increase in secretion titer with increasing concentrations of inducer used (Figure 4.2a). As anticipated, the overexpression of *hilD* did not result in detectable DH secretion in a Δ*invA* strain control. In addition, there was also no detectable increase in lysis as assessed by western blotting against a cytoplasmic protein, GroEL in the supernatant. Differences in GroEL would indicate differential lysis but this was not the case here.

## *Synthetic overexpression of hilD increased HilD levels*

We hypothesized that the increase in secretion observed was due to increasing intracellular concentrations of HilD protein. To test this hypothesis, both the *hilD* on the plasmid and the chromosome were tagged with a FLAG epitope to enable detection by western blotting. Cells harboring only these constructs, but this time without the export plasmid, were grown and induced under the same levels of IPTG as for the prior experiment.

The contribution to HilD levels for such cells comes from both the synthetic overexpression and the native copy of *hilD*. We found that HilD levels indeed increased with increasing IPTG added (Figure 4.2b). We reasoned that the shape of the curve in Figure 4.2 b was due to the

unique regulation of *hilD*, as *hilD* is autoregulated. The increased levels of HilD will drive the expression of *hilD* under its native promoter and thus changes in HilD level are not truly linear and not solely a function of the placUV5 promoter. Combining the DH secretion titers and HilD expression results, we concluded that despite higher intracellular HilD levels, there were no significant increases in secretion titer at IPTG concentrations greater than 50μM. In addition, more IPTG also resulted in lower $OD_{600}$ (Table 4.1). This confirms observations from the literature, which suggest that once a threshold of HilD level is reached, SPI-1 is fully activated and the cells are committed to the SPI-1 program. Increasing IPTG levels beyond this threshold resulted in decreasing final $OD_{600}$.
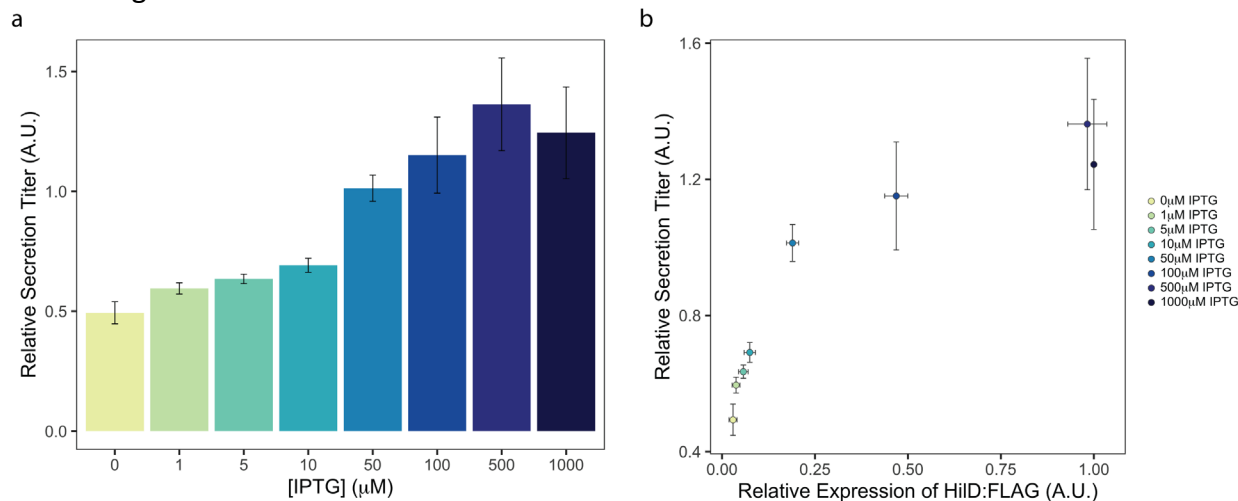


Figure 4.2 Increasing the IPTG concentration led to increasing secretion titer as measured by western blotting (a) normalized by $OD_{600}$. The increase in secretion titer due to increases in HilD levels as measured by western blotting (b). Marginal increases in secretion titer can be observed once a critical level of HilD is reached.

| STRAIN | IPTG (μM) | MEAN $OD_{600}$ | S.E. |
|---|---|---|---|
| pLac *hilD* | 0 | 3.58 | 0.32 |
| pLac *hilD* | 1 | 3.42 | 0.14 |
| pLac *hilD* | 5 | 3.44 | 0.17 |
| pLac *hilD* | 10 | 3.43 | 0.12 |
| pLac *hilD* | 50 | 3.28 | 0.12 |
| pLac *hilD* | 100 | 3.15 | 0.16 |
| pLac *hilD* | 500 | 2.72 | 0.11 |
| pLac *hilD* | 1000 | 2.82 | 0.11 |

Table 4.1 Final $OD_{600}$ of strains harboring pLac *hilD* induced with different level of IPTG after 8 hours of induction.

## HilD levels drives different levels of activation of pHilA promoter

One of the promoters directly controlled by HilD is the promoter for the *hilA* operon. We reasoned that the *hilA* promoter activity would be a more accurate readout for the activation of the system than direct measurement of HilD levels by western blotting, because HilD can be prevented from binding to promoters via sequestration by other proteins such as HilE and FliZ[154,155]. We transformed cells harboring the pLac *hilD* plasmid with a second plasmid containing the promoter of *hilA* fused to GFPmut2. We then used flow cytometry to assess the fluorescence of each cell, which represents the activation of the *hilA* promoter, pHilA, as a function of

increasing IPTG concentration. The fluorescence level of each cell should directly correlate to the extent to which HilD is binding to its cognate promoters such as pHilA.

As we increased the concentration of IPTG, we can observe an earlier activation of the pHilA promoter (Figure 4.3a). This reflects the trend seen in the secretion titer (Figure 4.2) when inducing *hilD*. We describe the population with high levels of GFP fluorescence as the "on" population. A small population of cells (~10%) with very low GFP fluorescence could be observed at high IPTG levels of 500 and 1000 μM after 1 hour, and we hypothesized that these are dead cells. Taken together with the lower $OD_{600}$ (Table 4.1), these results indicate that there is a metabolic cost to inducing the system with >100μM of IPTG.

The GFP geometric mean of the "on" population under pHilA also increased with increasing levels of IPTG up to 50 μM of IPTG (Figure 4.3b). However, the geometric mean over time for 100 μM IPTG lay above that of the 50 μM IPTG. Taken with the slightly earlier activation of pHilA at 100 μM IPTG, and the lack of further increase in secretion titer at concentration >50 μM of IPTG, we chose 100 μM of IPTG as the optimal level of induction concentration for balancing growth and secretion. This is the concentration of IPTG that was used for the rest of the studies.
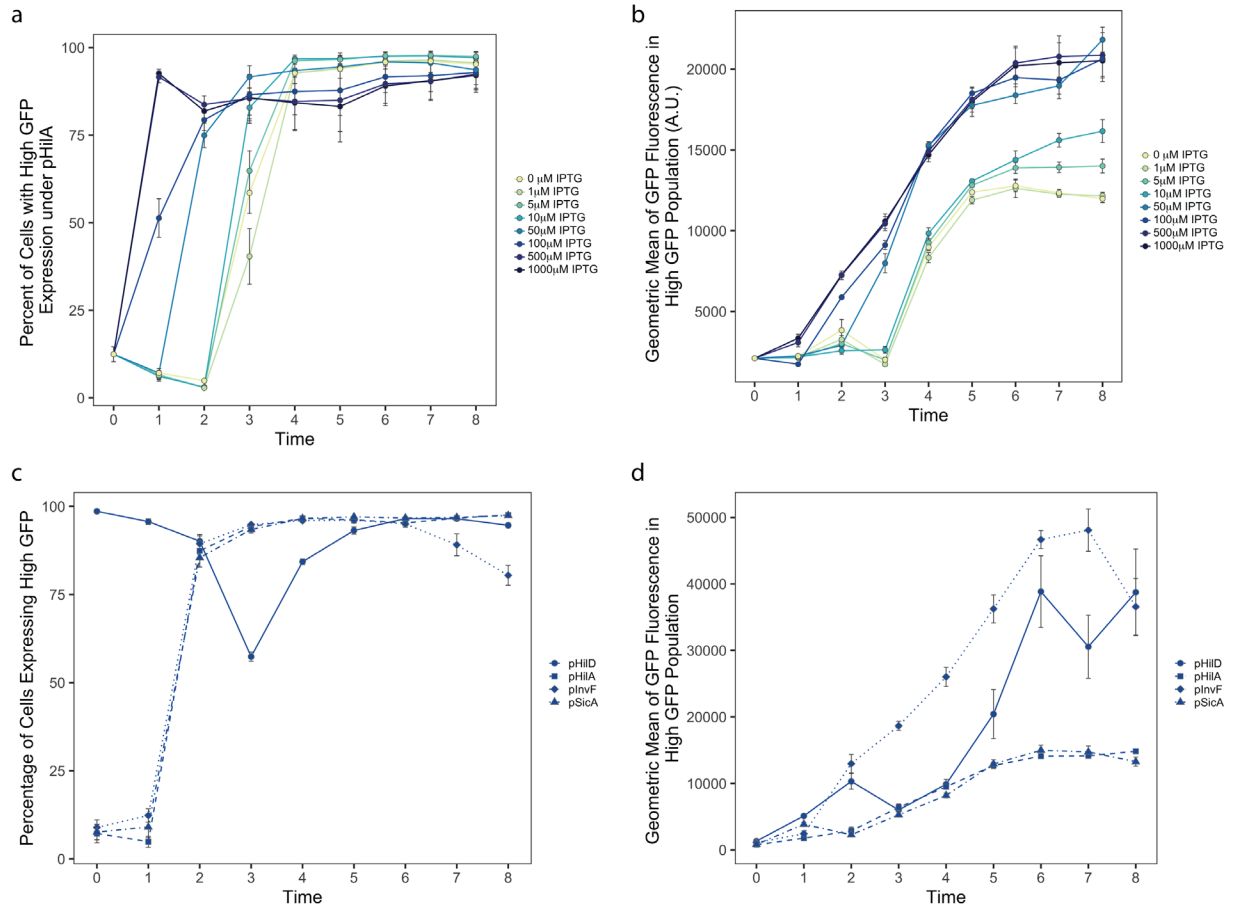


**Figure 4.3 Plots depicting the percentage of "on" population expressing high GFP under pHilA (a) and the GFP geometric mean of the on population (b) as a function of time. The percentage of "on" population expressing high GFP under all the SPI-1 promoters tested are shown in (a) and the GFP geometric mean of the "on" population in (b). Representative plots can be found in Figure 4.4 and Figure 4.5. Most of the SPI-1 promoters were highly active after 1 hour when *hilD* was overexpressed.**
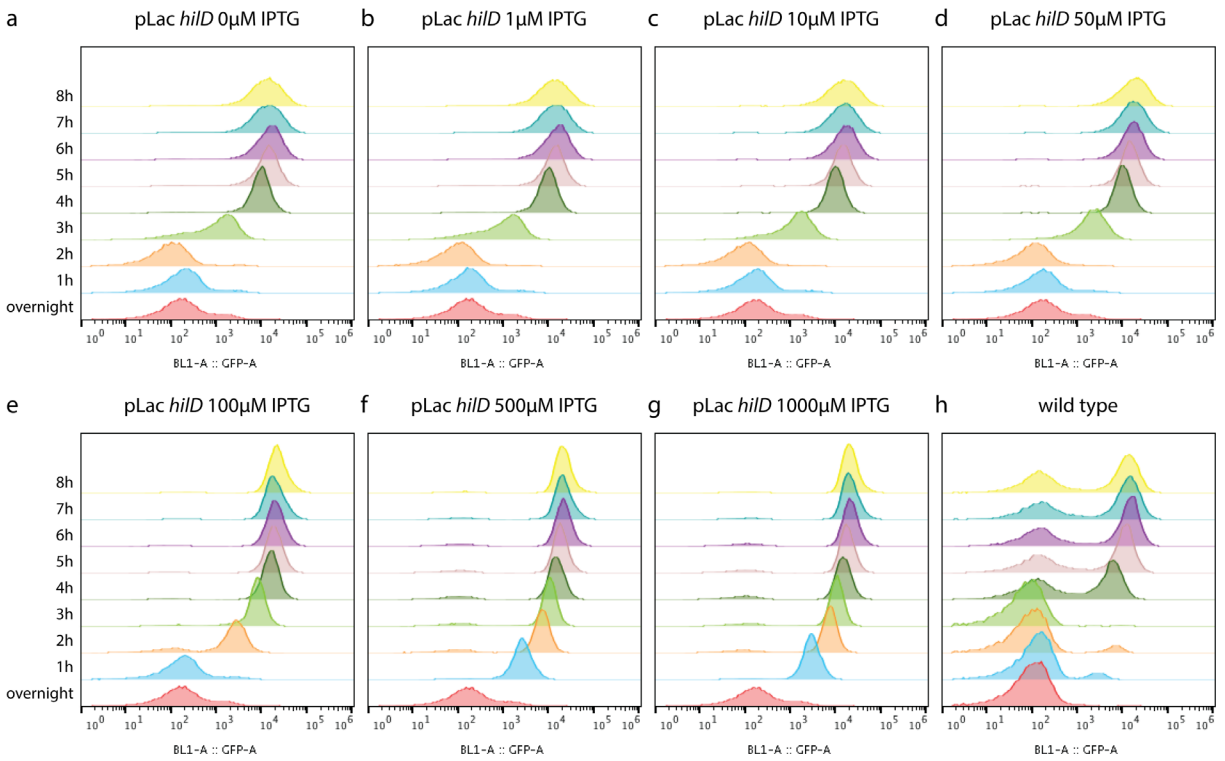
**Figure 4.4 Representative flow plots for the different levels of *hilD* induction. The wild type behavior for each promoter is also included for comparison.**
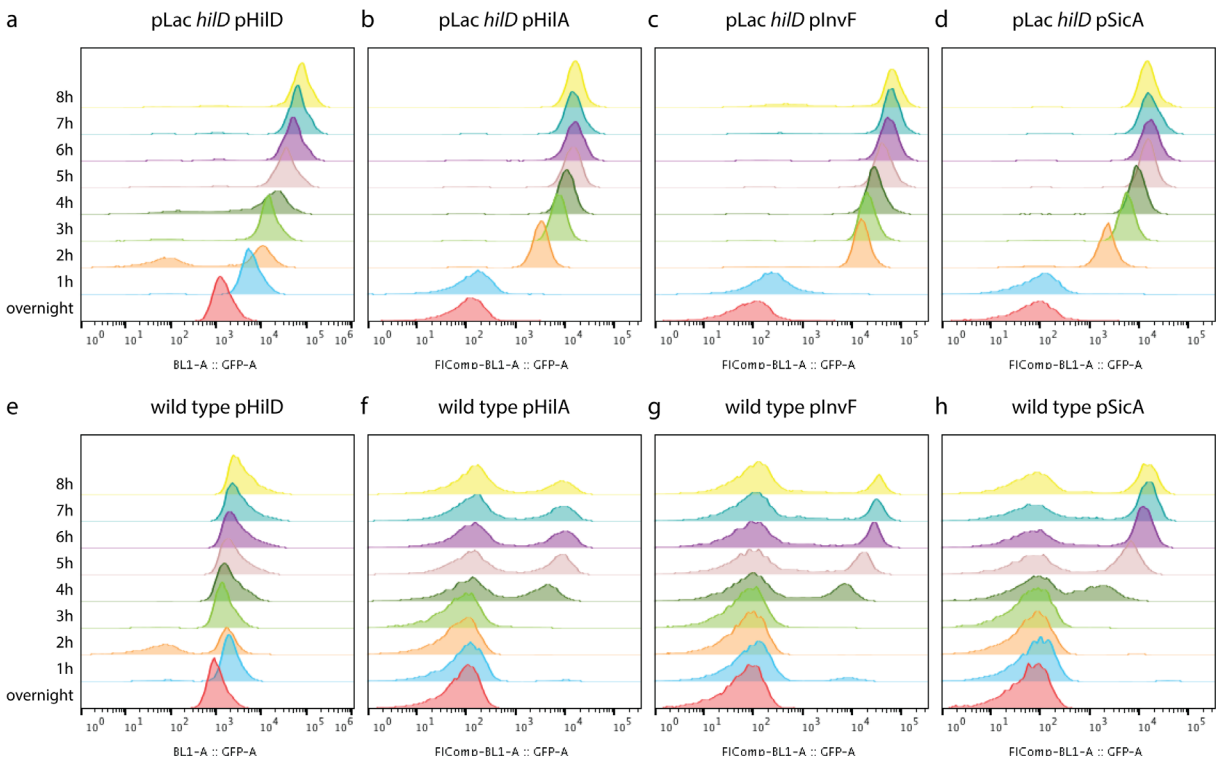


**Figure 4.5 Representative flow plots for different promoters when *hilD* is overexpressed. The wild type behavior for each promoter is also included for comparison.**

41

## Overexpression of hilD results in earlier activation of SPI-1 promoters

We further wanted to examine the activation of other key promoters in the system – pHilD, pInvF and pSicA – when the plasmid-borne *hilD* was induced with 100 μM IPTG. As with the pHilA experiment, we used GFPmut2 as the reporter protein and cloned it downstream of each of the three promoters, and then transformed cells harboring pLac *hilD* with each of these additional plasmids. Flow cytometry was again used to monitor fluorescence as a function of IPTG at various time points. Interestingly, for the two downstream promoters, pInvF and pSicA, more than 90% of the population expressed high levels of GFP after 2 hours (Figure 4.3c) compared to previously reported data of 3 hours when overexpressing *hilA*[43]. Maximum GFP expression under the different promoters was achieved at 6 hours (Figure 4.3d) which was similar to previous reports when overexpressing *hilA*. The earlier activation of the SPI-1 promoters could result in the higher secretion titer observed in Figure 4.2b when overexpressing *hilD* as compared to overexpressing *hilA*.

The decrease in percentage of the "on" population for pHilD (Figure 4.3c) from 1 to 3 hours was similar to the behavior observed without overexpression of the system (comparing Figure 4.10 a with Figure 4.10 e). This could be due to other cellular factors that can modulate the activity of the *hilD* promoter.

## Transcriptional fusions to hilD on the genome lead to an increase in secretion titer

We reasoned that the use of the plasmid-based system for pHilD was not an accurate representation of the native *hilD* transcription and constructed a transcriptional fusion of GFPmut2 to *hilD* on the genome to better capture the native behavior. Before assessing fluorescence, we transformed this strain with our export plasmid to confirm that secretion titer is not affected by our strain engineering.

Unexpectedly, the addition of the GFPmut2 coding sequence after *hilD* resulted in an increase in secretion (Figure 4.7) While this means that the genomic transcriptional fusion of GFPmut2 was also not a reliable measure of the *hilD* transcription, it is a fortuitous finding given that our overall goal is to increase heterologous secretion titer. We hypothesized that changing the DNA sequence found at the 3' end of the *hilD* transcript could be used to modulate the SPI-1 T3SS. We tested this hypothesis by appending DNA sequences of different compositions and lengths after the *hilD* sequence. We chose mCherry, halotag, maltose binding protein (MBP) and the *lacZα* to transcriptionally fuse to *hilD* on the genome.

The addition of a long DNA sequence (i.e. >700 bp) encoding soluble proteins – GFPmut2 (711 bp), mCherry (711 bp), halotag (894 bp), and MBP (1164 bp) significantly increased secretion and was comparable to the titers achieved by overexpressing *hilD* with 100 μM IPTG (Figure 4.6a). The addition of the *lacZα* sequence, which is much shorter at 270 bp in length, increased secretion titer to a much lower extent (Figure 4.6a). The result here indicates that there may be an optimal length of DNA to insert downstream of *hilD*.
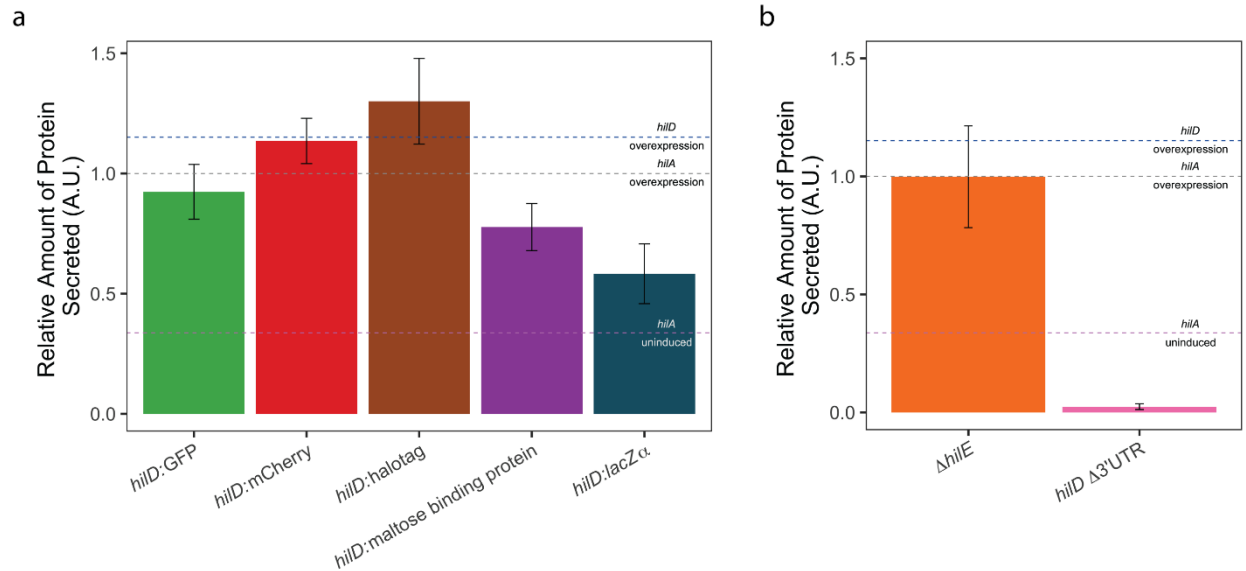
**Figure 4.6 (a) Transcriptionally fusing different protein-encoding sequences led to increases in secretion titer as measured by western blotting.. (b) Deleting _hilE_ conferred an increase in secretion titer but not when the _hilD_ 3'UTR was deleted. For both plots, results are normalized to the secretion titer obtained by overexpressing _hilA_ off the pLac _hilA_ vector with 100 μM IPTG (gray dashed line). For comparison, a blue dashed line is used to depict the relative secretion titer obtained by overexpressing _hilD_ from a pLac _hilD_ vector with 100 μM IPTG is shown. A purple dashed line is also used to depict the relative secretion titer obtained from DW01 harboring the pLac _hilA_ plasmid that had not been induced by IPTG.**

## _Removing negative regulators of HilD can also activate the system_

Inspired by these results and the low levels of HilD needed to activate the system (Figure 4.2 b), we decided to remove existing negative regulators of _hilD_ translation and activity. Our expectation was that this would also increase heterologous secretion to the levels observed with induced, plasmid-borne _hilD_.

Two important negative regulators of _hilD_ are known – _hilE_ and the 3'UTR of _hilD_. HilE binds to HilD, preventing the HilD dimer from binding to DNA[69,154]. Thus, by knocking out _hilE_, we expected increased binding of HilD to its cognate promoters and an overall increase in SPI-1 activation and heterologous secretion.

In contrast, Hfq can bind to the 3'UTR of _hilD_, preventing its translation[153]. We reasoned that by deleting the 3'UTR responsible for destabilizing the _hilD_ transcript, we would be able to observe a similar increase in heterologous secretion that was observed with the transcriptional fusions in Figure 4.6a. Deleting the 3'UTR could have an synergistic effect when combined with transcriptional fusions to the variety of proteins.

To test these hypotheses, we created two strains: one in which _hilE_ is deleted and another in which the 3'UTR of _hilD_ is deleted. We then transformed these strains with our export plasmid and carried out the secretion assay. Knocking out _hilE_ resulted in secretion titer similar to the synthetic induction of _hilD_ (Figure 4.6b). Surprisingly, deleting the _hilD_ 3'UTR resulted in no increase in secretion titer. As _hilD_ is upstream of the operon containing _hilA_, we suspect that this deletion might be having a polar effect on HilA levels; alternatively, there may be regulation within this section of the transcript that is important for SPI-1 regulation.

## *Genomic alterations increased HilD levels*

To explore the role of HilD level on secretion titer, we made several of the genomic manipulations described above in the strain harboring *hilD*:1XFLAG on the genome. Four strains were made – Δ*hilE hilD*:1XFLAG, *hilD*:1XFLAG:GFP, *hilD*:1XFLAG:mCherry, and *hilD*:1XFLAG:*lacZα*. HilD levels were determined by western bloting against the FLAG tag. Higher HilD levels were observed in all the engineered strains , and the levels of HilD in Δ*hilE hilD*:1XFLAG were approximately five times greater than in the *hilD* transcriptional fusions (Figure 4.7). Despite the higher HilD level, knocking out *hilE* had comparable secretion titers to the other strains (Figure 4.6). We hypothesized that the increased HilD levels observed in *hilD*:1XFLAG:mCherry, and *hilD*:1XFLAG:*lacZα* could be due to the stabilization of the *hilD* mRNA from the addition of DNA sequences to the 3' end of the transcript.
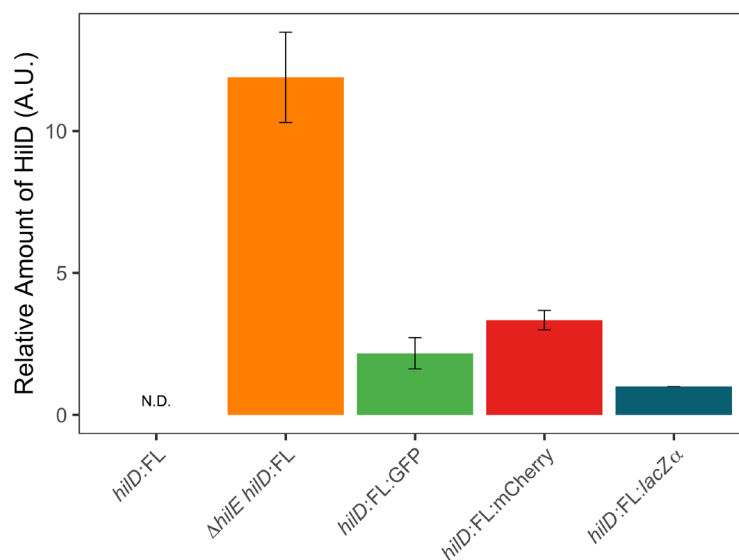


**Figure 4.7 The different engineered strains had increased HilD levels above the *hilD*:1XFLAG strain. Although knocking out *hilE* resulted in more HilD, this did not translate to higher secretion titer.**

## *Genomic engineering approaches turn on different SPI-1 promoters differently*

We wanted to determine how the different strain engineering approaches affect the behavior of the different SPI-1 promoters – pHilD, pInvF and pSicA. As GFPmut2 was used as the reporter protein, only *hilD*:mCherry, *hilD*:halotag, *hilD*:MBP and Δ*hilE* were transformed with the respective promoter fusion plasmids and assayed.

There was a higher background activation of pHilA, pInvF and pSicA in the strains with long coding sequence transcriptional fusions to *hilD*, as compared to the background activation in the wild type strain shown in Figure 4.5 (Figure 4.8). This is surprising as our protocol includes 0.4% glucose in the overnight culture prior to the experiment to repress such residual activity. This could be due to the stabilization of the *hilD* transcript. The overall temporal activations of the different promoters were largely similar in pattern in all the strains (Figure 4.8 a – c). There was a smaller percentage of the "on" population turning on for all the promoters in the strains harboring *hilD:lacZα*, which corresponds to the lower secretion titer observed in Figure 4.6 (Figure 4.8 d – f).
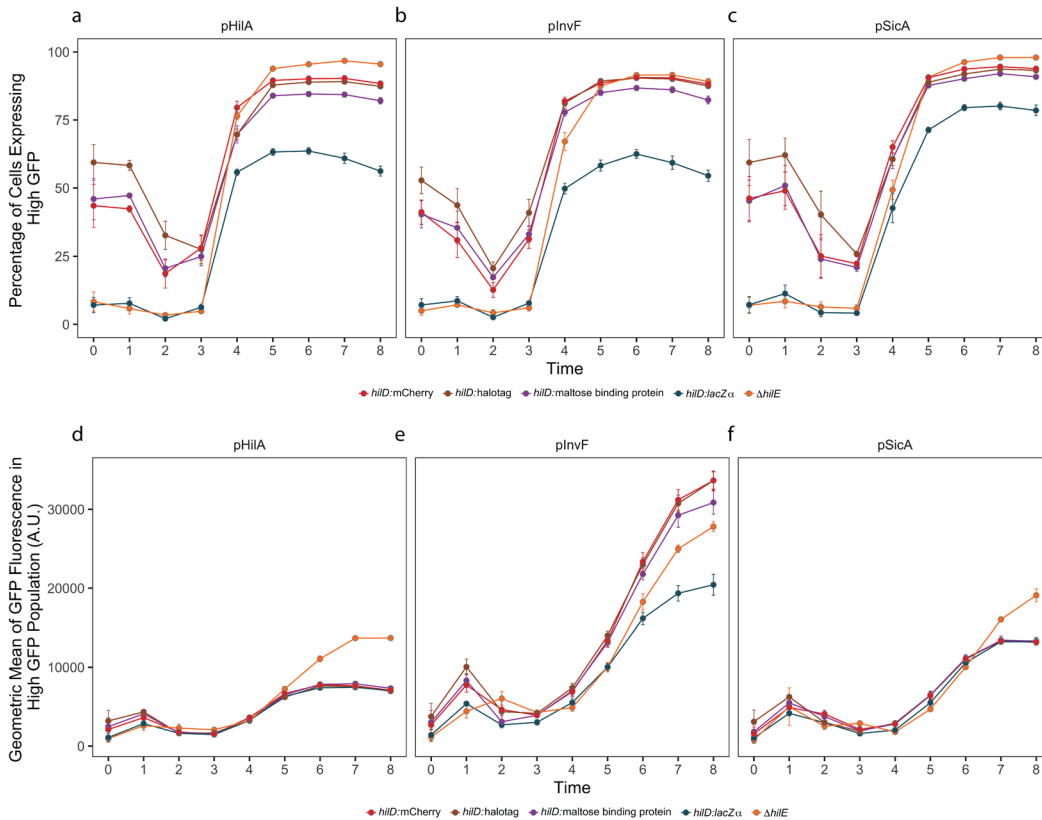
**Figure 4.8 The percentage of the "on" population expressing high GFP under (a) pHilA, (b) pInvF, and (c) pSicA for the different genomic mutations. The respective GFP geometric mean of each population is shown in (d), (e) and (f) for each of the promoters. Representative plots can be found in Figure 4.9 and Figure 4.10. The higher percentage of "on" population and higher geometric mean corresponded to increased activation of SPI-1 and thus higher observed titer.**

Knocking out *hilE* resulted in a higher pHilA GFP geometric mean compared to the fluorescence of the strains harboring the *hilD* transcriptional fusions. The GFP geometric mean conferred by the pHilA GFP reporter for Δ*hilE* as seen in Figure 4.8d was similar to GFP geometric mean of DW01/pLac-*hilD* at 50μM IPTG (Figure 4.3b), suggesting comparable HilD activity in these two conditions. Despite the higher pHilA activity, this did not result in higher pInvF GFP geometric mean; instead, it is slightly lower than that of the long transcriptional fusions (Figure 4.8e). Again, knocking out *hilE* unexpectedly resulted in higher geometric mean conferred by the pSicA GFP reporter compared to the long *hilD* transcriptional fusions (Figure 4.8f). Despite the differences observed in the extent of activation in the different SPI-1 promoters, the secretion titers from strain harboring Δ*hilE* and the long transcriptional fusions were similar (Figure 4.6).
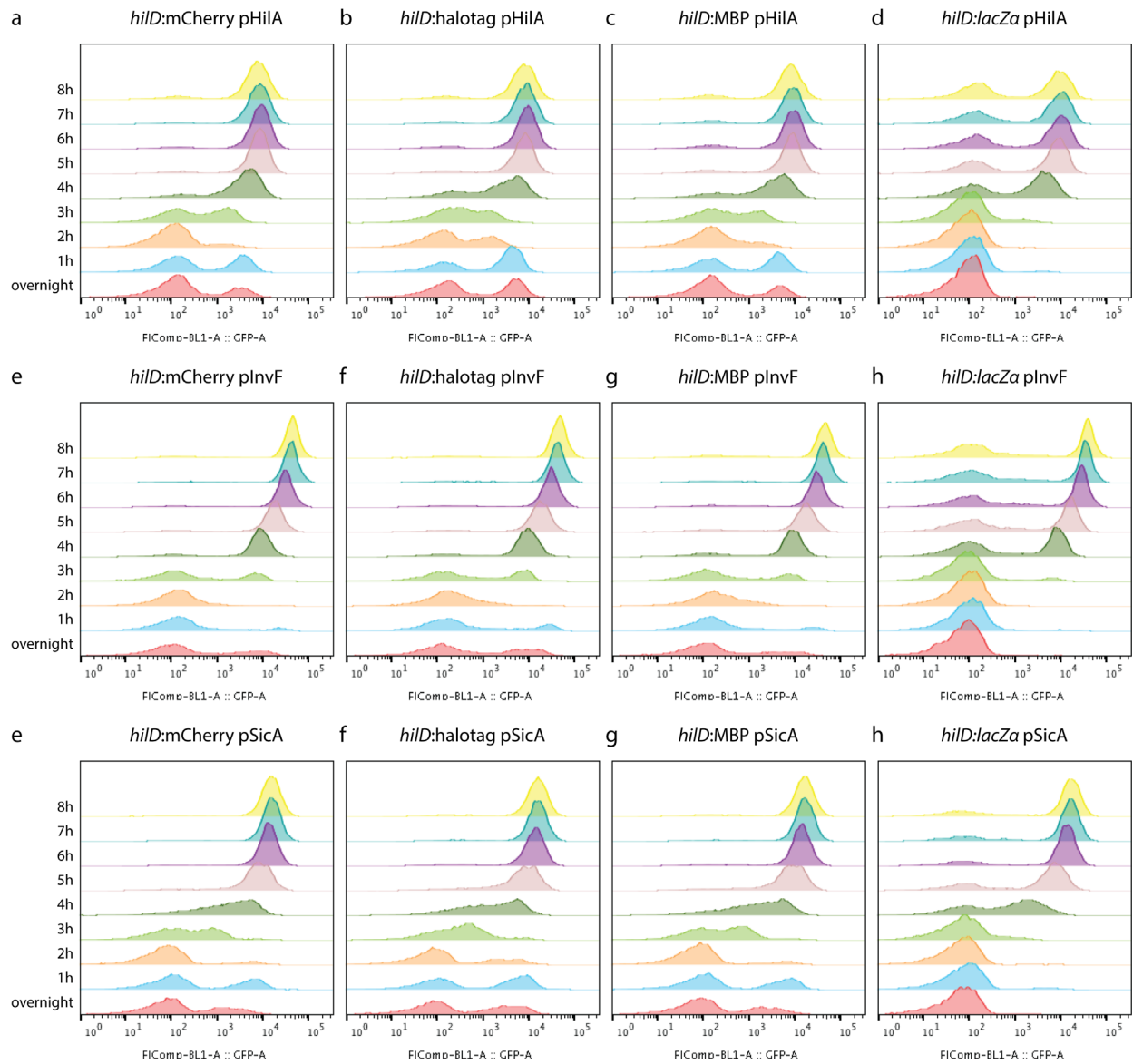
**Figure 4.9 Representative flow plots of different promoters for each *hilD* transcriptional fusion.**
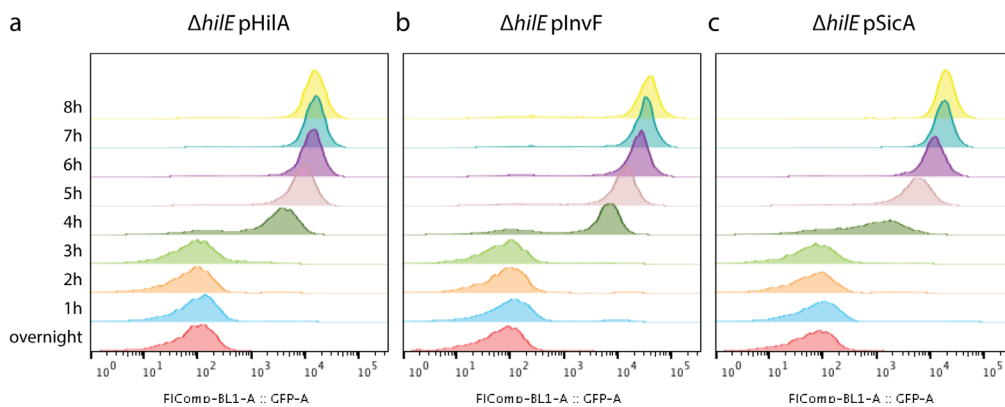


**Figure 4.10 Representative flow plots of different promoters for Δ*hilE*.**

## Catabolite repression of the different strains

Glucose inhibits SPI-1 expression in a process mediated by cAMP receptor (CRP) protein[156]. This repressive mark is seen even when *hilA* is overexpressed although the main effect of catabolic repression is on the translational efficiency *hilD*. Catabolic repression of SPI-1 is mediated by the *sirA/barA* system which disrupts the translation of the *hilD* transcription through the activity of Hfq[156–158]. Hfq is a RNA chaperone that facilitates translational regulation by binding to sRNAs and mRNAs. We added increasing levels of glucose – 0.25%, 0.5%, 0.75% and 1% - to the genomic strains harboring the pSicA GFPmut2 plasmid. For comparison, we also added these levels of glucose to the strain harboring pSicA GFP mut2 and pLac *hilA*.

When *hilA* was overexpressed, the addition of glucose reduced GFP fluorescence levels under the pSicA promoter to 20% - 30% of the GFP fluorescence levels without the addition of glucose (Figure 4.11). Glucose is able to completely shut down pSicA expression in the Δ*hilE* strain. This is not surprising as HilE prevents HilD from binding to different promoters and does not affect the translational efficiency of the *hilD* transcripts. In contrast, the different strains containing the different transcriptional fusions to *hilD* appeared to overcome some of the catabolic repression, suggesting that the *hilD* transcript was likely to be more available for translation. This provides further evidence that the transcriptional fusions of the coding sequence to *hilD* were likely stabilizing the *hilD* transcript.
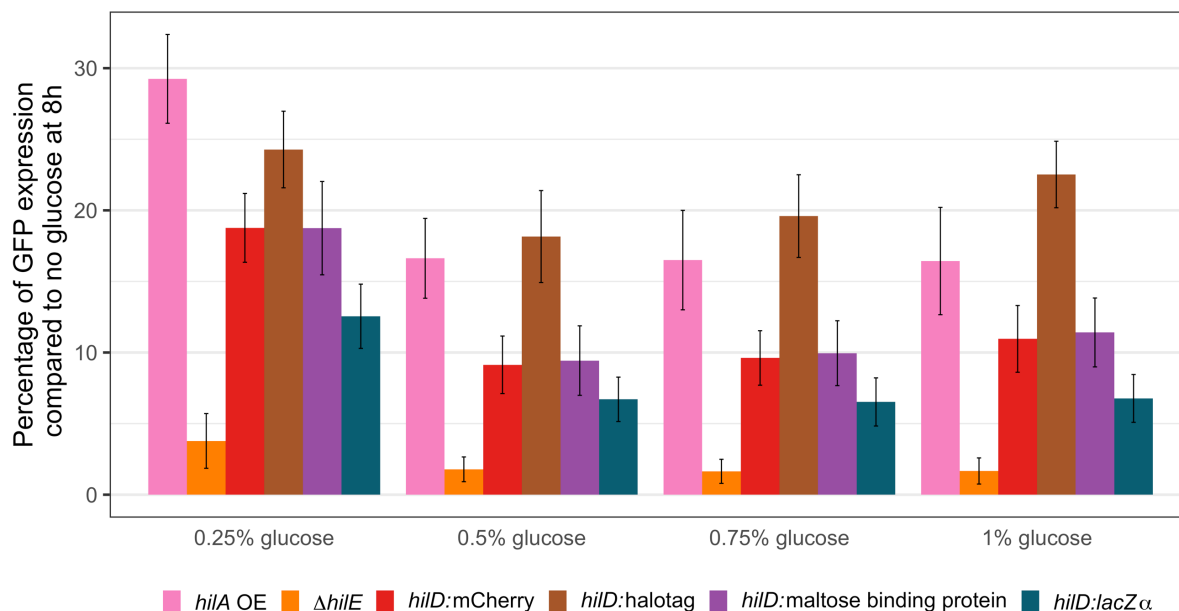


**Figure 4.11 The GFP fluorescence under pSicA was measured by flow cytometry at 8 hours. In each strain, the GFP fluorescence level under different glucose levels was divided by the same strain without glucose. The GFP expression in pSicA was reduced even with *hilA* was overexpressed. Each of the engineered strains are repressed to a different extent by glucose with GFP expression in Δ*hilE* almost completely turned off.**

## Discussion

Activation of SPI-1 occurred at a much earlier time when *hilD* was overexpressed from a plasmid compared to when *hilA* was overexpressed[43]. This resulted in higher secretion titers observed when synthetically overexpressing *hilD*. This earlier activation of SPI-1 promoters was

unexpected as *hilD* is upstream of *hilA* and thus more time should be needed for the overall T3SS system to become induced. The earlier activation could be due to other genes outside SPI-1 that HilD is known to regulate. Alternatively, the pHilA promoter could be much stronger than pLac and result in higher *hilA* transcript levels despite there being fewer copies of the *hilA* gene in DW01/pLac-*hilD* compared to DW01/pLac-*hilA*. The proximity of *hilD* and *hilA* on the genome might also play a part in ensuring more efficient binding of HilD to pHilA as less time may be spent by HilD searching for its cognate promoter sequence.

Previous studies suggested that once HilD levels exceeded a threshold, the whole system would be activated. Instead of the clean switch-like (or digital) behavior that we expected, the activation behaved more like a dimmer switch (or analog system) with maximum activation and secretion occurring at around 50-100 μM even when more HilD proteins were present. The minimal protein level needed might have already been exceeded due to the extra copies of *hilD* under a leaky promoter such as pLac. As the prior work focused on the infectivity of *S. enterica*, slight differences in the SPI-1 activation might not result in detectable difference in infectivity but here we could detect differences in heterologous secretion.

The unexpected activation of SPI-1 when different coding sequences were fused to *hilD* is likely due to increased mRNA stability of the transcript. Adding a foreign DNA sequence to the end of the mRNA could potentially disrupt the native regulation of the *hilD* transcript due to the 3'UTR. Moreover, these coding sequences are used for protein expression and thus likely to form stable transcripts.

The slightly different effects of the different genomic alterations made in knocking out *hilE* and transcriptionally fusing reporter proteins to *hilD* on SPI-1 promoters suggests that combining them may have an additive effect. This experiment was performed by another graduate student and indeed combining the two strategies to engineer the strain improved secretion titers further.

As using IPTG on an industrial scale is costly, so we wanted to develop strains that can be used without IPTG induction. In addition, using a plasmid-based system to overexpress *hilD* would not be optimal. The final strains (Δ*hilE, hilD:mCherry* and Δ*hilE hilD:mCherry*) each showed promise as a general strain for bacterial protein secretion of different proteins. Although the reported titers here are lower than the maximum reported titer using synthetic overexpression of *hilD*, these titers are comparable to those observed with overexpressing *hilA* from a chemically induced promoter. In addition, the systems introduced here require one less plasmid, opening up the potential for co-expressing other helper factors that can benefit protein production and secretion. Lastly, problems such as plasmid instability, metabolic burden of plasmid maintenance and potential recombination of the plasmid with the native copy on the genome can be avoided with the use of these engineered strains.

Given that there are other layers of control on *hilD*, it might be possible to increase titer further even more by either further strain engineering or production optimization with process engineering. Specific options include the deletion of FnrS, a sRNA known to inhibit *hilD* translation [134]; and FliZ, a protein known to bind to HilD and prevents HilD from carrying out its function[155]. Further improvements can also be achieved by tuning environmental factors such as $O_2$ levels and ferric levels that are known to affect HilD levels[73,134]. Combining strain engineering, media optimization and growth optimization will lead to industrially relevant production titers.

## Materials and methods

### Cloning of various constructs

The pLac plasmids were cloned using Gibson assembly[159]. *hilD* insert was prepared using PCR with overlaps flanking the 5' and 3' ends. Insert were then thermocycled with pLac p15a vector linearized by Phusion PCR using standard procedures as described previously. New primers were used to linearize the pLac *hilD* to add the FLAG tag and then assembled by Gibson assembly. The primers are found in Table 4.2. 5 μl of the Gibson assembly reaction was then used to transform chemically competent *E. coli* DH10B cells by heat shock. The chemically competent cells were subjected to 20 minutes on ice, followed by 60 seconds at 42°C, 2 minutes on ice and then recovery for 1 hour at 37°C with 350 μl of SOC media. 50 μl of the transformed cells were plated on LB agar plates with kanamycin. All genes were sequenced-verified by Quintara Inc. (Boston, MA).

### Strains, media, growth and harvest of bacteria

*Salmonella enterica* strains were grown overnight in LB Lennox (Dot Scientific Catalog no. DSL24066-500) with the appropriate antibiotics as per Metcalf 2014 at 37°C and 225rpm in 24-well blocks. The overnights were sub-cultured to an $OD_{600}$ of 0.05 in 24-well blocks for all experiments (Axygen Catalog no. PDW10ML24C). Cells were electrotransformed with the required plasmids.

For secretion assay, cultures were grown for 8 hours at 37°C and 225 rpm. The overexpression of *hilA* was induced by 0.1 mM IPTG (Dot Scientific Catalog no. DSI56000-5) at the point of subculture. Antibiotics were added as needed. At the point of harvest, 20 μL of the culture was added to 40 μL of 4X Laemilli buffer for the whole culture lysate (WCL) samples. The blocks were then spun at 4000 g for 10 minutes to collect the supernatant (Sup) fractions. 40 μL of the supernatant was added to 16 μL of 4X Laemilli buffer.

### Generation of strains used

Genomic modifications made in this study were done using the Court lab recombineering method[145,146]. Briefly, strains were transformed with pSIM6 and then grown overnight at 30°C, 225 rpm. The cells were then sub-cultured at a 1:100 dilution and grown for ~2hrs to an OD of 0.4 – 0.8. The lambda red system was then induced at 42°C for 15 minutes and then the cultures were cooled in an ice water bath for 10 minutes. The cells were spun down at 4600 g for 3 minutes at 4°C, washed thrice in $ddH_2O$ and finally resuspended in 200 μL of $ddH_2O$. 200ng of DNA products or 5 μL of 10mM primers were electroporated into 50 μL of cells as needed. The DNA products used in round one of recombineering were generated by Phusion PCR using the TUC01 genomic DNA as a template using the primers in Table 4.2. The PCR reaction was then cleaned up with the Wizard SV Gel and PCR Clean-up kit (Promega Catalog no. A9282).

For round one of recombineering, the cells were recovered in 350μL of SOC at 30°C for an hour and then plated on LB agar plates with 10 μg/L of chloramphenicol. For round two, the cells were recovered in 10 mL of LB Lennox at 37°C for 4 hours. Serial dilutions of the recovery were then plated on 6% sucrose plates. Patch plating on LB agar plates with 10 μg/mL of chloramphenicol or 30 μg/mL carbenicillin were used to determine successful recombination and

loss of pSIM6 plasmid respectively. Colony PCR was then carried out using GoTaq (Promega Catalog no. M3008) to isolate the gene of interest to be Sanger sequence verified by Quintara BioSciences. The verified clones were then electroporated with the relevant plasmids as needed to generate the strains.

The PCR products for inserting the transcriptional fusions into wild type were made with the primers in Table 4.2 with different plasmids using Phusion. To insert the transcriptional fusion into the *hilD:1XFLAG* background, the primers hilD FL rec F and hilD 3'UTR seq R were used to PCR out each fusion with the previously engineered strains using Phusion.

## Western blotting

Samples were run on a 12.5% SDS PAGE gel at 150V for 60 minutes. 2 μL and 4 μL of samples of WCL and supernatant were loaded respectively. The gels were then equilibrated in Towbin buffer with 20% methanol for 15 minutes. The samples were then transferred to a PVDF membrane (Millipore Catalog no. IPVH00010) using the Owl HEP-1 blotter at 0.3A for 40 minutes. After which, the membrane was blocked in 5% milk in TBST for 1 hour. The membrane was then decorated with anti-FLAG M2 antibodies from mice (Sigma Catalog no. F3165-1MG) in 1% milk (1:6666) overnight, followed by three 5-minutes washes in TBST (0.1% Tween-20). The membrane was then decorated with anti-mouse secondary antibody from goat conjugated with HRP (Thermo Fisher Catalog no. 32430) in TBST (1:1000) for 1 hour, followed by three 5-minutes washes in TBST. A different secondary antibody (Jackson ImmunoResearch Catalog no. 515-035-003) was used for the HilD:FLAG blots. Blots were then imaged using SuperSignal West Pico (Thermo Scientific Catalog no. 34080) on the Bio-Rad Chemidoc. The secretion titer was then quantified by densitometry using the Image Lab software v5.2.1 provided by Bio-Rad. The values were then normalized by $OD_{600}$ and to pLac *hilA* induced by 100 μM IPTG.

The blots were then striped using the mild stripping protocol modified from Abcam to reprobe for GroEL. The membranes were incubated in mild stripping buffer for 8 minutes twice, washed twice in ddH$_2$O for 10 minutes and then twice in TBST for 5 minutes. The stripped blot was blocked in 5% milk in TBST as above. The membrane was processed as above but using anti-GroEL antibodies from rabbit (Sigma Catalog no. G6532-.5ML) in 1% milk (1:10000) and anti-rabbit secondary from goat conjugated with HRP (Thermo Fisher catalog no. 32460) instead.

## Flow cytometry

The cells were grown overnight in LB Lennox (Dot Scientific Catalog no. DSL24066-500) with the appropriate antibiotics supplemented with 0.4% glucose. The overnight culture was diluted to an $OD_{600}$ of 0.05 for subculturing and induced according as above. Samples were taken every hour and diluted to an OD of approximately 0.03 in PBS with 2 mg/mL of kanamycin sulfate. The samples were kept overnight at 4°C and protected from light before running it on the Attune NxT Flow Cytometer (BY) using the autosampler. Data was collected for at least 20 thousand cells and processed with FlowJo v10.5.3. The resulting data was visualized with R and the ggplot2 package[137].

In order to have a cleaner background to measure the amount of catabolic repression on each strain, there was an additional outgrowth of 3 hours after the first subculture. 150 μL of the first subculture was added to 5 mL of LB-L supplemented with the respective amount of glucose.

*Primers used in Chapter 4*

| Cloning of hilD constructs | |
|---|---|
| pLac F GA | TAAAAATGGCGAACCATTAAGGATCCAAACTCGAGTAAGG |
| pLac R GA | ATTATCCCTTTGTTGATGTTAGATCTTTTGAATTCTGAAATTGTT |
| hilD F GA | TTTCAGAATTCAAAAGATCTAACATCAACAAAGGGATAATATGG |
| hilD R GA | CCTTACTCGAGTTTGGATCCTTAATGGTTCGCCATTTTTATGA |
| pLac hilD:FL F GA | GATTATAAAGATGACGATGACAAGTAAGGATCCAAACTCGAGTAAG |
| pLac hilD:FL R GA | CTTGTCATCGTCATCTTTATAATCATGGTTCGCCATTTTTATGAA |
| Recombineering Round 1 Primers | |
| hilD:catGsacB F | AACTACGCCATCGACATTCATAAAAATGGCGAACCATTAATGTGACGGAAGATCACTTCG |
| hilD:FLAG catsacB R | AATGGCGAACCATGATTATAAAGATGACGATGACAAGTAATGTGACGGAAGATCACTTCG |
| hilE catsacB F | ACGAAATGGCTGGAAAATGGAACGTTCTTTCATTGTTGGCTGTGACGGAAGATCACTTCG |
| hilE catsacB R | GTCCTCATCGCCACAGCGCCTGTCGGTGAAGAGGCCGCCATCAAAGGGAAAACTGTCCAT |
| Recombineering Round 2 Primers | |
| hilD 3'UTR KO R | GCCGGCCTTAATCCACAGGGTTAAAGCCGGTTAATGGTTCGCCATTTTTATGAATGTCGA |
| hilD:FLAG R | TTACTTAAGTGACAGATACAAAAAATGTTACTTGTCATCGTCATCTTTATAATCATGGTTCGCCATTTTTATGAATGTCGATGG |
| hilE KO | ATGGCTGGAAAATGGAACGTTCTTTCATTGTTGGCGGCGGCCTCTTCACCGACAGGCGCTGTGGCGATGA |
| hilD halotag F | AACTACGCCATCGACATTCATAAAAATGGCGAACCATTAAATTAAAGAGGAGAAAGGTCATGGGATCCGAAATCGGTACTG |
| hilD halotag R | ATAAAAATCTTTACTTAAGTGACAGATACAAAAAATGTTAACCGGAAATCTCCAGAGTAG |
| hilD MBP F | AACTACGCCATCGACATTCATAAAAATGGCGAACCATTAAATTAAAGAGGAGAAAGGTCATGAAAATCGAAGAAGGTAAACTGGTA |
| hilD MBP R | TAATAAAAATCTTTACTTAAGTGACAGATACAAAAAATGTTAGTTTTCCTCGATCCCGAG |
| hilD lacZalpha F | AACTACGCCATCGACATTCATAAAAATGGCGAACCATTAAATTAAAGAGGAGAAAGGTCATGACCATGATTACGGATTCACTG |
| hilD lacZalpha R | TTAATAAAAATCTTTACTTAAGTGACAGATACAAAAAATGTTATTCGCCATTCAGGCT |
| hilD FL rec F | AATGGCGAACCATGATTATAAAGATGACGATGACAAGTAAATTAAAGAGGAGAAAGGTCATG |
| Primers to amplify PCR product for Sanger Sequencing | |
| hilD 3'UTR seq F | ATCGGCAAGAATGAATCAGG |
| hilD 3'UTR seq R | CAAGCGTGACTGTTTCGGTA |

| | |
|---|---|
| hilE seq F | TCTATATTCCGATTCGGTGG |
| hilE seq R | TGTGTTTCATCGCTTTTCC |
| Sanger Sequencing Primers | |
| hilD 3'UTR int seq F | AGCACGTCCTACTTCATTCAA |

**Table 4.2 Primers used in Chapter 4**

## Chapter 5 Understanding Cellular Regulation in Heterologous Secretion

As described previously in Chapter 1, the regulation of SPI-1 is tightly regulated such that each SPI-1 operon activates at different times along the regulation cascade. Currently, our lab uses a set of plasmids encoding promoter fusions to GFP to study the temporal dynamics of the SPI-1 T3SS activation and these were used extensively in Chapter 4[43,60]. The use of promoter fusions suffers from several drawbacks. Firstly, only the promoter region of each operon is encoded on the plasmid which is problematic as other control elements like the untranslated regions can affect translational rates. Secondly, the plasmid is a high copy plasmid which can have between 10-100 copies per cell as opposed to one native promoter on the genome which leads to an overestimation of its activity. Thirdly, the homology to genome from the promoters can led to plasmid instability due to recombination[160]. Lastly, from the turbidostat experiments (described later in Chapter 6, detectable expression from the SPI-1 pSicA on a plasmid does not always correspond with a functional SPI-1 injectisome expressed from the genome.

One gene that would be particularly difficult to study using promoter fusions is *hilD* and the challenges for this locus were detailed in Chapter 4. As previously mentioned, *HilD* is controlled on numerous levels from transcription, mRNA stability, translation and protein activity. The use of pHilD-GFP to study the expression of *hilD* is not ideal as the stability of the *hilD* transcript depends on its 5'UTR and 3'UTR regions and this is not found on the promoter fusion. Moreover, the DNA sequence composition of GFP is very different from the genes in SPI-1. In addition to the promoters, SPI-1 expression can be modulated by global regulators that are dependent of the DNA sequence composition. For example, AT-rich regions in SPI-1 can be silenced by the histone-like nucleoid restructuring protein such as *hns*. Different microRNAs that involved in the tight regulation also depend on the transcript sequence.

Here, I address the limitations of plasmid-based promoter fusions. First, I engineered eight strains containing transcriptional fusions to different SPI-1 genes that represent the different layers of the regulation cascade. In addition, two other RNA-based methods were adapted to study changes in gene expression of the SPI-1 genes during heterologous secretion over time: RT-qPCR and RNA sequencing. Better understanding of these cellular regulatory changes during heterologous secretion are likely to provide insights to engineering the SPI-1 secretion platform for increase robustness and productivity. I also describe some preliminary experiments on the impact of overexpressing key cellular factors on heterologous secretion.

### Developing genomic based transcriptional reporters

I designed transcriptional fusions of two different reporter proteins, GFPmut2 and mCherry, to five different genes in SPI-1 – *hilD*, *invE*, *prgH*, *spaS* and *sipC*. The fluorescent proteins were inserted in intergenic regions and near the middle of the operon to minimize disruption and polar effects as follows: *hilD*:GFP, *hilD*:GFP, *invE*:GFP, *spa*:mCherry, *prgH*:GFP, *prgH*:mCherry, *sipC*:GFP, and *sipC*:mCherry. The cells harboring these genomically incorporated fusions were tested under inducing conditions of low oxygen and high salt. These conditions were not known to affect the maturation of the fluorescent proteins. For comparison, I also tested the relevant plasmid-based promoter fusions – pHilD, pInvF, pPrgH and pSicA that were previously used in literature[43,60].

The temporal dynamics of the genomic transcriptional fusions were largely similar to the plasmid-based ones (Figure 5.1). The separation between the populations expressing mCherry

were less well resolved in certain cases, likely because this fluorescent protein is dimmer[161]. The transcriptional fusions did not alter secretion titer for invE, spaS, prgH, and *sipC* fusions but the *hilD* fusions had increased secretion titer (recall that this feature was explored in Chapter 3). Attempts to place promoter-GFPmut2 fusions at a neutral locus, *putAP*, were not successful[56]. There was a short homology between GFPmut2 and the chosen location which resulted in incorrect recombination that occurred at a much higher rate than the desired recombination.
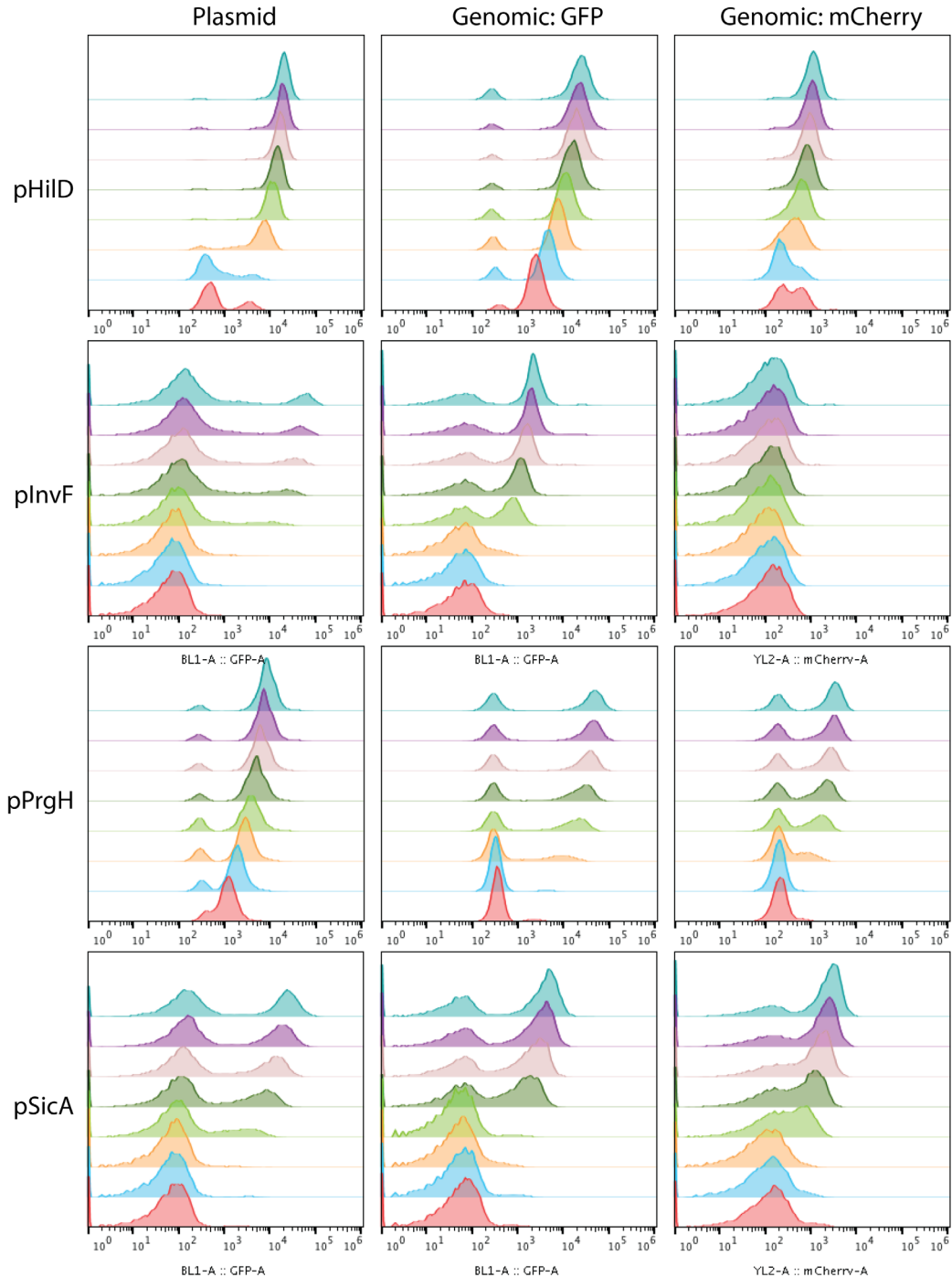


**Figure 5.1 Flow plots showing the changes in FP fluorescence over time. Both plasmid-based ones and genomic transcriptional fusions showed similar trends.**

These strains harboring this comprehensive set of transcriptional fusions can be used to rapidly test different culturing conditions. Another graduate student used them extensively to prototype and design media for increased heterologous secretion. As there was little difference in the temporal dynamics between the plasmid-based promoter fusions and the genomically encoded ones, the plasmid system is still useful for initial prototyping in engineered strains as transformation is a much easier procedure than recombineering.

## Developing RT-qPCR of SPI-1 genes

Direct measurements of gene expression levels can provide information on both transcriptional and post-transcriptional layers of regulation in SPI-1 mediated heterologous secretion. As such, I developed a reverse transcriptase quantitative PCR (RT-qPCR) approach to measure the expression of the genes in the SPI-1 operon by direct measurement of the RNA transcript levels. A key consideration was the choice of housekeeping gene as this would be used to measure RNA levels across different timepoints and different growth phases. Six housekeeping genes were chosen based on previous reports from the literature – rpoD, gmk, dnaN, sucA, glnS and eno. Seven SPI-1 genes were chosen to represent the important operons of SPI-1 – *hilD*, *hilA*, *invF*, *prgH*, *prgK*, *sicA* and *sipA*. Two sets of primers were designed for each gene and tested for efficiency using purified genomic DNA of DW01. The cycle threshold, $C_t$ at which amplified DNA can be detected by the system is a measure of gene expression. Linear regression was carried for each $C_t$ value to the respective number of genomes to determine the efficiency. All of the final primers had highly similar efficiencies (Figure 5.2).
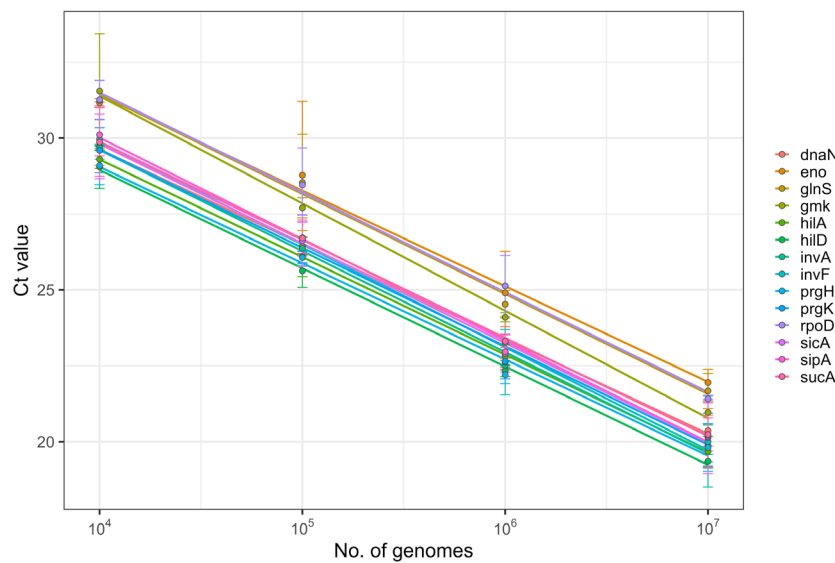


**Figure 5.2 The slopes of these primer (representing the PCR efficiencies) were similar. $C_t$ is the cycle threshold, which is a measure of gene expression levels. The slight differences in $C_t$ values observed between the different sets of genes were due to either using different preparation of standards and threshold settings variations from the thermocycler.**

The expressions of these genes were compared across three timepoints, namely 2, 4, and 6 hours, and two strains, wild-type and Δ*prgI* in non-inducing conditions. Two of the housekeeping genes, *glnS* and *eno* were used as the reference to calculate the ΔC$_t$ of each SPI-1 gene as they behave the most similarly and covaried across the different samples and had C$_t$ values closest to the SPI-1 genes. The mean C$_t$ of the reference genes were subtracted from each gene to derive

the $\Delta C_t$. $\Delta\Delta C_t$ values were subtracting the wild type 2-hour sample $\Delta C_t$ [115]. The expression of genes in the same operon followed similar patterns in the wild type but not in the $\Delta prgI$ strain (Figure 5.3). Interestingly, there was a spike in *prgH* level for $\Delta prgI$ at 6h that is not seen in *prgK*. The increase in expression of *sicA* for $\Delta prgI$ at 4h was also not seen in *sipA*. Given that the *prgI* gene lies between *prgH* and *prgK*, there might be some polar effects altering the expression levels of genes downstream of *prgI*. Although the primary promoter driving expression of *sicA* is the same as *sipA*, *sicA* can also be expressed as part of the *inv* operon. This might explain the difference in expression between the genes on the *sic/sip* operon.
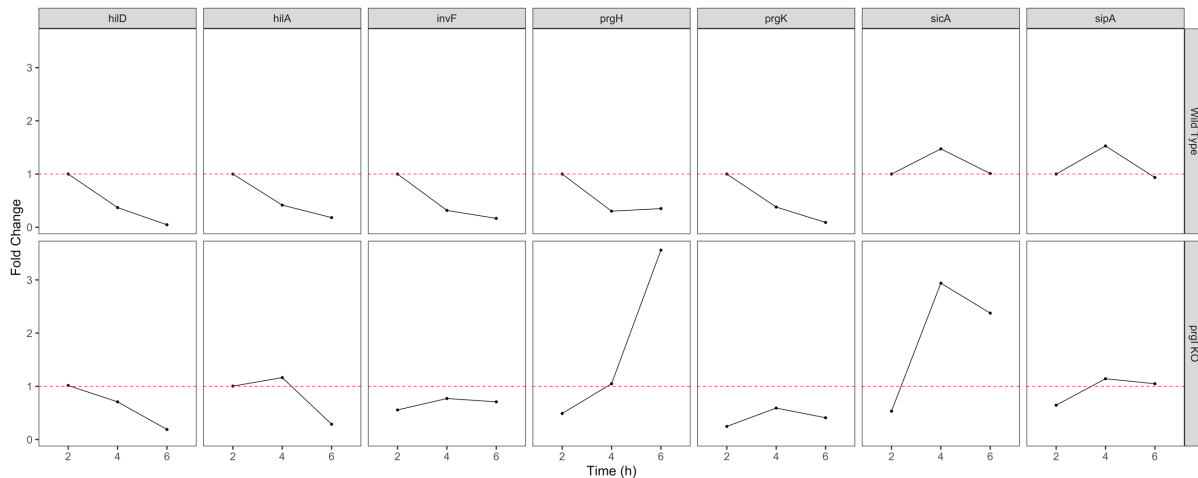


**Figure 5.3 Fold changes of the different SPI-1 genes over time normalized to the 2h timepoint for each strain. The mean $C_t$ of *glnS* and *eno* were used in the $\Delta\Delta C_t$ calculation.**

The accuracy of RT qPCR depends heavily of the choice of reference genes. It is difficult to determine if difference in $C_t$ values between samples are due to biological differences or technical error. There was little to no evidence in literature on the choice of housekeeping genes in most studies. Assuming that all of the housekeeping genes are indeed invariant in different conditions, they should all show similar patterns between samples. This was not the case here and thus a different approach to measure RNA transcripts level is needed.

## Understanding T3SS with transcriptomics

To circumvent the problem of selecting the right reference genes to normalize the different samples, I utilized RNA sequencing instead. There are well established bioinformatical approaches in RNA sequencing to normalize for differences between the libraries[162,163]. In addition, RNA sequencing would provide data in all the genes rather than just a select number of genes. The wealth of information from RNA sequencing can be used to 1) uncover covarying genes with SPI-1, 2) better understand which regulons are affected by *hilA* overexpression, 3) examine how metabolism might be altered by *hilA* overexpression, and 4) study changes in the *S. enterica* transcriptome over time. Given the decreasing cost of high throughput sequencing and greater availability of bioinformatics tools, this is no longer a specialized technique and can be successfully used in most labs. Three different timepoints were chosen; representing the start of SPI-1 activation (3 h), maximum secretion (6 h) and shutting down of the system (10 h) instead of the previous time points used in RT-qPCR.

Cultures were subjected to a total of 15 conditions as listed in Table 5.1. As the choice of growth medium had been shown to have an impact on both the transcriptome and phenotype, we wanted to include the base defined medium being developed for enhanced heterologous secretion, NCE supplemented with amino acids[164]. To account for the increase in lag phase when grown in NCE with amino acids, the cultures in NCE were sampled an hour later than those grown in LB-L, ensuring similar $OD_{600}$ during each sampling.

| Media | Strain | Timepoints |
|---|---|---|
| **LB-L** | pLac none, pLac *hilA* | 3, 6 and 9 hours |
| **NCE with amino acids** | pLac none, pLac *hilA* | 4, 7 and 10 hours |
| **LB-IM** | pLac none | 3, 6 and 9 hours |

**Table 5.1 A total of 15 conditions and 45 samples were sent for RNA sequencing**

Shown in Figure 5.4 are the cultures grown in LB-IM under inducing conditions similar to the conditions used in RT-qPCR (Figure 5.3) to compare the two methods. The counts per million (CPM) for each gene was normalized using DESeq2[150]. The expression patterns from RNA sequencing were completely different from RT-qPCR (Figure 5.4). This could be explained by the choice of housekeeping genes used to normalize the fold change in Figure 5.3. None of the housekeeping genes were invariant across time points (Figure 5.4) and thus the aberrant results from RT-qPCR were due to normalizing samples across different growth phases using *glnS* and *eno*. Given the narrow spread of normalized read counts across the samples at each time point, we concluded that the genes, *dnaN*, *eno*, *glnS*, *gmk*, and *rpoD* can still serve as useful reference genes in limited situations, but not those of interest to our work.
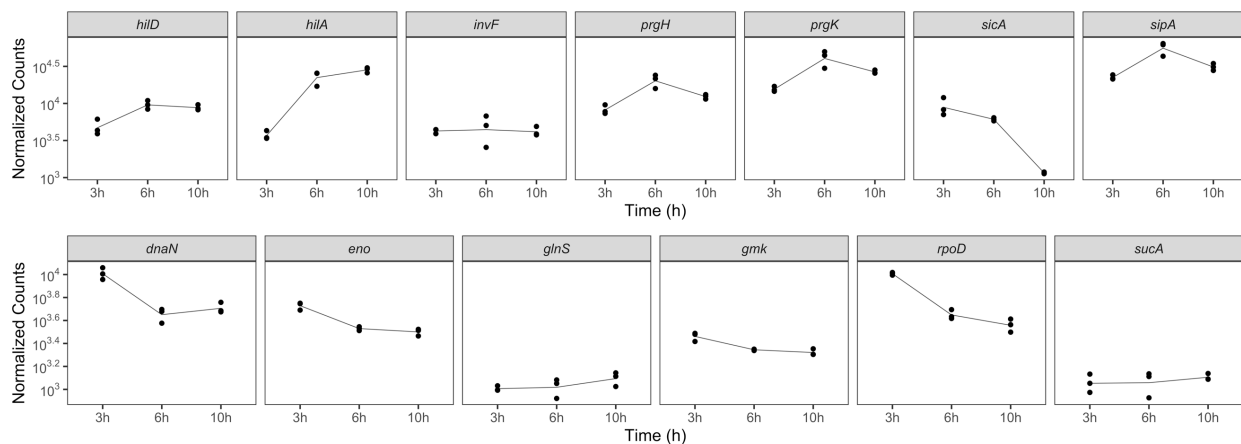


**Figure 5.4 Normalized counts per million of SPI-1 genes and housekeeping genes that were used in RT-qPCR over time in LB-IM. Longer genes will result in more counts and thus across gene comparison should not be done using this data.**

To look at the impact of *hilA* overexpression on SPI-1, we compared the normalized CPM between DW01/pLac empty vector and DW01/pLac *hilA* grown in LB-L. As expected, the *hilA* expression levels of DW01/pLac *hilA* did not change over time (Figure 5.5 d). This provided further evidence that synthetic expression from the pLac promoter in *S. enterica* is not growth dependent. The expression of the SPI-1 genes downstream of *hilA* in the regulatory cascade were higher when *hilA* was overexpressed across all time points. However, we observe a general

decrease in transcript levels over time. The *inv* and *prg* operons under the direct control of HilA decreased in expression over time while the *sic/sip* and *sop* genes under the control of InvF stayed on between 3 to 6 hours (Figure 5.5). In addition, there were some observable differences between the two conditions in the SPI-1 genes that lie above *hilA* in the regulation cascade – *hilD*, *hilC*, and *rtsA* (annotated as STM4315 in the NCBI database)[118,165]. Differential gene expression analysis would need to be carried out to determine if the observed differences are significant. There was more evidence that the housekeeping genes identified for RT-qPCR were not suitable for normalization (Figure 5.6).
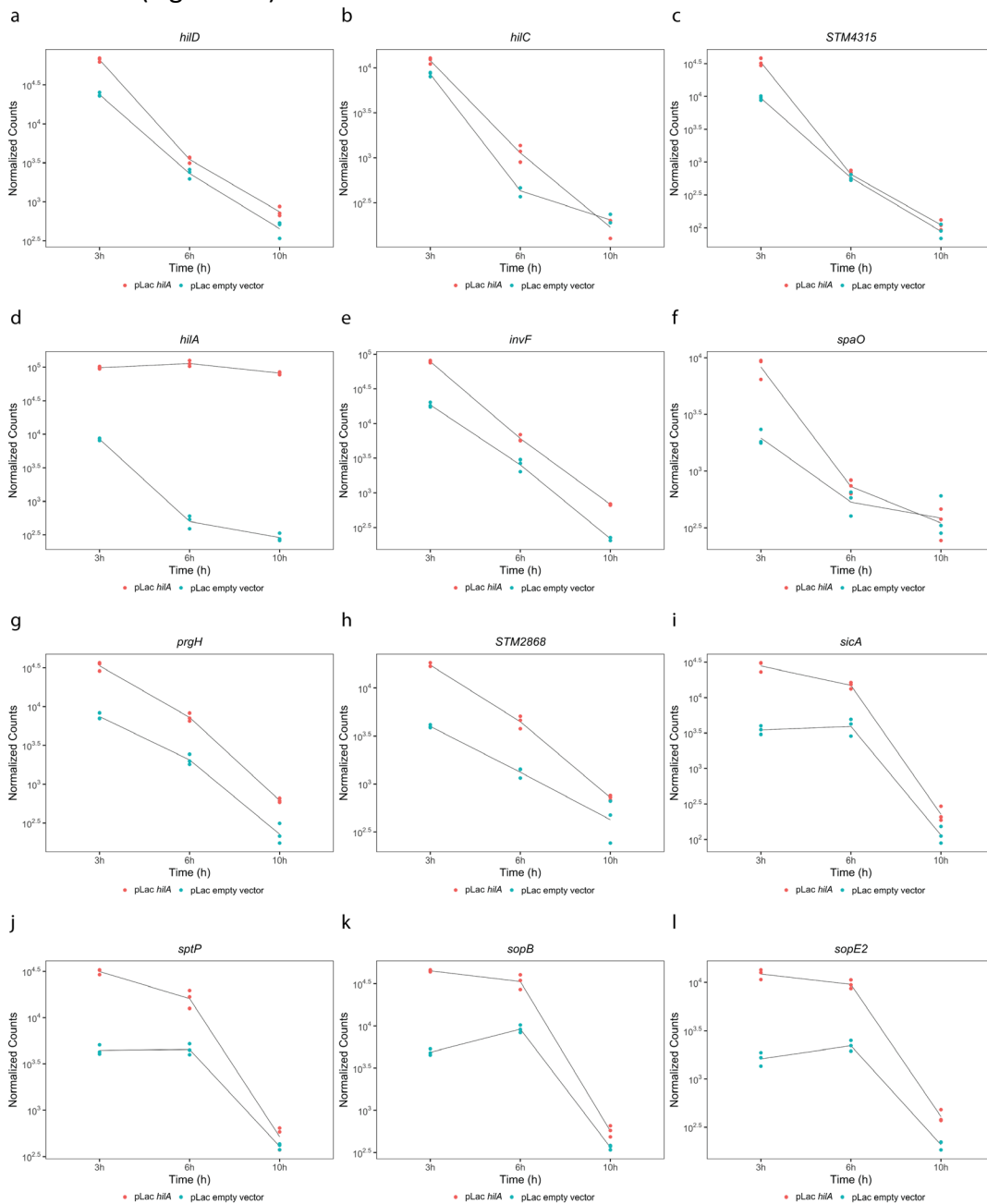


**Figure 5.5 Normalized CPM of different SPI-1 genes were derived from DESeq2. Most of the gene expression in pLac *hilA* were higher than pLac empty vector. Two genes from each operon are shown here, one at the start and one near the end of the operon. The expression patterns within each operon were similar as expected except between *invF* (e) and *spaO* (f).**

58
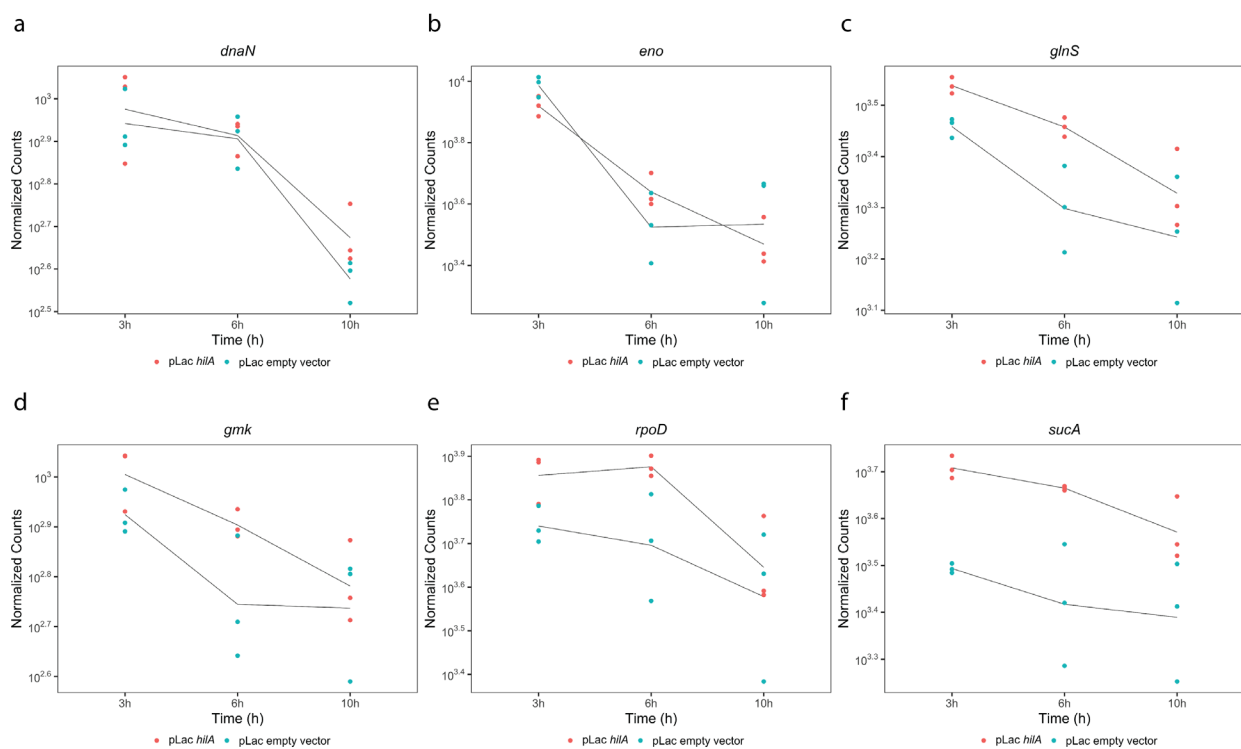
**Figure 5.6 Normalized CPM of the different housekeeping genes in pLac *hilA* and pLac empty vector over time showed a fair amount of variation.**

## *Exploratory analysis of transcriptomics data*

RNA sequencing provides a wealth of data, in which thousands of variables are measured. This allowed for the use of unsupervised machine learning algorithms to have a better grasp of the overall differences/similarities between the samples. I utilized principal component analysis (PCA) to look at the relationship between the samples using the top 500 varying genes (Figure 5.7 and Figure 5.8).

Principal component (PC) 1 separated out the samples by time and explained 60% of the variance. PC2 separated out the samples by strain, including whether *hilA* was overexpressed or not, and explained 11% of the variance. Both PC3 and PC4 were needed to separate out samples by media and together they explained 10% of the variance. Based on this data, we concluded that the strongest effect on the transcriptome of *S. enterica* is time. Both media and strain explained similar amounts of variance observed in the transcriptomic data.

These samples can also be clustered using hierarchal clustering. Hierarchal clustering was used to determine the relatedness of the different samples (Figure 5.9). The resulting heatmap shows the samples clustering in two major groups which reflects the grouping seen along PC1 of Figure 5.7). It is surprising that the LB-IM samples cluster together with sampled from first timepoint of cells grown in LB-L (3 hours) and NCE (4 hours). This could be due to OD$_{600}$ or growth phase of the culture in LB-IM being closer to the cellular state of cells at first timepoint in LB-L and NCE.

59

Using both PCA and hierarchal clustering provides insight on how similar or dissimilar the samples are to each other. Experimental outliers can also be detected visually through the use of these tools. The PCA plot also allows us to visualize the overall effect of experimental covariates and batch effects. Note that time could also be a reflection of cell density or growth phase of the culture. Changes in transcriptome due to cell density is mediated by the quorum sensing systems of *S. enterica* and is a known input to SPI-1[166–168].
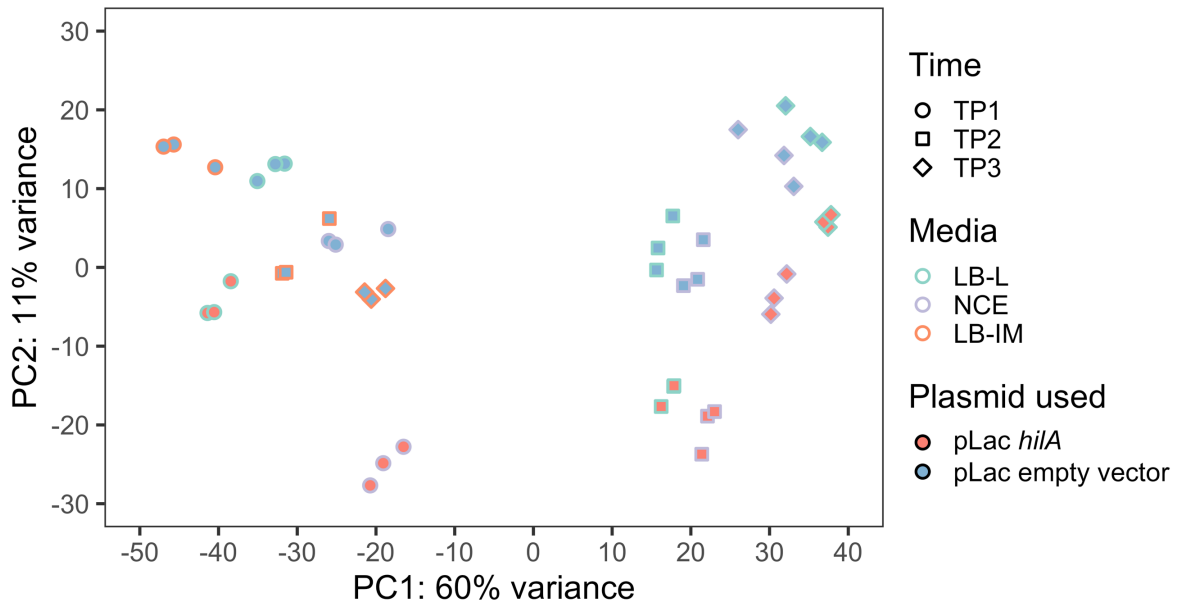


**Figure 5.7 PCA plot with PC1 and PC2 of all 45 samples using the top 500 varying genes as determined by DESeq2. 72% of the total variance can be explained by these two principal components.**
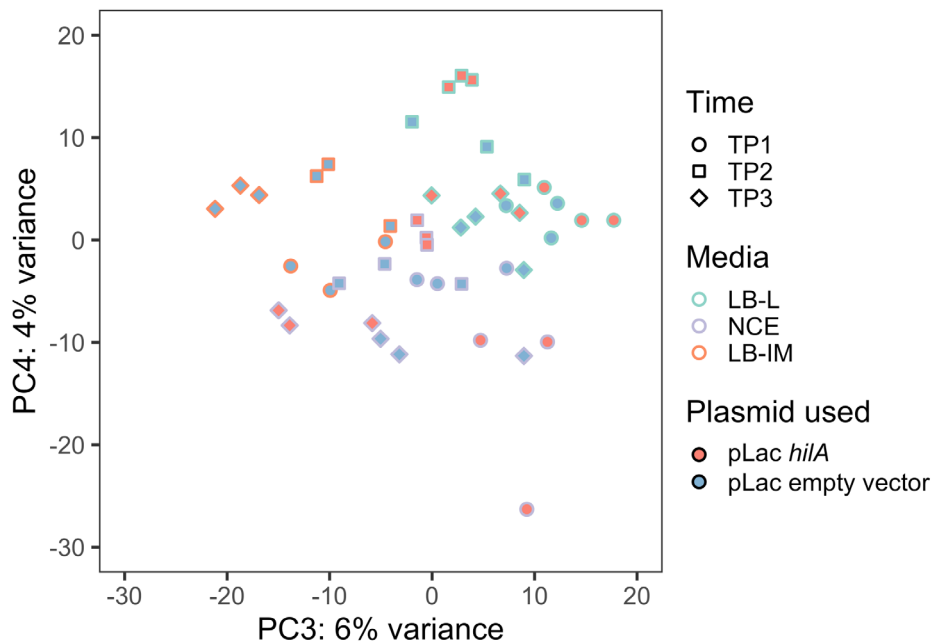


**Figure 5.8 PCA plot with PC3 and PC4 of all 45 samples using the top 500 varying genes as determined by DESeq2. Only 10% of the total variance can be explained here. Both are needed to separate out the samples by media.**
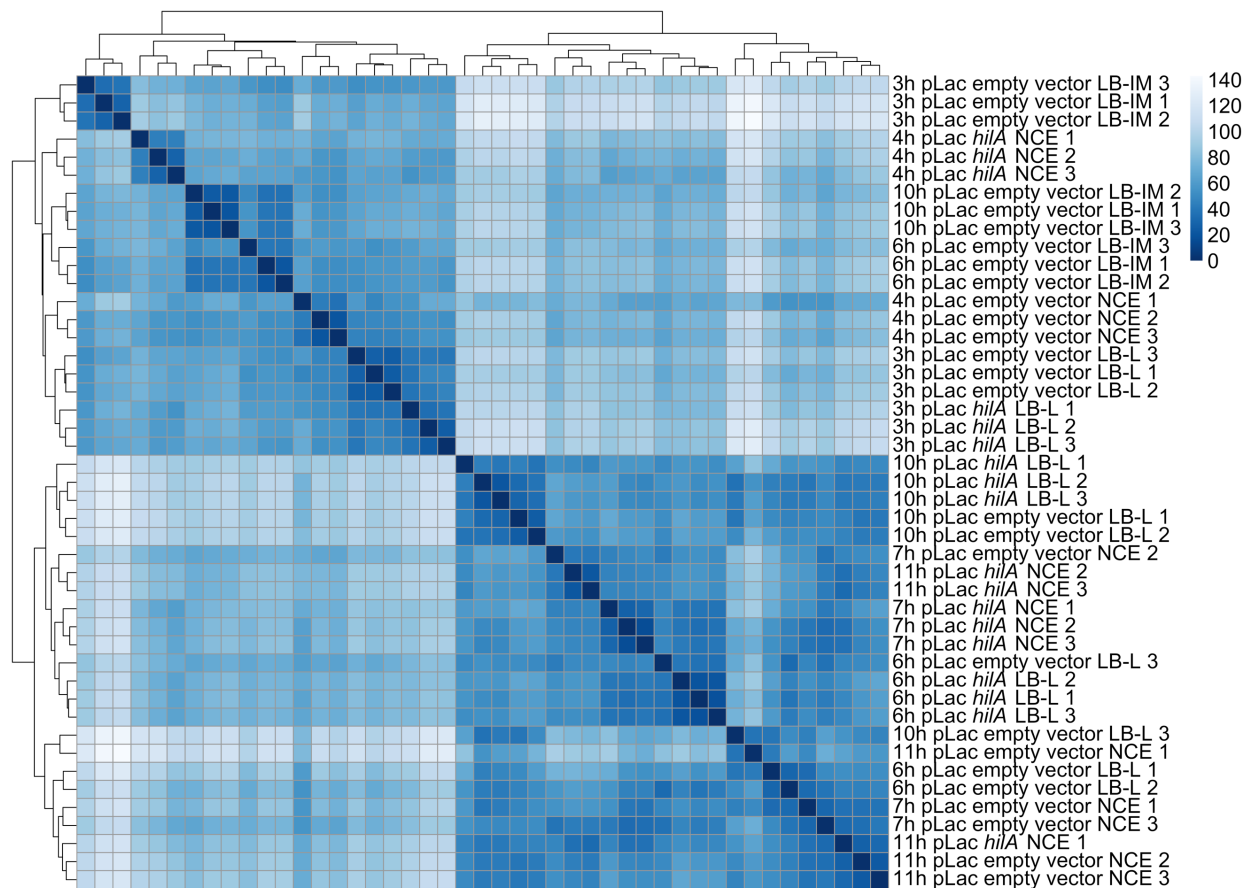
**Figure 5.9 Hierarchal clustering of the 45 samples used in RNA sequencing. The distance matrix was based on the normalized counts from DESeq2[150]. This shows the relatedness of each sample to each other with the bigger numbers showing a greater difference/distance between the samples.**

## Determining cluster of covarying genes using weighted correlation network analysis

The data in Figure 5.5 suggests that differences in cellular physiology and state might be affecting SPI-1 expression even in the presence of constant *hilA* expression. Pulling out what these cellular factors may be can help guide engineering to extend the productive phase of our system. We reasoned that we could use weighted correlation network analysis (WCGNA), we would be able to uncover clusters of genes that covary with SPI-1 genes. Such clusters of genes may be under the direct control of specific global regulators and contribute to the cellular state associated to SPI-1 heterologous secretion.

The WGCNA R package was used to identify covarying clusters/modules using all 45 samples[169]. All the genes were clustered into different modules through hierarchal clustering and dynamic tree cuts (Figure 5.10). A total of seven modules were identified and subsequently correlated to the three different explanatory variables – time, media and strain (Figure 5.11). The MEgreen module was found to be significantly correlated to differences in strain – DW01/pLac empty vector and DW01/pLac *hilA*. Interestingly, most of the SPI-1 genes could be found in the MEgrey module. The MEgrey module was significantly correlated with time. This provides further evidence that time/growth-driven changes to SPI-1 are due to specific cellular programming.
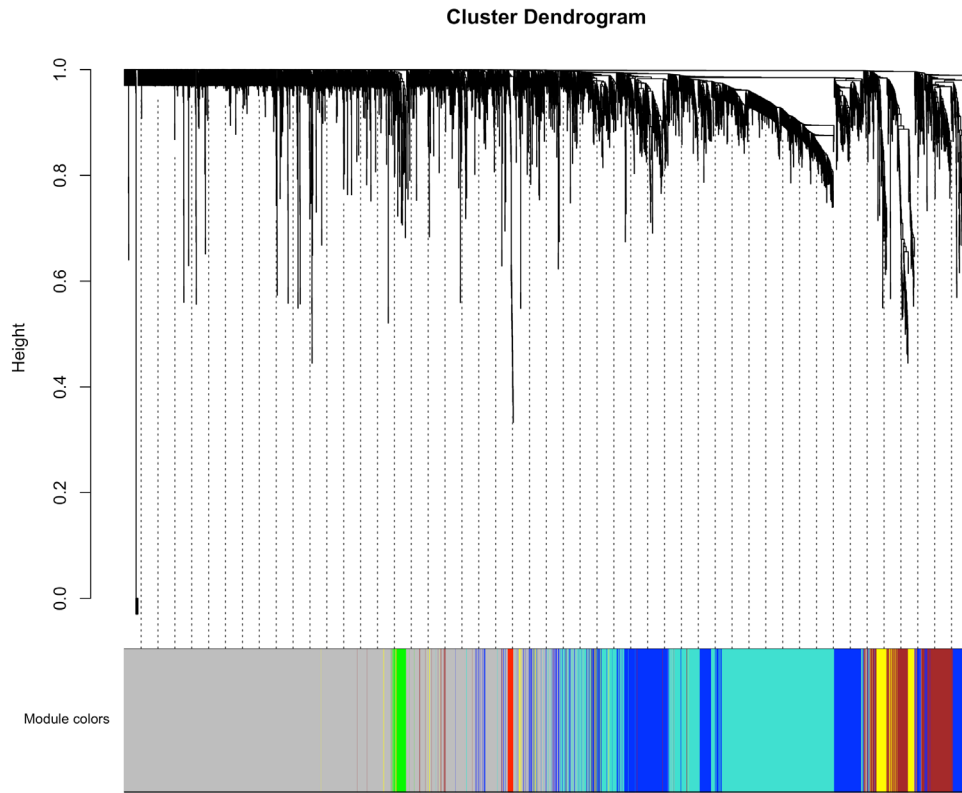
**Figure 5.10 Hierarchal clustering was carried out by WGCNA. After which, dynamic tree cuts were made to derive the different cluster of covarying genes or modules.**
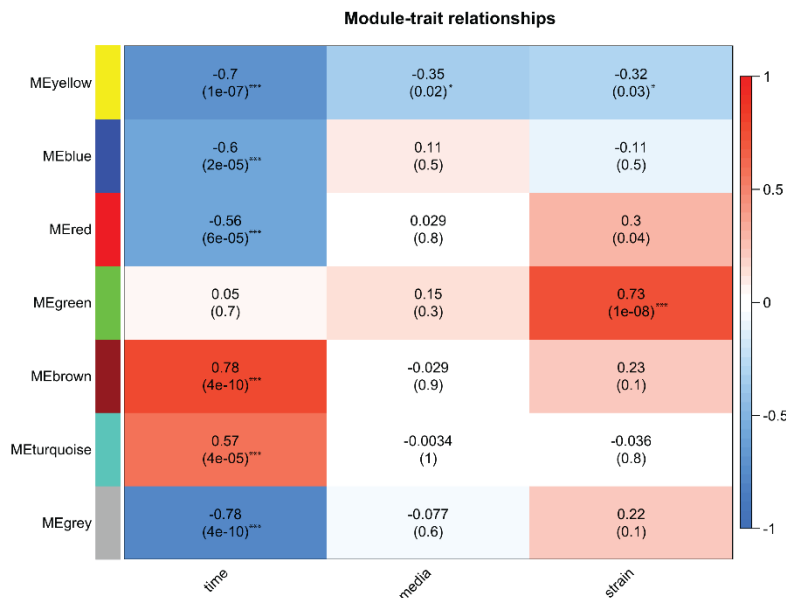


**Figure 5.11 Each module called by WGCNA relationship was tested for correlation to the three different factors – time, media and strain. Surprisingly none of the modules were significantly correlated to media. An asterisk indicated p < 0.05, two asterisks indicated p < 0.01 and three asterisks indicated p < 0.005.**

To have a better understanding of the overall network topology, we perform hierarchal clustering of the different network modules identified by WGCNA. This allows us to identify meta-

modules and determine which modules are connected[169]. The clustering is based on the eigengene calculated by WGCNA (Figure 5.12 a). The eigengene can be considered the representative gene expression profile for each module as calculated by PCA. An eigengene was also calculated for strain. There were three major meta modules that were connected by the MEgreen module and strain.
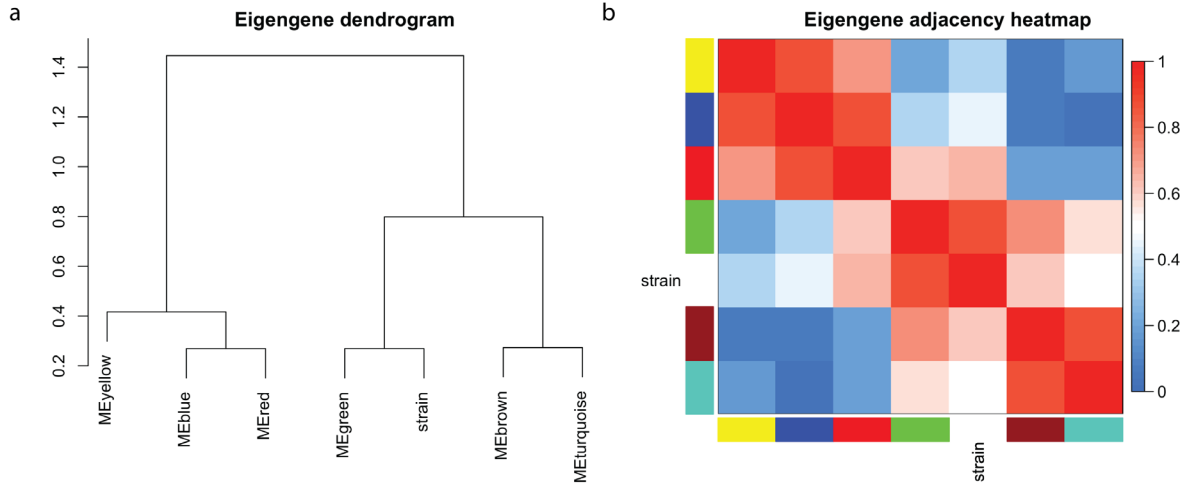


Figure 5.12 The different modules are hierarchically clustered together with the eigengene of the strain variable. The dendogram in (a) shows how closely related the modules were. (b) A heatmap showing three major meta-modules that are interconnected. Only values > 0.5 was considered to be connected. The MEgrey module was dropped by the WGCNA package.

The MEgreen module consisted of genes involved in periplasmic remodeling such *pagP* and *pdgL* and genes regulated by PhoP. The complete network is shown in Figure 5.13. The genes with the most connection were *ydjA* and *clpB* (70 connections each). The weight of the edge signified the strength of the correlation between the different gene nodes.
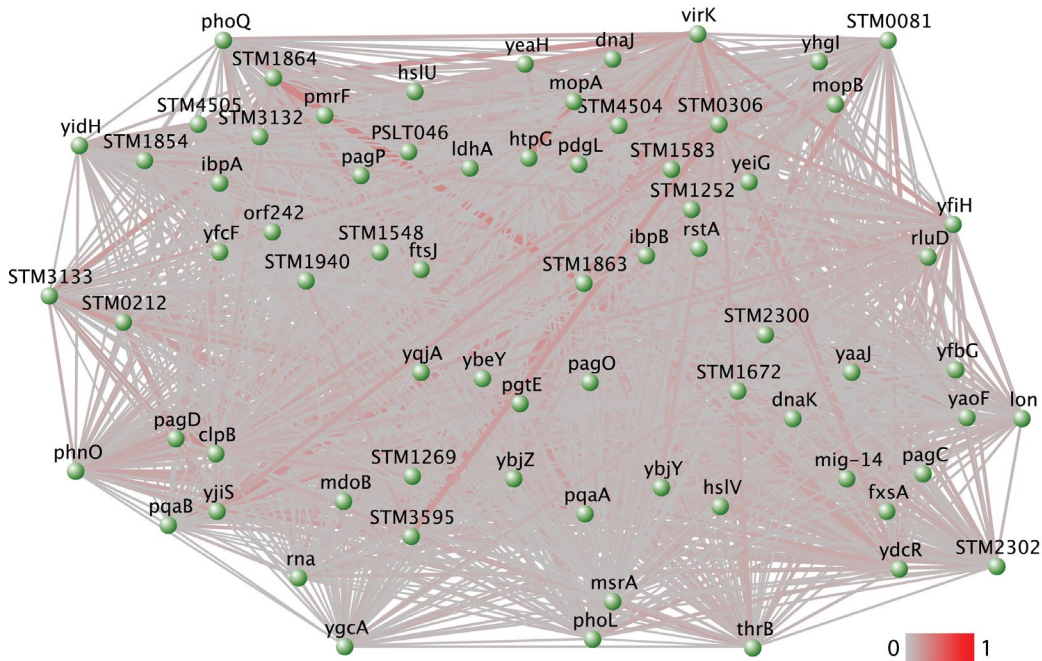


Figure 5.13 Visualization of the network in the MEgreen module using VisANT[170]. The edges were colored by the weight of the edge calculated by WGCNA.

63

## Determining the regulatory nodes for gene clusters

As mentioned earlier, it is possible to pick out genes that were under the regulation of PhoP. In order to systematically identify the transcriptional regulation of the genes in each cluster, I employed the Salmonet database. The Salmonet database is an integrated network for *Salmonella* with information on metabolic, transcriptional regulatory and protein-protein interaction. Using the transcriptional regulatory layer, I mapped each gene to its transcriptional regulator to determine the regulatory master node of the MEgreen module. The genes in the MEgreen module were mainly regulated by *phoP* and *hilA* (Figure 5.14).
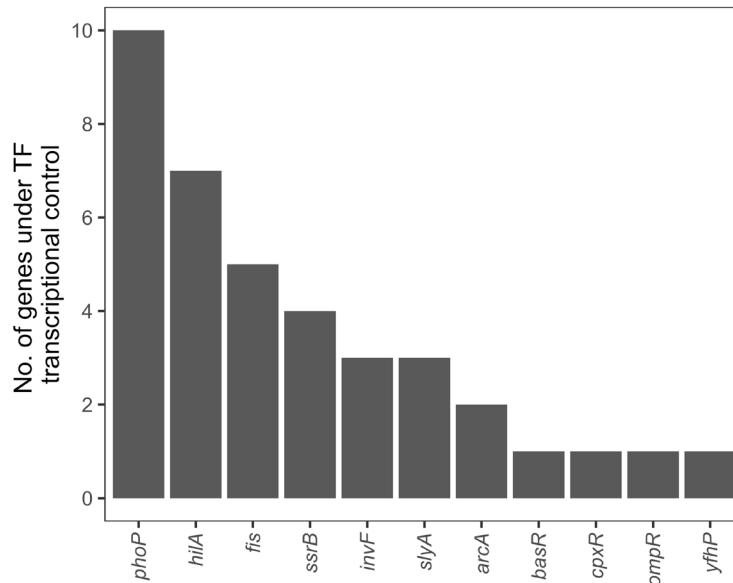


**Figure 5.14 The transcriptional regulators of the different genes in MEgreen module.**

## HilA overexpression alters the transcriptome

To focus on the impact of *hilA* overexpression on the transcriptome, we focus on two sets of conditions – DW01/pLac empty vector and DW01/pLac *hilA* grown in LB-L – and carried out similar exploratory data analysis. Figure 5.15 shows the PCA plot using the top 500 varying genes as determined by DESeq2[150]. Here we can clearly see that PC1 accounted for 79% of the variance and separated the samples by time points. PC2 accounted for 12% of the variance and separated the samples by strain. The variance explained by PC2 was similar to that observed in Figure 5.7, while the variance explained by PC1 was increased. Both PCs accounted for ~100% of the variance observed between the samples.
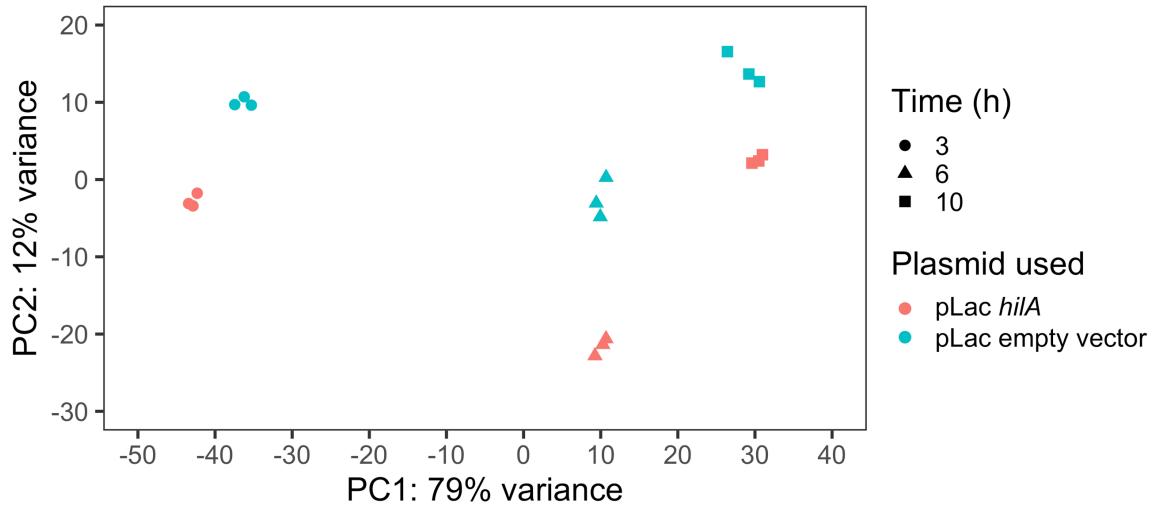
**Figure 5.15 PCA analysis was carried using top 500 varying genes determined by DESeq2. When more genes were included, the patterns shown here do not change and the separation between the different conditions was less pronounced.**

Each PC was derived through the summarization of the original characteristics – the top 500 varying genes. By looking at the loadings of each PC, we are able to look at the influence of each gene on the PC. Furthermore, by looking at how the loadings affect the PC, we are able to draw inferences about the correlation between the loadings. Loadings that group together and differ by a small angle are likely to be positively correlated. Loadings that diverge and are close to 180° are likely to be negatively correlated. Loadings that differ by an angle close to 90° are not correlated. The contribution of loadings for both PCs can be seen in Figure 5.16. The top 10% loadings for PC2 were plotted to show each loading's direction (Figure 5.16c). The SPI-1 genes, including *hilA*, were found in the third quadrant as were recently identified secretion effectors such as STM1239[171]. Four gene, *ymdF*, *yciG*, *ynfM*, and *yiaG*, were negatively correlated to the SPI-1 genes (first quadrant). *YmdF* and *yciG* are paralogs; *yciG* is known to be induced by an herbicide, salicylate and high salt conditions. *YnfM* is an uncharacterized member of the major facilitator superfamily of transporter. *YiaG* is a putative transcriptional regulator.
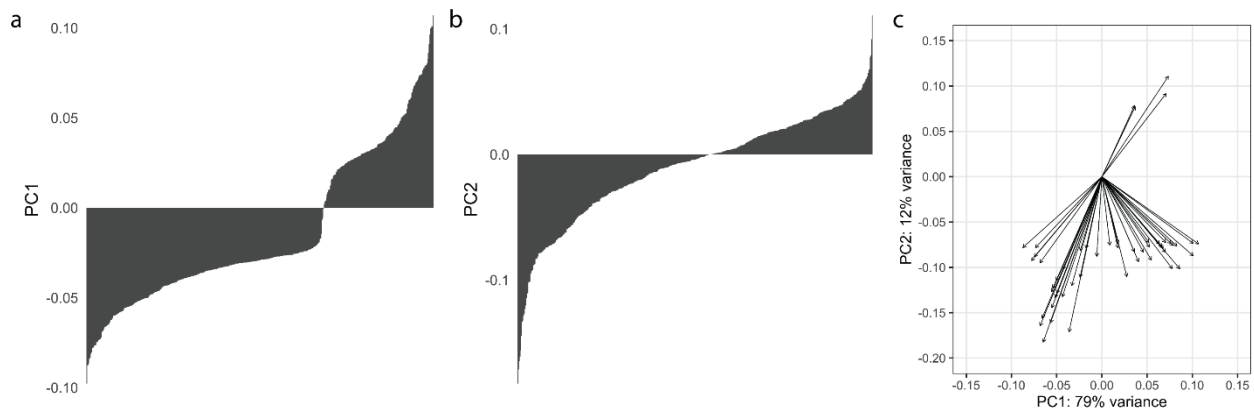


**Figure 5.16 Histogram of each loading's contribution to (a) PC1 and (b) PC2. (c) Each arrow represents the loading of a gene. Only 50 genes with the greatest contribution to PC2 were shown here. The gene names are not shown here due to the tight clustering of the arrows.**

65

Interestingly the hierarchal clustering revealed that the DW01/pLac *hilA* samples at 10 hours clustered tightly with DW01/pLac empty vector at 10 hours and with DW01/pLac empty vector at 6 hours (Figure 5.17). This could represent the shutting down of the SPI-1 cellular program.
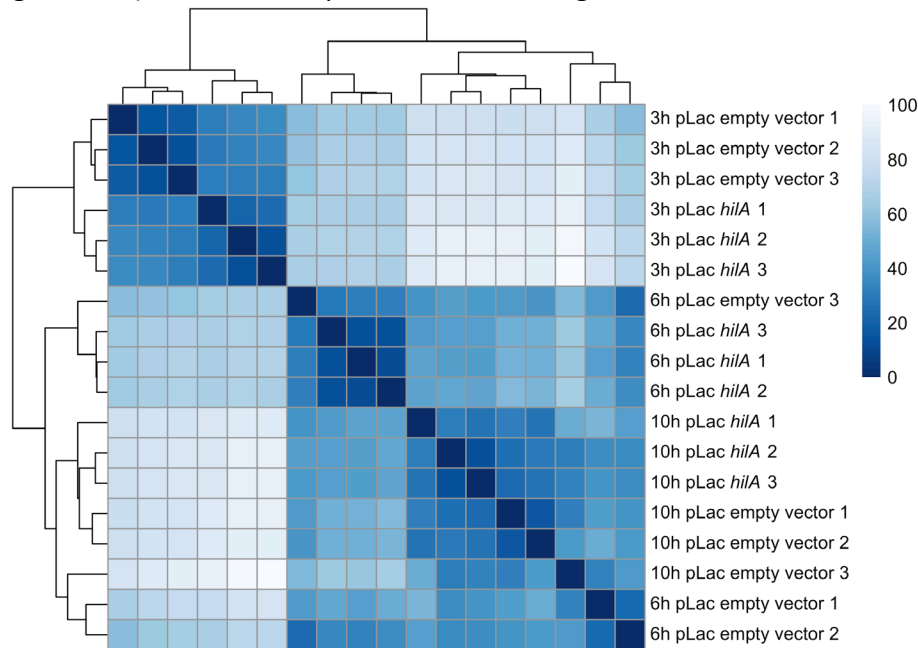


**Figure 5.17 Hierarchal clustering of the samples using normalized counts from DESeq2[150]. This shows the relatedness of each sample to each other with the bigger numbers showing a greater difference/distance between the samples.**

### Differential gene expression caused by hilA overexpression

To determine changes in the transcriptome caused by *hilA* overexpression, the full DESeq2 pipeline was run. Hypothesis testing was carried out using a Wald test, with time as a factor in the overall model used by DESeq2. 658 genes were upregulated, and 469 genes were downregulated, representing 14% and 10% of all the genes tested respectively (FDR > 0.05).

As there were many genes differentially regulated, I hypothesized that the cells are experiencing specific cellular programming. To determine which regulons may be involved in shaping the cellular state, I took the approach of mapping the genes back to their transcriptional regulators as described above. Figure 5.18 showed strong roles played by *fis*, *phoP*, *crp*, *ssrB*, *arcA*, *fur* and *fnr*. This is not surprising as most of these systems are known to have an impact on SPI-1[67,68,70–72,134,156].
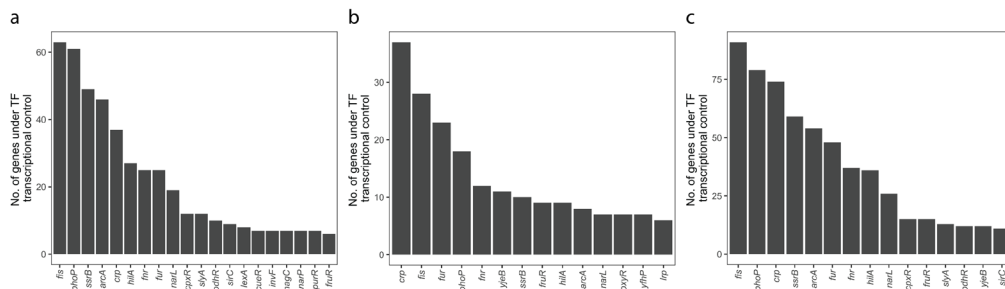


**Figure 5.18 The transcriptional regulators of the different differentially expressed genes. A cutoff of 5 was used for (a) upregulated genes and (b) downregulated genes. A cutoff of 10 was used for (c) all differentially expressed genes.**

The normalized expression of the differentially expressed genes are shown for each transcriptional regulation node of *fis* (Figure 5.19), *phoP* (Figure 5.20), *crp* (Figure 5.21), *ssrB* (Figure 5.22), *arcA* (Figure 5.23), *fur* (Figure 5.24) and *fnr* (Figure 5.25).
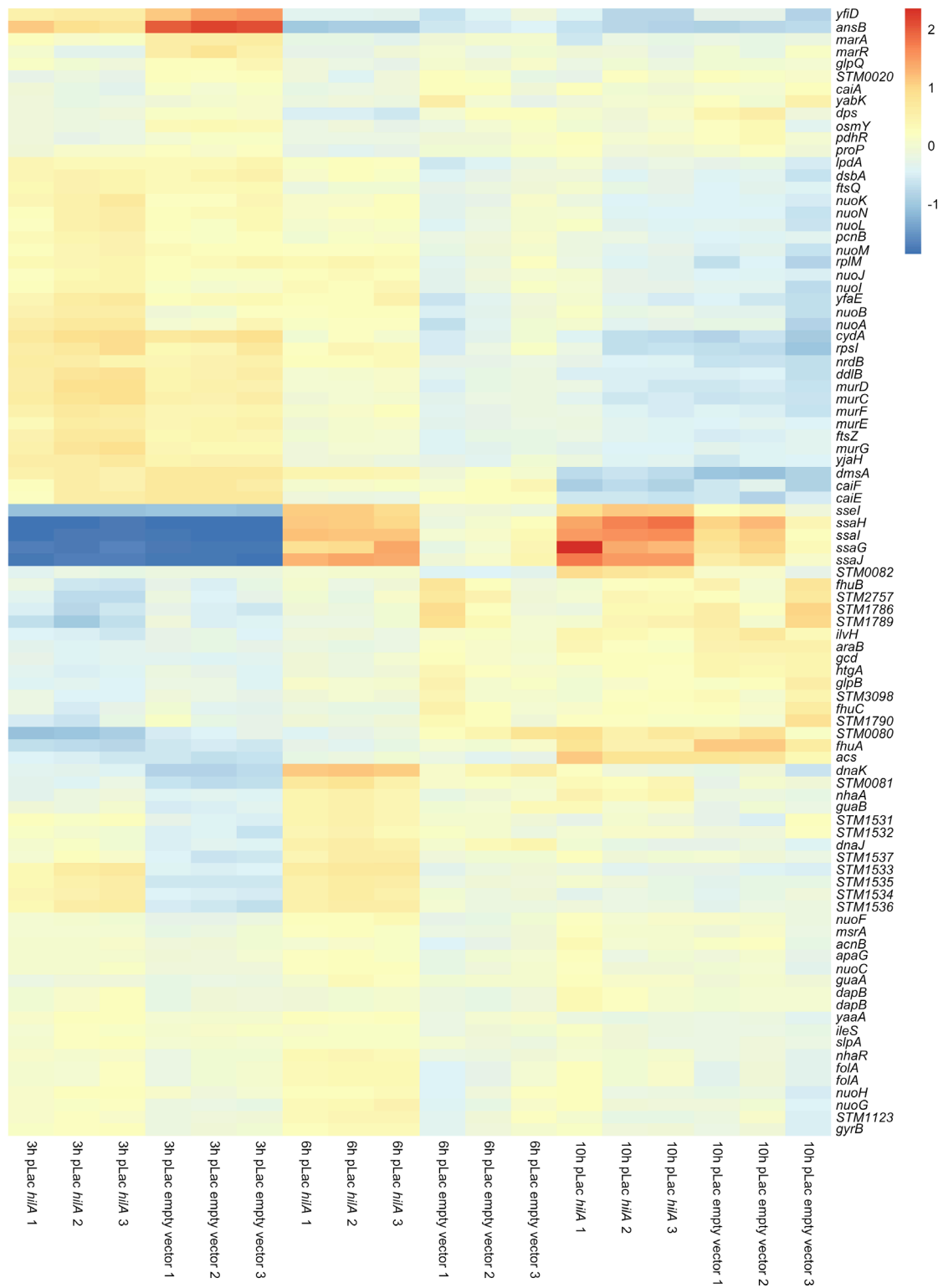


**Figure 5.19 Heatmap showing the regularized log transformed read counts of each gene under the nucleoid factor *fis*.**
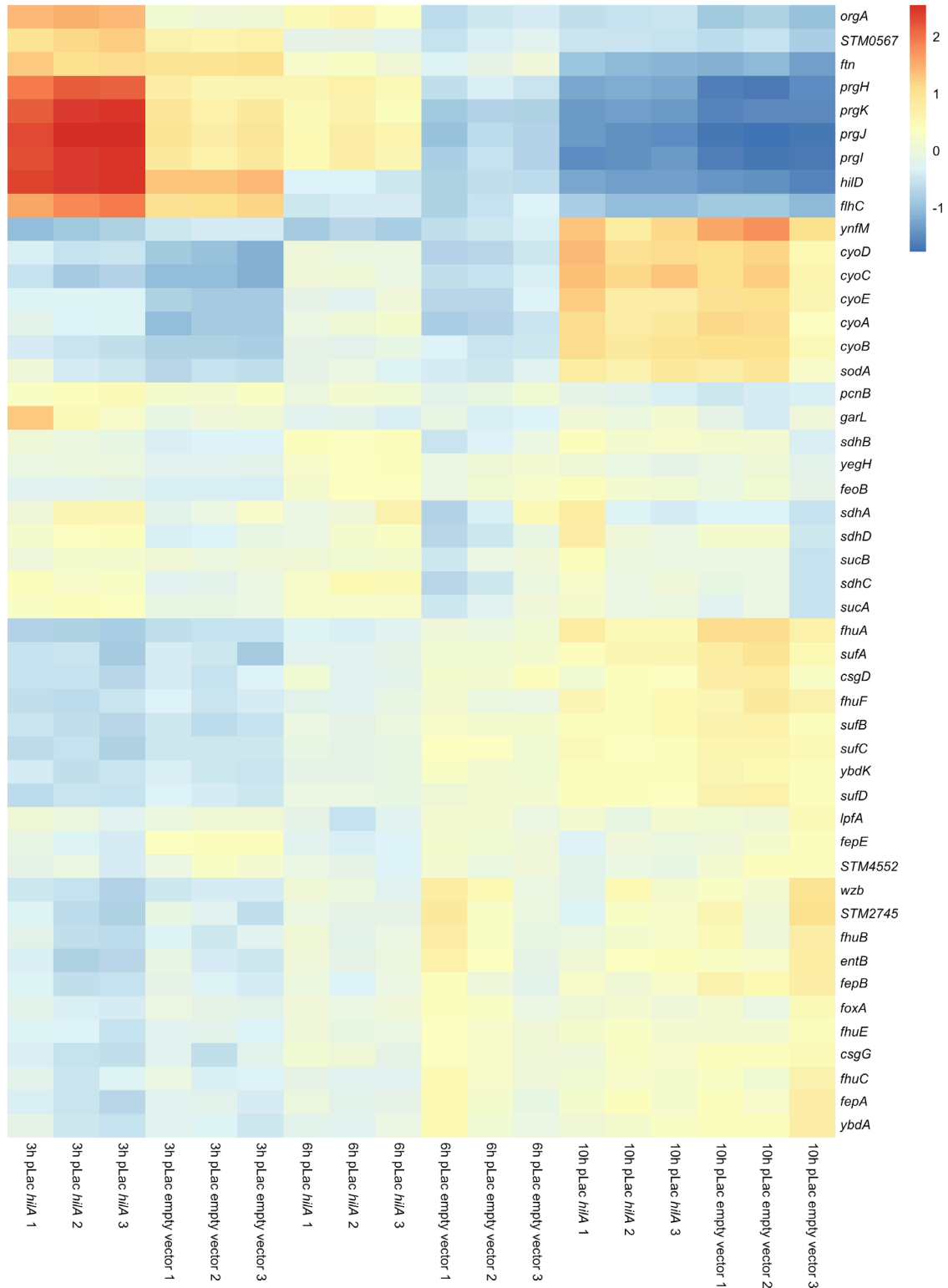
**Figure 5.20 Heatmap showing the regularized log transformed read counts of each gene under the transcriptional regulatory protein *phoP*.**
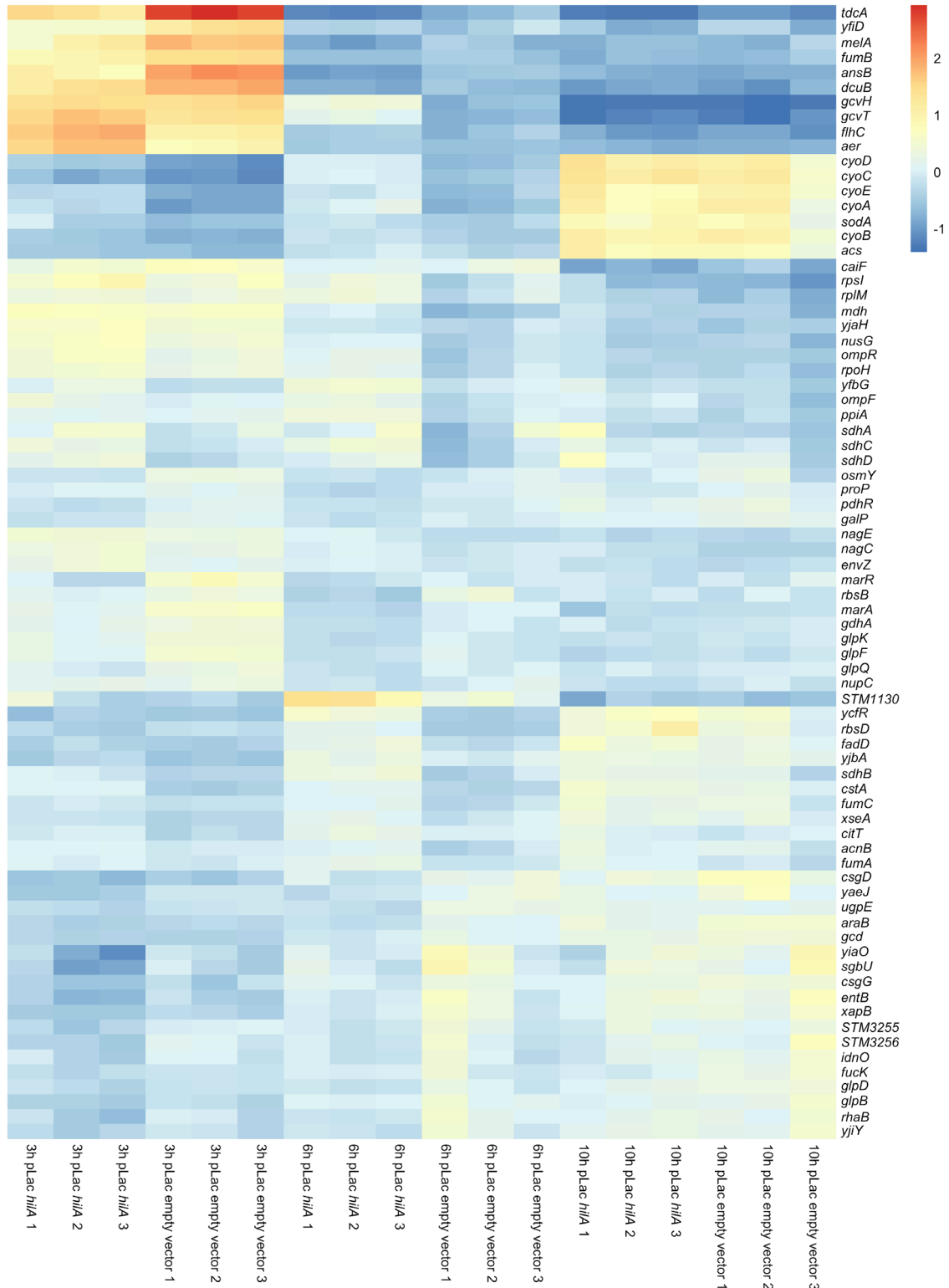
**Figure 5.21 Heatmap showing the regularized log transformed read counts of each gene under global transcriptional regulator *crp*.**
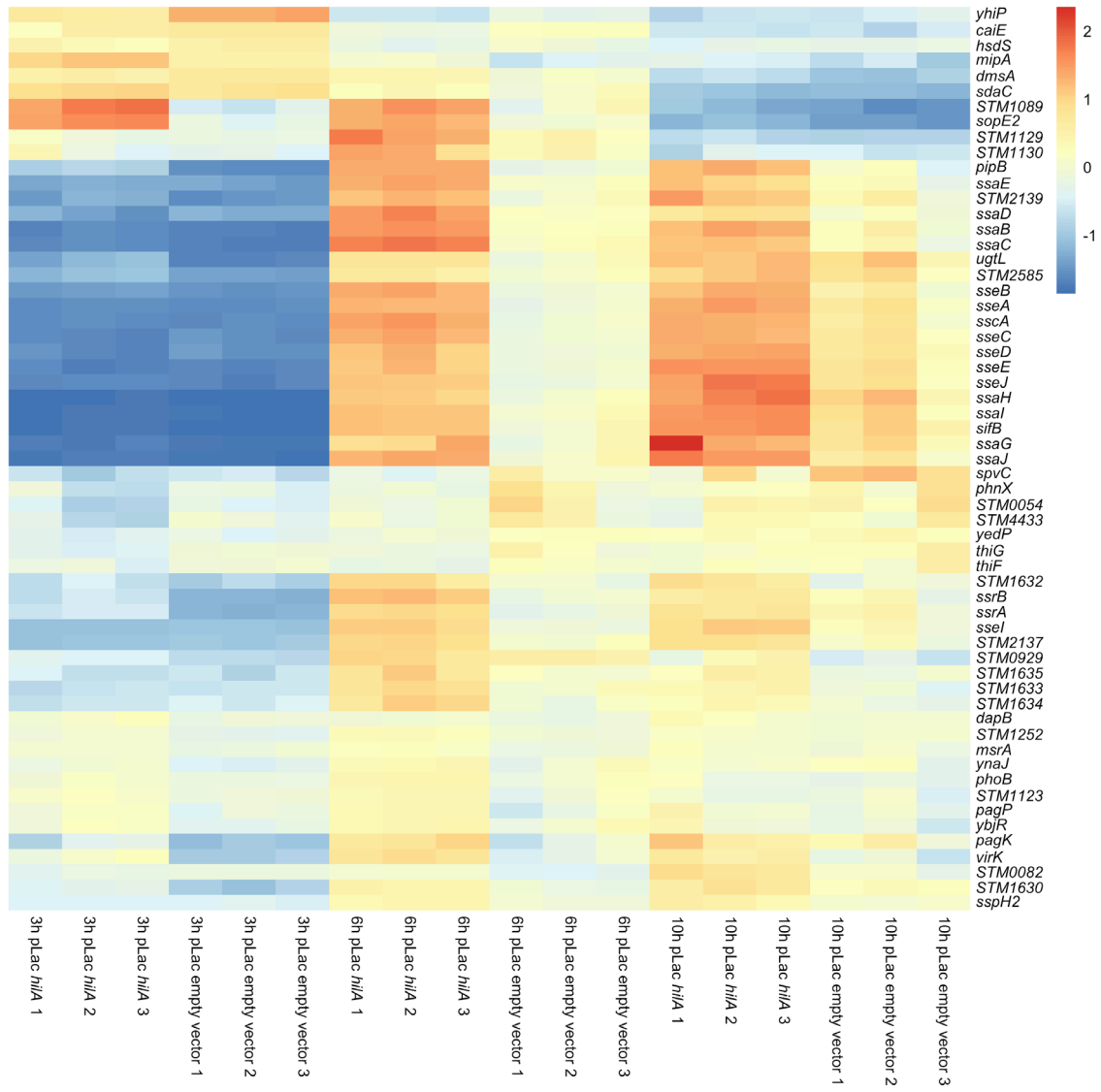
**Figure 5.22 Heatmap showing the regularized log transformed read counts of each gene under SPI-2 transcriptional activator *ssrB*.**
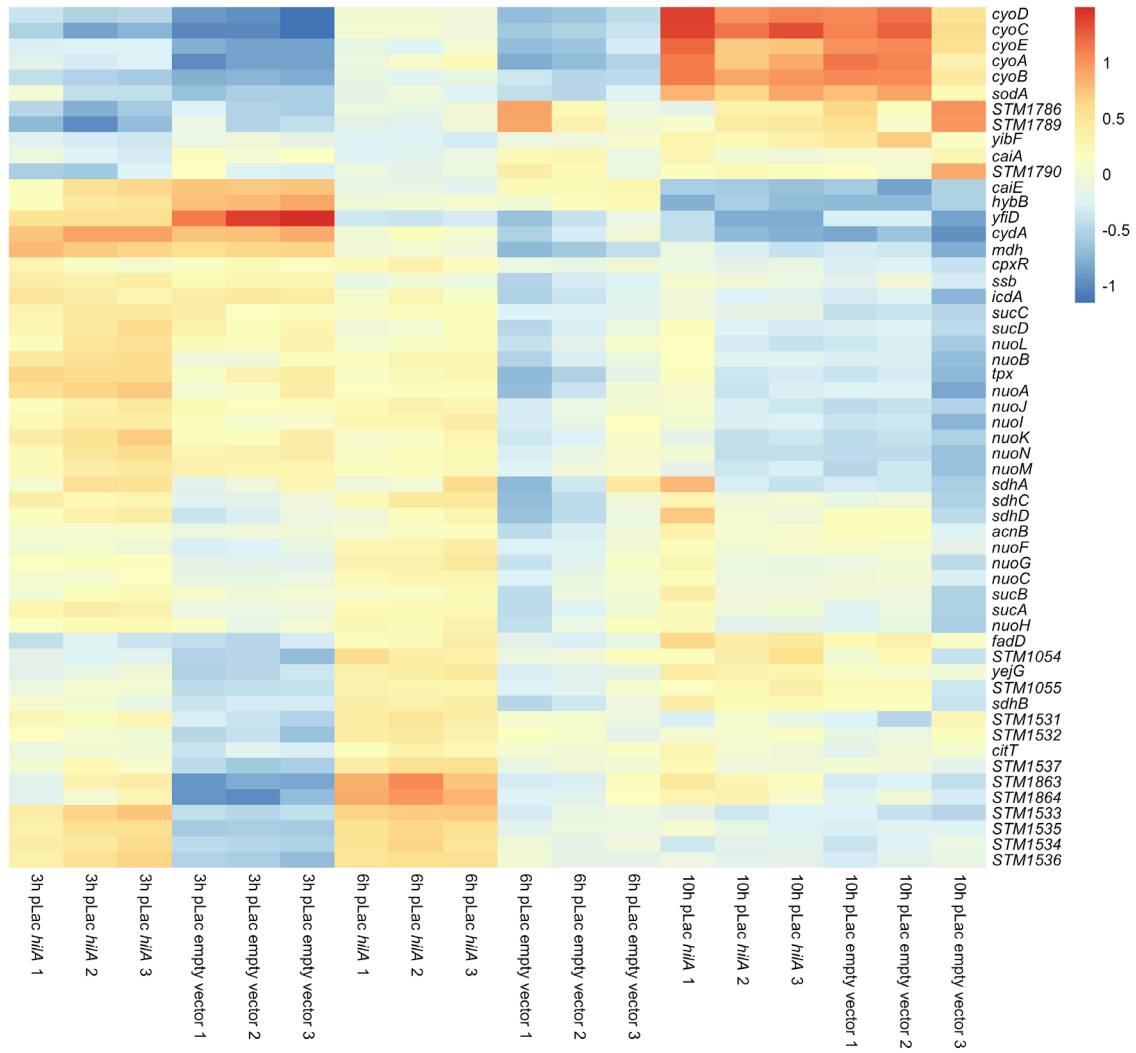
**Figure 5.23 Heatmap showing the regularized log transformed read counts of each gene under transcriptional regulator *arcA*.**
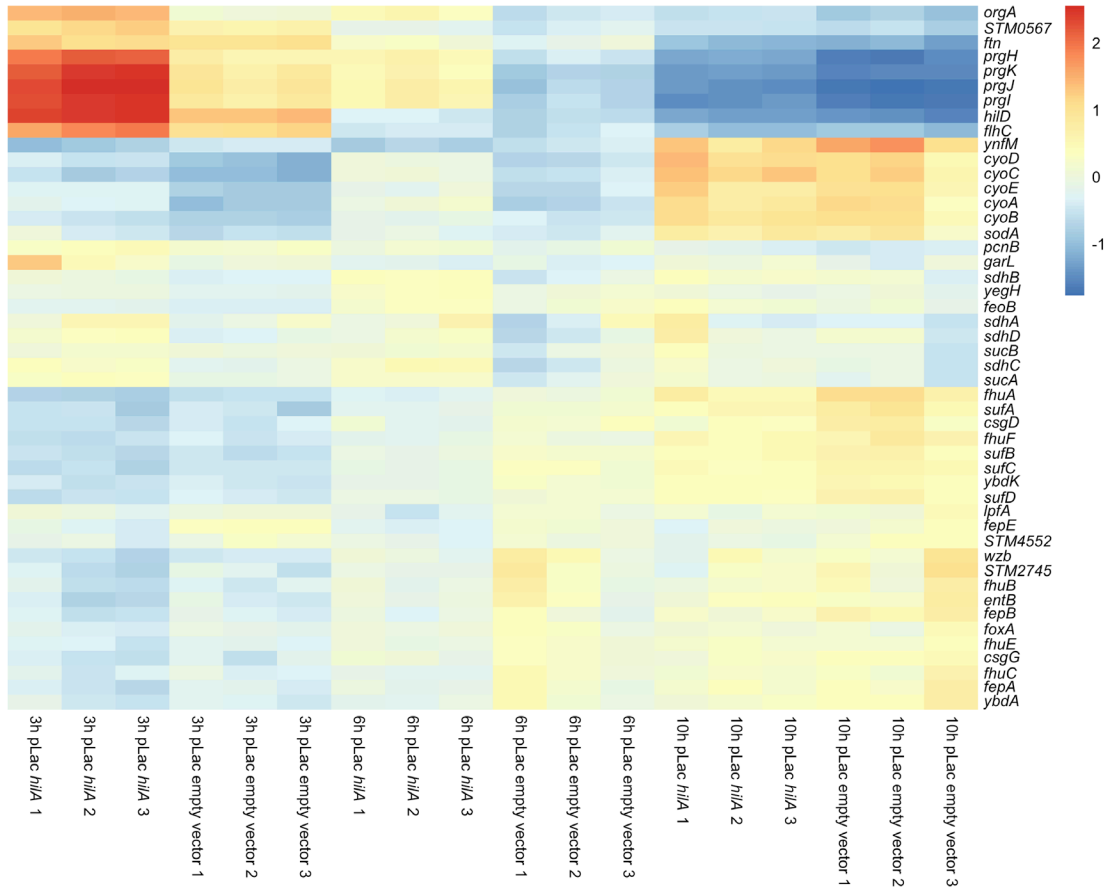
**Figure 5.24 Heatmap showing the regularized log transformed read counts of each gene under transcriptional regulator *fur*.**
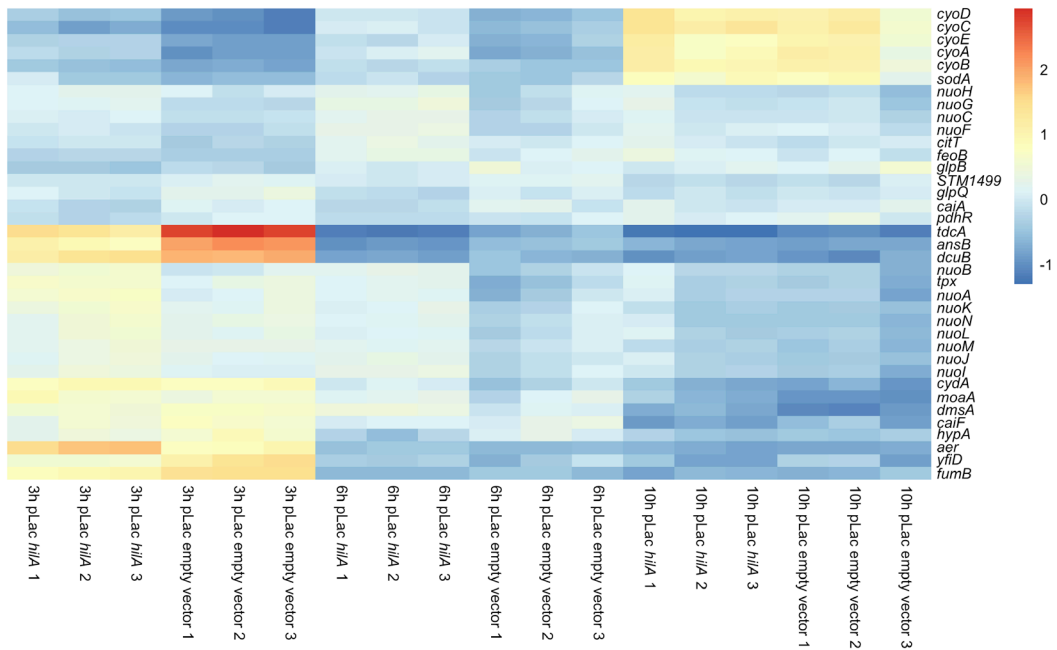


**Figure 5.25 Heatmap showing the regularized log transformed read counts of each gene under transcriptional regulator *fnr*.**

## Pathway analysis revealed enrichment of genes involved in two KEGG pathways

The KEGG database contains well-curated sets of pathways that are important to cellular metabolism[172,173]. We used gene set enrichment or pathway analysis (GSEA) to provide insights to changes in the metabolome due to *hilA* overexpression. To map changes due to *hilA* overexpression in LB-L, we used the gage library to perform GSEA[174]. The pathways were then visualized with Pathview[175].

As expected, there was an enrichment in gene expression changes in the bacterial secretion system (Figure 5.26). The other genes shown in Figure 5.26 had small fold changes that may not be significant. There changes in the protein composition of the membrane may be necessary to accommodate the large SPI-1 needle complex.
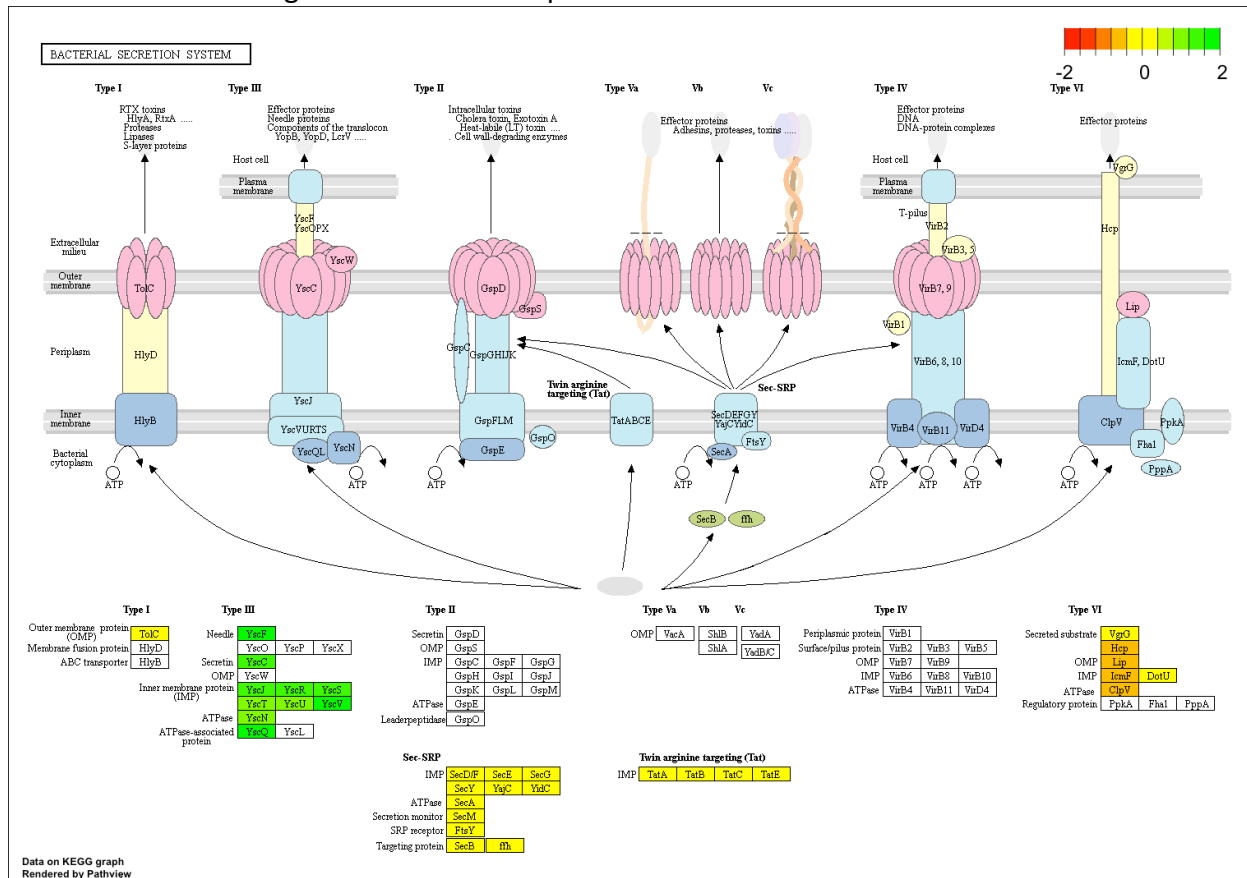


**Figure 5.26 GSEA analysis showed an enrichment of genes in the bacterial secretion systems. The numbers represent the fold change of *hilA* overexpression over the empty vector. Note that the unified nomenclature for T3SS is used by KEGG.**

The other pathway with an enrichment of genes was the oxidative phosphorylation pathway (Figure 5.27). The oxidative phosphorylation pathway is key to energizing the bacteria. Given that heterologous secretion requires the expression of numerous proteins, there is likely a depletion in ATP levels. Moreover, the SPI-1 injectisome is powered partially by ATP[39]. Of note, many of these proteins are also membrane bound. However, we did not identify any enrichment in differentially expressed genes at the different subcellular locations using the PSORT database[176].

**Figure 5.27 GSEA analysis also showed an enrichment of genes in the *S. enterica* oxidative phosphorylation pathway. There was an increased in expression of genes in the oxidative phosphorylation pathway. The numbers represent the fold change of *hilA* overexpression over the empty vector.**

## Clustering of genes by Gene Ontology provides information on key cellular functions

Gene Ontology (GO) is another good database that can be used to look for enrichment analysis in gene set using similar GSEA. Gene Ontology provides a classification scheme for gene functions and is subdivided into biological process, cellular component or molecular function[177,178]. By grouping the differentially expressed genes into GO terms, we might be able to identify specific biological functions being altered when *hilA* is overexpressed.

As with the KEGG pathway analysis, the gage library was used to determine enrichment in each GO term. A total of 2205 GO terms were mapped out by this process of which 834 GO terms (for upregulated genes) and 318 GO terms (for downregulated) were significantly enriched. Given the potential redundancy of GO terms, REVIGO was used to summarize these terms using the semantic similarity measure, SimRel[179]. To visualize the data, the outputs from REVIGO were used to draw individual treemaps for biological process (upregulated genes in Figure 5.28, downregulated genes in Figure 5.29), cellular component (upregulated genes in Figure 5.30, downregulated genes in Figure 5.31), and molecular function (upregulated genes in Figure 5.32, downregulated genes in Figure 5.33)[180,181]. The size of the space occupied by each GO term in the treemap represents the number of genes in the cluster. The complete set of summarized GO terms can be found in Appendix I List of Gene Ontology terms summarized by REVIGO. While there is much to be learnt from this analysis, this will be pursued by a future researcher in the Tullman-Ercek laboratory.

74

## Gene Ontology Treemap of Biological Process for Upregulated Genes



**Figure 5.28 Treemap of biological process for upregulated genes as determined by REVIGO. There were originally 453 GO terms that were significantly enriched.**

## Gene Ontology Treemap of Biological Process for Downregulated Genes



**Figure 5.29 Treemap of biological process for downregulated genes as determined by REVIGO. There were originally 99 GO terms that were significantly enriched.**

## Gene Ontology Treemap of Cellular Component for Upregulated Genes



**Figure 5.30 Treemap of cellular component for upregulated genes as determined by REVIGO. There were originally 65 GO terms that were significantly enriched.**

## Gene Ontology Treemap of Cellular Component for Downregulated Genes
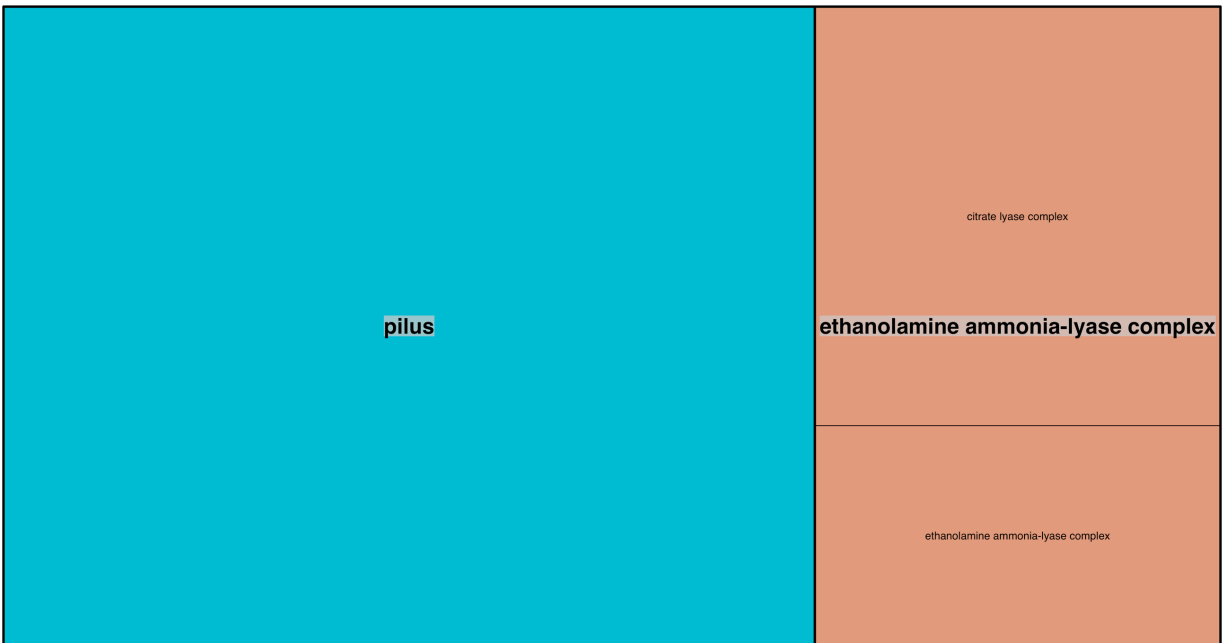


**Figure 5.31 Treemap of cellular component for downregulated genes as determined by REVIGO. There were originally 3 GO terms that were significantly enriched.**

## Gene Ontology Treemap of Molecular Function for Upregulated Genes
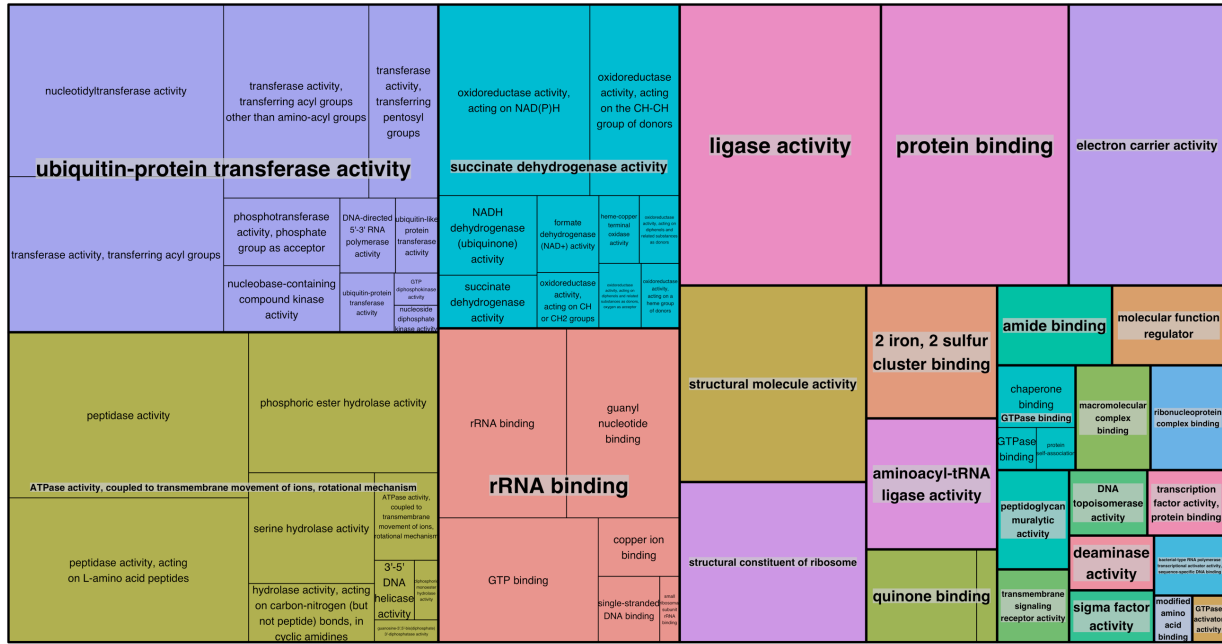


**Figure 5.32 Treemap of molecular function for upregulated genes as determined by REVIGO. There were originally 142 GO terms that were significantly enriched.**

## Gene Ontology Treemap of Molecular Function for Downregulated Genes
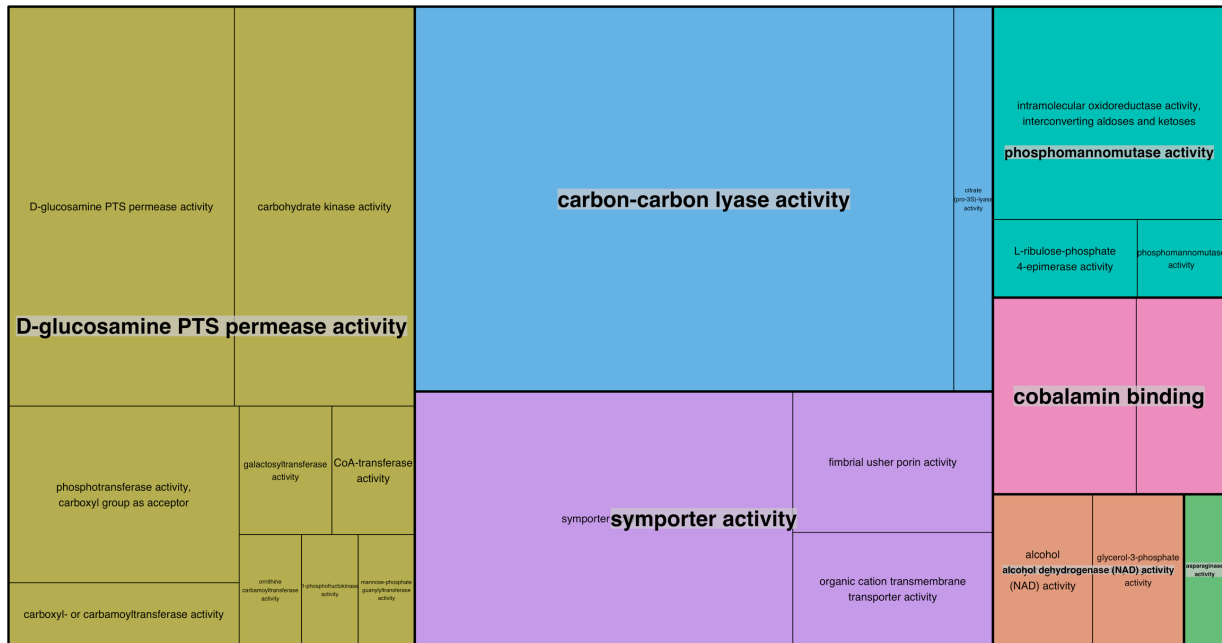


**Figure 5.33 Treemap of molecular function for downregulated genes as determined by REVIGO. There were originally 52 GO terms that were significantly enriched.**

## Altering transcriptional regulators levels may improve secretion titer

The differential gene analysis showed that *hilA* overexpression drives changes in the transcriptome which can be coordinated by transcriptional regulators. We hypothesized that by altering the levels of these global regulators, we would be able to create the optimal cellular state needed for SPI-1 heterologous secretion. Many of these cellular factors are already known to have a strong impact on SPI-1, as discussed in Chapter 1.

Fis is a good candidate gene for which we could modulate expression to impact secretion activation. We reasoned that the overexpression of *fis* could confer an ideal cellular state for SPI-1-mediated heterologous secretion[67]. Another important cellular factor is H-NS, which is a known antagonist to Fis as both proteins compete to bind on similar DNA sequences and have different impacts on gene expression. In addition, SPI-1 expression is subjected to H-NS silencing[182]. H-NS, a nucleoid-associated protein, binds to AT-rich sequences and helps suppress horizontally acquired genes which includes pathogenicity islands[68,183–187]. H-NS silencing is alleviated by overexpressing of a non-functional HN-S (e.g. *hns*[G113D]). This drives the formation of H-NS heterodimers that no longer bind to AT-rich DNA sequences[188,189].

To test the hypothesis that altering the concentration of nucleoid binding proteins can alter heterologous protein, I overexpressed *fis*, *hns*, and *hns*[G113D] in a strain harboring *sptP*[168-543]*::DH:2XFLAG* on the genome. The effect of overexpressing these cellular factors were also studied in concert with *hilA* overexpression. Overexpression of either Fis, H-NS, or H-NS[G113D] resulted in higher total protein expression and secretion of the FLAG-tagged protein of interest (POI), DH (Figure 5.34). Unexpectedly the highest secretion titer was observed when both *hns* and *hilA* were overexpressed, leading us to conclude that there might be a potential artefact from the experimental design.

| pTet | *fis* | *hns* | *hns*[G113D] | *fis* | *hns* | *hns*[G113D] | *fis* | *hns* | *hns*[G113D] | *fis* | *hns* | *hns*[G113D] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| aTc (1µM) | - | - | - | - | - | - | + | + | + | + | + | + |
| pLac hilA | - | - | - | + | + | + | - | - | - | + | + | + |
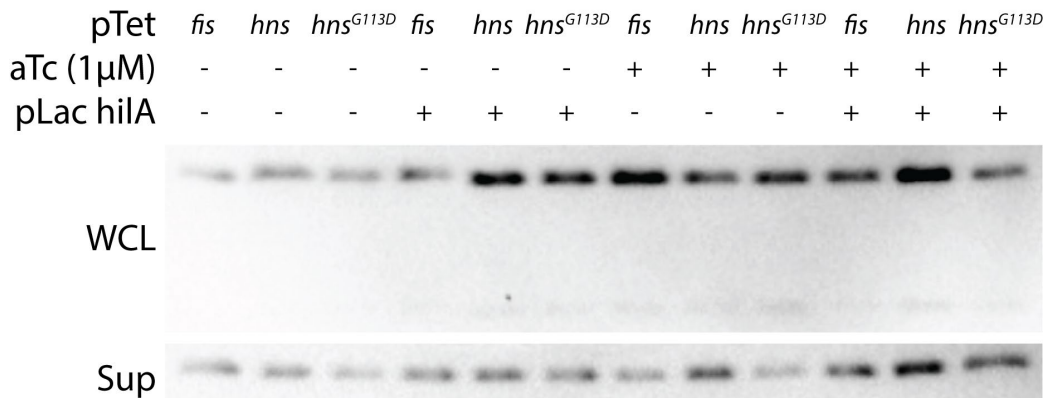


**Figure 5.34 Western blots of whole culture lysate and supernatant against the FLAG epitope.**

The impact of other genes such as *rfaH* and *rpoE* was also carried out in the lab but are not reported here due to the potential artefact from the experimental design.

## Iron can alter expression and secretion of heterologous proteins

Fur was another important transcription regulator of SPI-1 identified from both the studies done here and from the literature[72,73]. When bound of ferrous iron ions, Fur bind to its cognate promoters and upregulates various genes[190]. We hypothesized that changing the availability of iron in the media could be used to alter the activity of Fur. Based on literature, an iron-specific

chelator, 2,2'-bipyridyl, can be used at 200mM to chelate the iron while 100mM ferric chloride can be used to enhance Fur activity[191]. I tested the effect of changing iron levels in the media on heterologous secretion by either chelating the iron or adding more iron. Secretion titers were determined by western blotting against the FLAG-tagged POI, DH.

When *hilA* overexpression was not induced, the addition of $FeCl_3$ slightly increase total protein expression and secretion while 2,2'-bipyridyl reduced both expression and secretion (Figure 5.35). This confirmed our hypothesis that iron can be used to drive SPI-1 mediated heterologous secretion. When *hilA* was overexpressed, there were no observed differences in expression but there was a decrease in secretion when 2,2'-bipyridyl was added to the media.
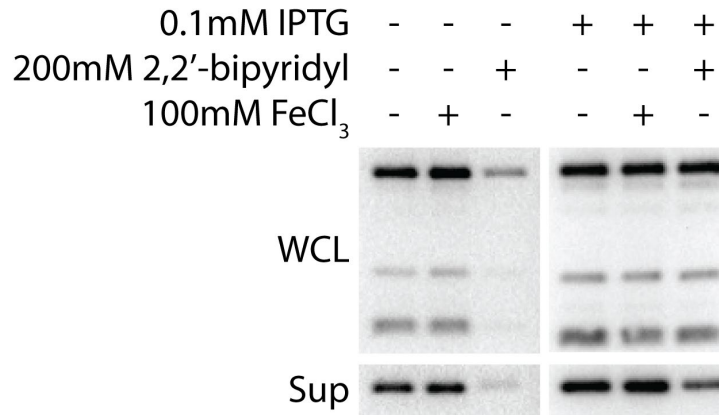


**Figure 5.35 Western blot against FLAG of whole culture and supernatant under different conditions.**

## Discussion

The use of RT-qPCR depends heavily on optimal selection of housekeeping genes as reference. The exploratory data analysis of the RNA sequencing carried out here showed that none of the identified housekeeping genes were suitable to correct for difference between samples. I also presented some initial work in using the transcriptomics data to gain insights into changes in regulation driven by *hilA* overexpression.

Numerous potential targets for engineering were identified by the initial analysis of the RNA sequencing data. These include *fis*, *phoP*, *fur* and *ssrB*. The importance of oxygen levels on SPI-1 regulation as reaffirmed in this study here provides an attractive target for optimizing growth conditions for heterologous secretion. The huge energetic cost of producing the SPI-1 complex was partially alleviated by upregulation of the electron transport chain. Employing fed-batch fermentation and additional carbon sources could also provide more energy for this costly process. Furthermore, the work here presents a starting point for rational metabolic engineering of *S. enterica* for enhanced heterologous secretion. Both potentially important and redundant metabolic pathways had been identified using GSEA. Additional detailed analysis of the data may provide changes in the transcriptome driven by time. This could be key to extending the activation of SPI-1 over a longer period of time.

Altering the packing of the nucleoid showed promise in further improving heterologous secretion. However, the data here may not be accurately depicting the effect of overexpressing the nucleoid bound proteins on heterologous secretion. Firstly, the pTet plasmid used here had

the same backbone as the export plasmid and thus its presence could affect the expression of gene from the genome at the SPI-1 locus as seen in Figure 1.5. Secondly, work from an undergraduate I mentored, Jason Wong, showed that the addition of aTc can alter heterologous secretion. As such, this work will need to be repeated using the induction-free strain described in Chapter 4 and new sets of plasmids.

Unexpectedly, altering the iron levels can still change secretion titers even with the overexpression of *hilA*. Given that Fur drives the autoinduction of *hilD*, any changes to *hilD* levels would not be important when *hilA* was already overexpressed[73]. The similar POI expression showed that the pSicA promoter at least was not affected. The reduction in iron led to lower secretion titers, which could be due to differences in the number of active SPI-1 injectisome. In order to better control the iron levels in the media and confirm these results, the use of a defined media would be needed; this is undertaken by another lab member. In addition, there is an increased need to more easily quantify the number of needle complexes under different conditions in order to correlate transcriptional activity with number of injectisomes built per cell.

## Materials and methods

### Cloning of various constructs

The plasmids with *fis* and *hns* were cloned Gibson assembly[159]. Gene inserts were prepared using PCR with 20 bp overlap flanking the 5' and 3' ends. Inserts were then thermocycled with pTet ColE vector linearized by Phusion PCR using standard procedures as described previously. 5 µl of the Gibson assembly reaction was then used to transform chemically competent *E. coli* DH10B cells by heat shock. The chemically competent cells were subjected to 20 minutes on ice, followed by 60 seconds at 42°C, 2 minutes on ice and then recovery for 1 hour at 37°C with 350 µl of SOC media. 50 µl of the transformed cells were plated on LB agar plates with the chloramphenicol. Quikchange was used to clone pTet *hns*$^{G113D}$. The Quikchange was carried out using the KOD polymerase with overlapping primers covering the mutation site. All genes were sequenced-verified by Quintara Inc. (Boston, MA). Primers used can be found in Table 5.2.

| Primers used in cloning | |
|---|---|
| pTet F GA | CTCGAGTAAGGATCTCCAGG |
| pTet R GA | GACCTTTCTCCTCTTTAAAGATCTTTTGAATTCTTTTCTCTATCACTGATAGGG |
| fis F GA | CTTTAAAGAGGAGAAAGGTCATGTTCGAACAACGCG |
| fis R GA | CCTGGAGATCCTTACTCGAGTTAGTTCATGCCGTATTTTTTTAA |
| hns F GA | CTTTAAAGAGGAGAAAGGTCATGAGCGAAGCACTTAAAATTCTG |
| hns RGA | CCTGGAGATCCTTACTCGAGTATTCCTTGATCAGGAAATCTTCC |
| hns G113D F QC | AAACCTGGACTGGCCAGGATCGTACACCGG |
| hns G113D R QC | CTGACCGGTCCTAGCATGTGGCCGACATTA |

**Table 5.2 Primers used for cloning cellular factors.**

### RNA and cDNA preparation for quantitative PCR

2 OD$_{600}$ of cells were spun down and resuspended in 500 µL of Trizol. The samples were then shaken vigorously after the addition of 100 µL chlorofoam and left at room temperature for 2

minutes. The samples were then centrifuged at 12000 g for 15 minutes at 4°C to phase separate. The resulting aqueous phase was isolated and placed into a new tube. 250 µL of isopropanol was added to the aqueous phase and left at room temperature for 10 minutes. The mixture was then spun again at 12000 g for 10 minutes at 4°C to phase separate. The supernatant was decanted, leaving behind a faint RNA pellet. The pellet was then washed with 500 µL of 70% ethanol and spun at 7500 g got 5 minutes at 4°C. The wash was decanted, and the pellet allowed to dry to remove any residual ethanol. The pellet was then resuspended in 30 µL of RNAse-free water.

Genomic DNA was further removed by treating 3 µg of RNA with RQ1 DNAse at 37°C for 1 hour. 1 µL of RiboLock was added. The reaction was stopped by using the provided Stop solution. The DNAse-treated total RNA was then reverse transcribed with iScript cDNA synthesis kit to make cDNA.

### *Quantitative reverse transcription PCR.*

qPCR was carried out using the SsoFast Evagreen supermix on the BioRad CFX96 system. A melt curve at the end of the cycle was used to determine the specificity of the amplification. Contamination of gDNA was determined using a no RT control. Each primer pair in Table 5.3 was tested for efficiency using 10000 – 10000000 copies of the DW01 genome and found to have similar efficiencies. Fold change was calculated using the delta delta Ct method. The Ct values of glnS and eno were used to calculate the mean Ct for reference gene. 6 housekeeping genes were tested – dnaN, gmk, rpoD, sucA, glnS and eno.

### *Generation of strains used*

Genomic modifications made in this study were done using the Court lab recombineering method[145]. Briefly, strains were transformed with pSIM6 and then grown overnight at 30°C, 225 rpm. The cells were then sub-cultured at a 1:100 dilution and grown for ~2hrs to an OD of 0.4 – 0.8. The lambda red system was then induced at 42°C for 15 minutes and then the cultures were cooled in an ice water bath for 10 minutes. The cells were spun down at 4600 g for 3 minutes at 4°C, washed thrice in ddH$_2$O and finally resuspended in 200 µL of ddH$_2$O. 200 ng of DNA products were electroporated into 50 µL of cells as needed. The DNA products used in round one of recombineering were generated by Phusion PCR using the TUC01 genomic DNA as a template and the primers listed in Table 5.3. For round 2, plasmids containing either GFPmut2 or mCherry were used as a template and primers listed in Table 5.3 were used. The PCR reaction was then cleaned up with the Wizard SV Gel and PCR Clean-up kit (Promega Catalog no. A9282) prior to electroporation.

For round one of recombineering, the cells were recovered in 350 µL of SOC at 30°C for an hour and then plated on LB agar plates with 10 µg/L of chloramphenicol. For round two, the cells were recovered in 10 mL of LB Lennox at 37°C for 4 hours. Serial dilutions of the recovery are then plated on 6% sucrose plates. Patch plating on LB agar plates with 10 µg/mL of chloramphenicol and 30 µg/mL of carbenicillin to determine successful recombination and loss of pSIM6 plasmid respectively. Colony PCR was then carried out using GoTaq (Promega Catalog no. M3008) to isolate the gene of interest to be Sanger sequence verified by Quintara BioSciences (Boston, MA) using the primers listed in Table 5.3. The fluorescence protein specific primers were used for the actual sequencing reactions.

| qPCR Primers | |
|---|---|
| dnaN F | GATTTCCCGAATCTTGACGA |
| dnaN R | CATCGAAAACTGGGTCGATT |
| eno F | ATCGAGAAAGGCATCGCTAA |
| eno R | CGTCTTTCGCCATCTTGATT |
| glnS F | TGCTGGACAACATCACCATT |
| glnS R | CAGCTTACGCTTGGACATCA |
| gmk F | CGCGTTAAGCGACTTGAAA |
| gmk R | TCAGTCTGCCAACAATTTGC |
| rpoD F | CGTACCCAGGAACGTCTGAT |
| rpoD R | CAGGTTTCGCTGGTTTCATT |
| sucA F | GTACCGGCGACGTGAAGTAT |
| sucA R | CACAATTTCCAGATGCGATG |
| hilD F | AGGAGCGCGTTTACAACATT |
| hilD R | AAGTTTCCGTTTGAGCGTTG |
| hilA F | ATGCGATTAAGGCGACAGAG |
| hilA R | GCAAACTCCCGACGATGTAT |
| invF F | CCGATAAATGGGTTTTGCTG |
| invF R | GCCGGAGAAGGCGTAATAAT |
| invA F | TGTCACCGTGGTCCAGTTTA |
| invA R | CTGTTTACCGGGCATACCAT |
| prgH F | TTATCCGCAGCTGGCTTATT |
| prgH R | TTCTTGCTCATCGTGTTTCG |
| prgK F | GGATTCGCTGGTATCGTCTC |
| prgK R | GCCCTCCATCGTCTGTAATG |
| sicA F | AATGCGTAAGGCAGCAAAAG |
| sicA R | TAGCGCCTCCAGATAGACCA |
| sipA F | GCGTAACCAGCAAGAGCATT |
| sipA R | TTCACAATCTCTGCCGTCTG |
| Recombineering Round 1 Primers | |
| hilD:catGsacB F | AACTACGCCATCGACATTCATAAAAATGGCGAACCATTAATGTGACGGAAGATCACTTCG |
| hilD:catGsacB R | ATAAAAATCTTTACTTAAGTGACAGATACAAAAAATGATCAAAGGGAAAACTGTCCATAT |
| invE:catGsacB F | AATGGCAGAACAGCGTCGTACTATTGAAAAGCTGTCTTAATGTGACGGAAGATCACTTCG |
| invE:catGsacB R | GAGAAAGCAGCACTATAGGTATCCTGTTAATATTAAAATCAAAGGGAAAACTGTCCATAT |
| spaS:catGsacB F | AGACGTTATTCAGCCACAAGAAACGAGGTACGGCATTGATGTGACGGAAGATCACTTCG |
| spaS:catGsacB R | AACGCCCAATGAATACATCGCTACTGCCTTACGCGGCATCAAAGGGAAAACTGTCCATAT |

| | |
|---|---|
| prgH:catGsacB F | AATGAGCCCAGGCCATTGGTATTTCCCAAGCCCACTTTAATGTGACGGAAGATCACTTCG |
| prgH:catGsacB R | AAGGTGTTGCCATAATGACTTCCTTATTTACGTTAAAATCAAAGGGAAAACTGTCCATAT |
| sipC:catGsacB F | GTTCGCCATCAGGAGCGCGATTAAATCACACCCATGATGGCGTATAGATGACCTTTCAGATGTGACGGAAGATCACTTCG |
| sipC:catGsacB R | AGCATTAACCAGTCGAAAGCATCCGCACTCGCTGCTATCGCAGGCAATATTCGCGCTTAAATCAAAGGGAAAACTGTCCATAT |
| Recombineering Round 2 Primers | |
| hilD:GFPmut2 F | AACTACGCCATCGACATTCATAAAAATGGCGAACCATTAAATTAAAGAGGAGAAAGGTCATGAG |
| hilD:GFPmut2 R | TTAATAAAAATCTTTACTTAAGTGACAGATACAAAAAATGTTATTTGTATAGTTCATCCATGCCATG |
| hilD:mCherry F | AACTACGCCATCGACATTCATAAAAATGGCGAACCATTAAATTAAAGAGGAGAAAGGTCATGGTTTCCAAGGGCG |
| hilD:mCherry R | TTAATAAAAATCTTTACTTAAGTGACAGATACAAAAAATGTTATTTGTACAGCTCATCCATGC |
| invE:GFPmut2 F | AATGGCAGAACAGCGTCGTACTATTGAAAAGCTGTCTTAAATTAAAGAGGAGAAAGGTCATGAG |
| invE:GFPmut2 R | GTAGAGAAAGCAGCACTATAGGTATCCTGTTAATATTAAATTATTTGTATAGTTCATCCATGCCATG |
| spaS:GFPmut2 F | AGACGTTATTCAGCCACAAGAAAACGAGGTACGGCATTGAATTAAAGAGGAGAAAGGTCATGAG |
| spaS:GFPmut2 R | AAAAACGCCCAATGAATACATCGCTACTGCCTTACGCGGCTTATTTGTATAGTTCATCCATGCCATG |
| prgH:GFPmut2 F | AATGAGCCCAGGCCATTGGTATTTCCCAAGCCCACTTTAAATTAAAGAGGAGAAAGGTCATGAG |
| prgH:GFPmut2 R | ACCAAGGTGTTGCCATAATGACTTCCTTATTTACGTTAAATTATTTGTATAGTTCATCCATGCCATG |
| prgH:mCherry F | AATGAGCCCAGGCCATTGGTATTTCCCAAGCCCACTTTAAATTAAAGAGGAGAAAGGTCATGGTTTCCAAGGGCG |
| prgH:mCherry R | ACCAAGGTGTTGCCATAATGACTTCCTTATTTACGTTAAATTATTTGTACAGCTCATCCATGC |
| sipC:GFPmut2 F | ATCCGCACTCGCTGCTATCGCAGGCAATATTCGCGCTTAAATTAAAGAGGAGAAAGGTCATGAG |
| sipC:GFPmut2 R | TTAAATCACACCCATGATGGCGTATAGATGACCTTTCAGATTATTTGTATAGTTCATCCATGCCATG |
| sipC:mCherry F | ATCCGCACTCGCTGCTATCGCAGGCAATATTCGCGCTTAAATTAAAGAGGAGAAAGGTCATGGTTTCCAAGGGCG |
| sipC:mCherry R | TTAAATCACACCCATGATGGCGTATAGATGACCTTTCAGATTATTTGTACAGCTCATCCATGC |
| Primers to amplify PCR product for Sanger Sequencing | |
| hilD seq F | AGCACGTCCTACTTCATTCAA |

83

| | |
|---|---|
| hilD seq R | AGCGTGTTAATGCGCAGTCT |
| invE seq F | TGATCATCACCATTAGTACCAGAA |
| invE seq R | TGCCTACAAGCATGAAATGG |
| spaS seq F | CGACTGCGTTTATCTGATGC |
| spaS seq R | GTCTTCTGGTTTGGCTGGAA |
| prgH seq F | TCCAGATAGCCTGACCAAGG |
| prgH seq R | GGGCGGAAGGTTATATCAAA |
| sipC seq F | AATATCCCCAGTTCGCCATC |
| sipC seq R | GTCGAAAGCATCCGCACT |
| Sanger Sequencing Primers | |
| GFP seq F | AAAGGAGAAGAACTTTTCACTGGA |
| GFP seq R | AGTTCATCCATGCCATGTGTA |
| mCherry seq F | AAGGGCGAGGAGGATAACAT |
| mCherry seq R | TTATTTGTACAGCTCATCCATGC |

**Table 5.3 Primers used for recombineering**

## *Strains, media, growth and harvest of bacteria*

*Salmonella enterica* strains were grown overnight in LB Lennox (Dot Scientific Catalog no. DSL24066-500) with the appropriate antibiotics as per Metcalf 2014 at 37°C and 225 rpm in 24-well blocks. The overnights were sub-cultured to an $OD_{600}$ of 0.05 in 24-well blocks for all experiments (Axygen Catalog no. PDW10ML24C). Cells were electrotransformed with the required plasmids.

For secretion assay, cultures were grown for 8 hours at 37°C and 225 rpm. The overexpression of *hilA* was induced by 0.1 mM IPTG (Dot Scientific Catalog no. DSI56000-5) at the point of subculture. 1 µM aTc was added at 3hours after subculture to induce the different cellular factors. Antibiotics were added as needed. At the point of harvest, 20 µL of the culture was added to 40 µL of 4X Laemilli buffer for the whole culture lysate (WCL) samples. The blocks were then spun at 4000 g for 10 minutes to collect the supernatant (Sup) fractions. 40 µL of the supernatant was added to 16 µL of 4X Laemilli buffer.

LB-IM refers to cultures there were grown in LB Lennox supplemented with NaCl to 15 g/L and grown at 37°C and 120 rpm in 24-well blocks. Cultures grown in NCE were supplemented with 1X Supplement EZ and grown at 37°C and 225 rpm in 24-well blocks.

## *Western blotting*

Samples were run on a 12.5% SDS PAGE gel at 150V for 60 minutes. 10 µL of samples of WCL and supernatant were loaded. The gels were then equilibrated in Towbin buffer with 20% methanol for 15 minutes. The samples were then transferred to a PVDF membrane (Millipore Catalog no. IPVH00010) using the Owl HEP-1 blotter at 0.3A for 40 minutes. After which, the membrane was blocked in 5% milk in TBST for 1 hour. The membrane was then decorated with anti-FLAG M2 antibodies from mice (Sigma Catalog no. F3165-1MG) in 1% milk (1:6666) overnight, followed by three 5-minutes washes in TBST (0.1% Tween-20). The membrane was then decorated with anti-mouse secondary antibody from goat conjugated with HRP (Thermo Fisher Catalog no. 32430) in TBST (1:1000) for 1 hour, followed by three 5-minutes washes in TBST. Blots

were then imaged using SuperSignal West Femto (Thermo Scientific Catalog no.) on the Bio-Rad Chemidoc.

## Transcriptomics analysis

RNA was extracted from the culture of DW01 in the different media and growth condition listed in Table 5.1 of the main text at 3, 6 and 10 hours by resuspending the cell pellet in Trizol (Thermo Fisher catalog no. 15596026). The Direct-zol RNA Miniprep kit (Zymogen catalog no. R2051) was then used to purify the RNA from the Trizol mixture. The isolated RNA was then sent to RTSF Genomics Core for library preparation using the TruSeq Stranded Total RNA Library preparation (Illumina, San Diego) with Ribo-Zero rRNA depletion. The resulting library was then run over two lanes of Illumina HiSeq 4000 SE50 (Illumina, San Diego) with other libraries.

The raw reads were concatenated and then trimmed to remove adaptor sequences using Trimmomatic v0.36 for a minimum quality of 14 in a sliding window of 4bp[138]. Reads less than 36bp were also filtered out. Quantification of the trimmed reads were done using Salmon v0.10.1 with 100 bootstraps with the reference LT2 transcriptome[149]. DESeq2 v1.22.2 was then used to normalize the reads[150]. The normalized counts per million were plotted using the ggplot2 package in R[137]. PCA analysis was carried out using top 500 genes with the greatest variance between the samples. Hierarchal clustering was done using the distance matrix calculated from all the genes. For PCA and hierarchal clustering, the counts were normalized using the variance stabilizing transformation in DESeq2[150]. For heatmaps, the rlog function of DESeq2 was used to log transform the data prior to plotting. Weighted correlation network analysis was carried out according to WGCNA tutorials[169]. GSEA was carried out according to gage workflow for RNA sequencing[174]. GO terms were filtered with a FDR <0.05 and submitted to the REVIGO server[179]. The treemaps were drawn using the treemap package in R[181].

## Flow cytometry

The cells were grown overnight in LB Lennox (Dot Scientific Catalog no. DSL24066-500) with the appropriate antibiotics supplemented with 0.4% glucose. The overnight culture was diluted to an $OD_{600}$ of 0.05 for subculturing and induced according as above. Samples were taken every hour and diluted to an OD of approximately 0.03 in PBS with 2 mg/mL of kanamycin sulfate. The samples were kept overnight at 4°C and protected from light before running it on the Attune NxT Flow Cytometer (BY) using the autosampler. Data was collected for at least twenty thousand cells and processed with FlowJo v10.5.3. The resulting data is visualized with R and the ggplot2 package[137].

## Chapter 6 Developing Tools to Better Study Secretion

*Salmonella enterica* is a well-studied model pathogen with many tools developed for genetics, molecular biology and biochemistry. As such, numerous genetic screens exist for studying the impact of SPI-1 on pathogenicity. However, these available genetic screens are not designed to quantitative measure SPI-1-mediated secretion and focus on disabling SPI-1-directed pathogenicity. Thus, there remains a need to develop tools for better quantification of heterologous secretion and high throughput screening to isolate better secretors.

Developing a high throughput assay will allow us to tap into the power of direct evolution. The current gold-standard biochemical assay for assessing secretion titers in a robust and quantitative fashion is western blotting. Although an extremely sensitive and powerful biochemical technique, western blotting is a laborious technique that is not suited as high-throughput assay. Enzyme assays had also been developed but these suffered from low dynamic range of detection due to low secretion titer of the enzymes. The optimal assay should be low cost, high throughput and with a wide dynamic range of detection.

With a high throughput assay, we can rapidly screening large strain libraries and this opens up the possibility of using high throughput library construction strategies such as transposon mutagenesis [192,193]. Furthermore, we would expect epistatic effects from combining different strain alterations. To design, test and build the different permutations of beneficial mutations could be time-consuming and laborious without the right assay. Here I am reporting the two approaches for more rapid screening SPI-1 mediated heterologous secretion.

### Growth-based bulk selection

Developing secretion assays can made difficult by a few factors – 1) not all proteins are secretable by T3SS; 2) secretion titers have been too low for most screens prior to this work; 3) only ~40-50% of all expressed protein is secreted. To the first point, recent unpublished data from the Tullman-Ercek lab shows that a greater proportion of proteins from the *E. coli* proteome is secretable than expected. We also found new proteins with interesting properties that are secretable. These include enzymes, transcription factors and anti-microbial peptides. Toward the second point, the Tullman-Ercek lab has steadily increased secreted protein titers over the past decade, which enables the use of enzyme assays as we were able to achieve reasonable protein concentrations secreted in the supernatant. However, the ideal goal would be a selection-based assay.

Designing a successful selection had proven difficult as there is a considerable intracellular pool of proteins trapped in the cell. Using toxic proteins such as barnase did not work due to the third (as-yet unsolved) challenge: the intracellular retention of expressed proteins retained sufficient toxicity.

Here, I am proposing to develop a bulk selection assay. Instead of toxic proteins, semi-toxic proteins can be developed as a potential selection. A few properties are desirable. These proteins should have an impact on growth intracellularly but not extracellularly. The toxicity of these proteins should be tunable and this can be achieved in a dose-dependent manner. Thus *S. enterica* mutants that secrete more proteins will have a better growth advantage and dominate the overall population over time.

*Turbidostat for continuous growth-based selection*

To increase the efficiency of bulk selection, we also need to shift away from batch growth to continuous growth. A continuous growth strategy allows us to exert a constant selection pressure that tracks with reduced toxicity from increased secretion[194–197]. Another advantage is the possibility of the build-up in numerous beneficial mutations in the system without the need for serial passaging. An example of the power of continuous selection is the phage-assisted continuous evolution (PACE) system, which was used to evolved T7 RNAP variants capable of using ATP for initiating transcription[198]. Moreover, many rounds of evolution could be carried out with minimal involvement of researchers.

The Ingolia Lab at UC Berkeley developed a turbidostat that relies of scattering of near-infrared light (IR) to monitor cell density[195]. When illuminated with near-infrared lights, cells would scatter the light which can be detected by a photodetector places at $90°$ from the light source. These systems cost less than \$500 to build and were useful for initial trials. Thus, I adapted this turbidostat system for SPI-1 heterologous secretion, in which new media is flowed in to maintain a desired optical density. This will allow me to maintain a constant selection pressure and accelerate adaptative lab evolution.

I first established the correlation of the IR values to $OD_{600}$ for *S. enterica* (Figure 6.1a). This allows us to set the desired optical density by IR measurements. The best correlation between IR to $OD_{600}$ can be seen from an $OD_{600}$ from 0.05 to 0.9 (Figure 6.1b), setting the boundaries at which cell densities can be monitored and maintained.

*S. enterica* can be grown in the turbidostat and its growth monitored online by IR measurements (Figure 6.1c). The IR measurements peaked at 5 hours which was expected based on the dynamic range observed in Figure 6.1a. Changes in culture volume, up to 5%, were observed after 20 hours. The relatively dry shaker unit used to house the turbidostat could have driven these volumetric changes. Bubbling the air that was sparged into the system first through water increase its humidity was not sufficient to prevent the decrease in culture volume. This could explain the significant drop in IR measurements after 8 hours.

I was able to reliably maintain a constant $OD_{600}$ for at least 24 hours below $OD_{600}$ of 0.8 when running the turbidostat in continuous flow mode. Considerable changes in volume were only observed at 48 hours (5-10% of the culture volume). The reduced loss of media could be a result of constant media addition during the run. Despite the evaporation observed, the $OD_{600}$ remained at 0.8 at 48 hours. However, changes to culture volume would lead to changes in oxygen transfer and thus might not be ideal for our oxygen-sensitive system.

Due to the above considerations, I grew *S. enterica* in the turbidostat over 24 hours at two different constant $OD_{600}$: 0.5 and 0.8. Expression of our model protein, DH, was observed in the cell pellet of cultures were grown in the turbidostat maintained at the $OD_{600}$ of both 0.5 and 0.8. The secretion of DH was only detected in the culture grown in the turbidostat maintained at the $OD_{600}$ of 0.8 (Figure 6.1d). This data shows that pSicA on the plasmid is active but a lack of functional injectisome when grown at the constant $OD_{600}$ of 0.5. I hypothesized that this could be due to lack of expression of SPI-1 from the genome or incorrect cellular state to build the complex SPI-1 injectisome. The previously described time dependence of the activation of T3SS in shaking cultures shown in Chapter 4 and Chapter 5 may also depend on growth phase/rate. Overall, the turbidostat is a robust growth system for SPI-1 mediated heterologous secretion in *S. enterica*.
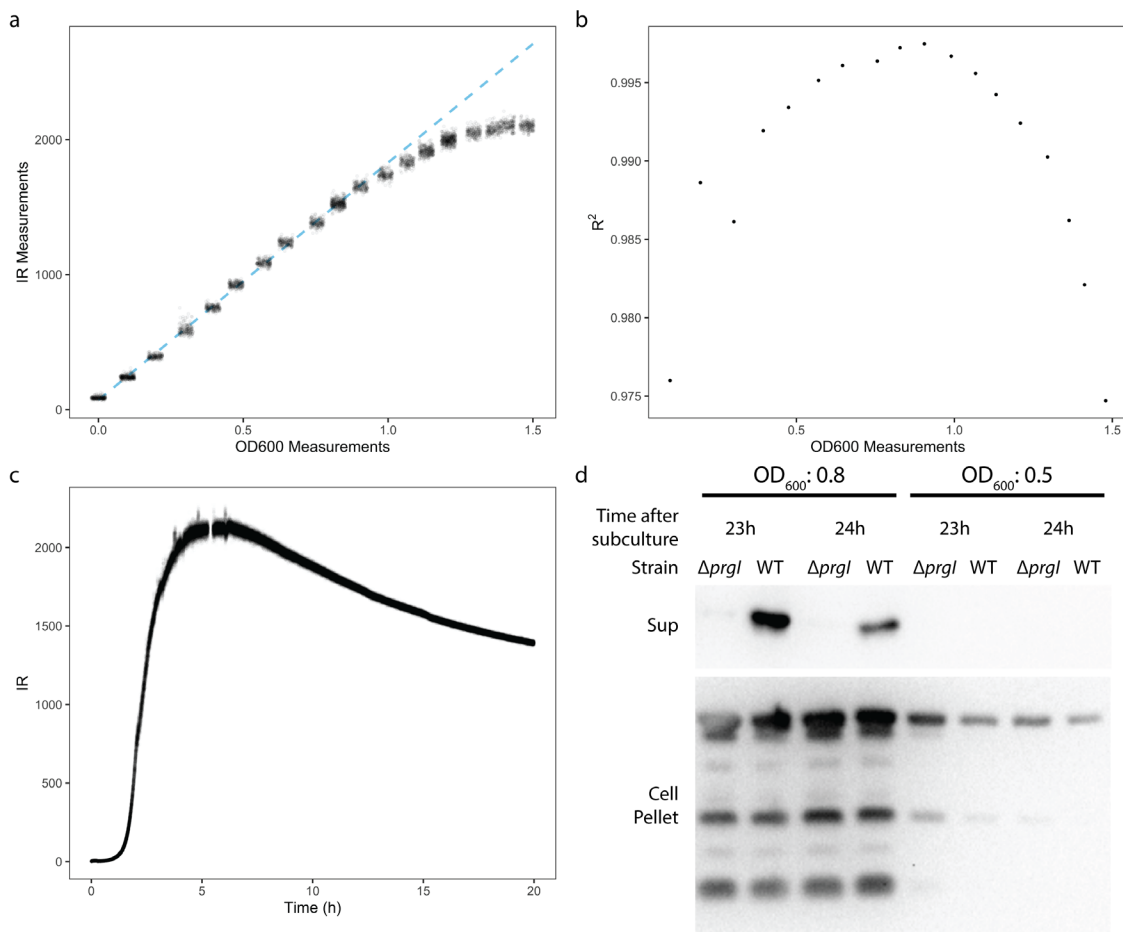
**Figure 6.1 The correlation between OD$_{600}$ and IR measurements is shown in (a) with the R² value of the different linear regression showed in (b). Correlation between OD$_{600}$ and IR decreased at OD$_{600}$ > 0.9. (c) Growth in turbidostat was monitored by IR measurements over time. (d) The cultures were grown the turbidostat at a fixed IR measurement. Western blots showed that secretion was only detected at OD$_{600}$ 0.8 but not 0.5 while expression could be detected in the cell pellet at both OD$_{600}$.**

## Heterologous secretion of semi-oxic genes

There were two proteins of interest available in the lab that could be adapted for a growth-based selection – Magainin-1 and T4 lysozyme. Magainin-1 is an antimicrobial peptide derived from frogs. When expressed as a fusion to a biomaterial-forming protein, it conferred anti-microbial properties to the resulting hydrogels. The expression of Magainin-1 is detrimental to cell growth but this is alleviated when the protein is secreted[42]. I tested the differences in growth of wild type and Δ*prgI* expressing 4 different antimicrobial peptides, Amp1, Amp2, Amp3 and Res-Cn-Mag1. Knocking out *prgI* prevents the formation of the SPI-1 needle complex and thus the expressed protein is retained intracellularly. Surprisingly, I did not observe the differences in growth between the wild type and Δ*prgI* strains that were described in prior reports (Figure 6.2).

Instead, the maximum separation observed in OD$_{600}$ between wild-type and Δ*prgI* was at 6 hours (Figure 6.2). As these properties are no longer ideal for a growth-based bulk selection, additional work would be needed to tune this system into a viable assay.
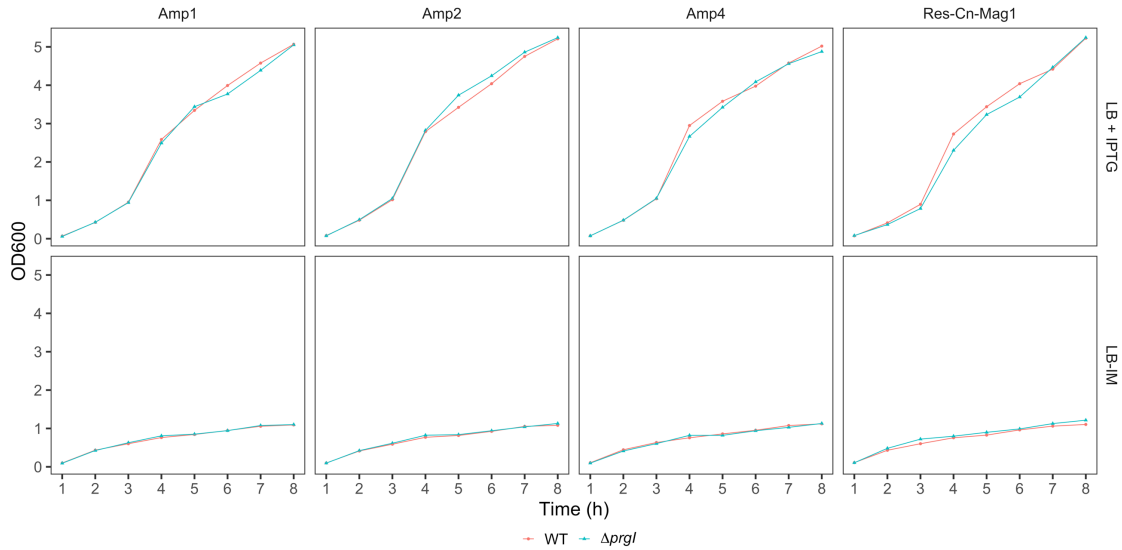
**Figure 6.2 Growth curves of the different antimicrobial peptides in WT and Δ*prgI* expressed off the pSicA export plasmid. The cytotoxicity of the different antimicrobial peptides was much lesser than previously reported.**

In light of the results with magainin-1, I searched for alternate semi-toxic proteins and settled on T4 lysozyme, which is a phage enzyme that is expressed by the T4 phage to lyse its host cell for viral production. T4 lysozyme mutants with different folding kinetics were well-characterized[199]. Prior lab members demonstrated (unpublished data) that the expression and secretion of the different mutants differed with only a single amino acid difference. The different mutants also resulted in different levels of lysis and can have different observed final $OD_{600}$. The cysteine-free T4 lysozyme mutant (T4L*) from Matsumura and Matthews was chosen for initial screening as it was one of the higher-secreting mutants[199,200]. This variant was also known to have good recombinant protein expression in *E. coli*. Initial trials on using T4L* for selection were carried out in an adaptive lab evolution experiment where cultures were passaged serially in blocks to create bottlenecks in the population. A single day passage resulted in vast improvements in $OD_{600}$ (Figure 6.3a).
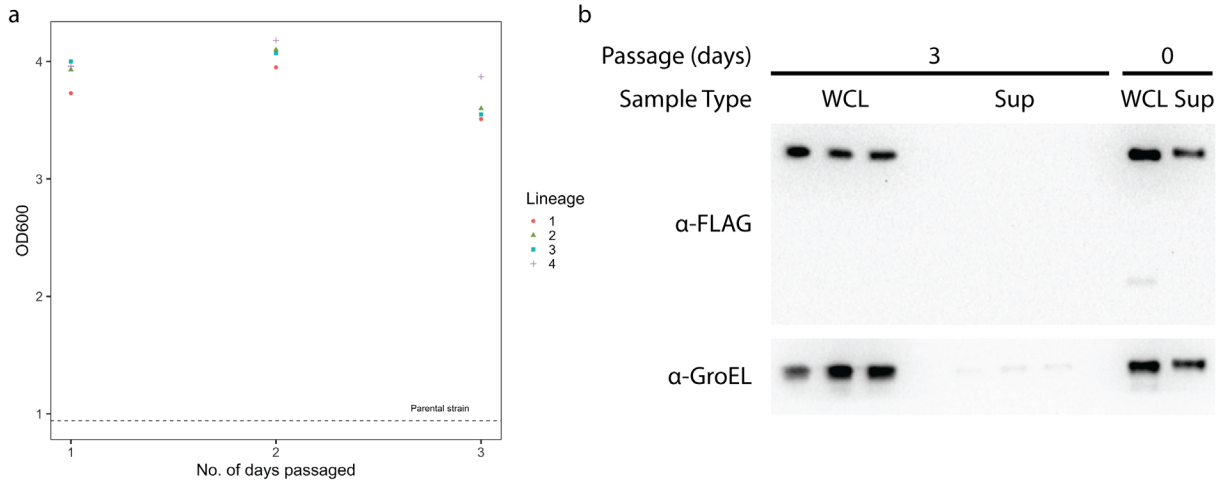


**Figure 6.3 The $OD_{600}$ in the 4 different cultures increase after a single passage (a). (b) Only 3 lineages of the evolved strains are shown here. GroEL blots showed that a significant decrease in cellular lysis. This could have resulted in the lower level of T4 lysozyme detected in the supernatant.**

89

There is an overall decrease in lysis as measured by western blotting for the presence of cytoplasmic GroEL in the supernatant of the evolved strains (Figure 6.3b). The expression levels of T4L* also decreased in the evolved strains (Figure 6.3b). The increases in $OD_{600}$ could be due to inactivating mutations arising in T4L*, *S. enterica* becoming resistant to lysozyme-mediated lysis, or T4L* decreasing in expression level. Moreover, adaptative laboratory evolution of *S. enterica* may lead to the inactivation of SPI-1, which was also observed during the adaptative laboratory evolution in Chapter 3.

One of the difficulties with growth-based selection is maintaining consistent expression of the POI under the pSicA promoter over numerous generations. To ensure the constant expression of the POI under pSicA, I developed a genetic construct by placing an antibiotics resistance cassette, *tetA*, in the same operon as the POI, placing them both under the control of the pSicA promoter. The dual operon was able to express sufficient copies of TetA to impart tetracycline resistance. Although expression from pSicA was only activated after 3 hours, adding tetracycline into the media at 0, 1 and 2 hours did not halt growth (Table 6.1). The resistance observed when tetracycline is added at 0-2 hours was not ideal and could be due to residual expression from the overnight or leaky expression from the plasmid. This can potentially remove the ability to link tetracycline resistance to *tetA* expression driven by pSicA and thus cannot be used to maintain a constant pSicA expression of the POI.

| Time at which tetracycline is added | $OD_{600}$ at time of addition | Final $OD_{600}$ |
|---|---|---|
| pSicA DH | | |
| 0 | 0.069 | 0.108 |
| 1 | 0.122 | 0.24 |
| 2 | 0.447 | 0.744 |
| 3 | 1.196 | 1.925 |
| 4 | 1.8 | 2.25 |
| 5 | 2.3 | 2.65 |
| 6 | 2.79 | 2.88 |
| No tetracycline added | | 3.53 |
| pSicA DH-*tetA* | | |
| 0 | 0.064 | 2.85 |
| 1 | 0.085 | 2.76 |
| 2 | 0.244 | 2.85 |
| 3 | 0.888 | 2.99 |
| 4 | 1.61 | 2.83 |
| 5 | 1.955 | 2.72 |
| 6 | 2.225 | 2.89 |
| No tetracycline added | | 3.02 |

**Table 6.1 OD of the different culture when tetracyline was added and at the point of harvest.**

Unexpectedly, the expression and the secretion of the protein of interest were severely impaired with the addition of the *tetA* gene (Figure 6.4). The dual operon plasmid bearing strain

had a lower overnight $OD_{600}$ as well as a lower $OD_{600}$ upon induction (Table 6.1). This could be due to the metabolic cost incurred when expressing another membrane protein[19].

| Time at which tetracycline added (h) | 4 | 5 | 6 | - | 0 | 1 | 2 | 3 | 4 | 5 | 6 | - |
|---|---|---|---|---|---|---|---|---|---|---|---|---|



Sup

WCL

pSicA DH          pSicA DH-*tetA*
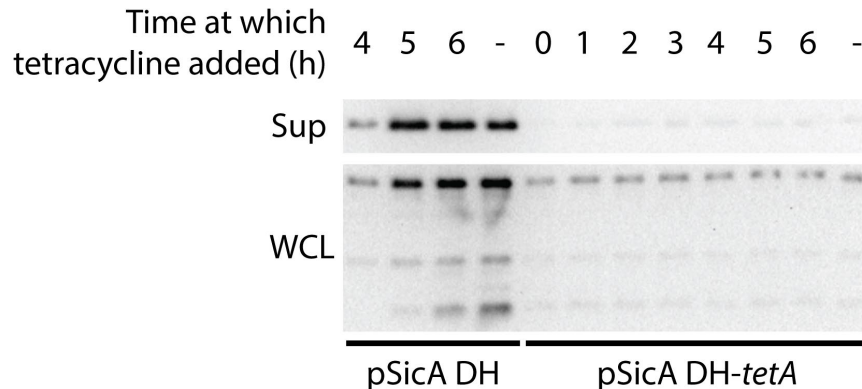
**Figure 6.4 Western blot against the FLAG-tagged DH showed that secretion and expression of DH was severely affected by the presence of *tetA*.**

Expression and secretion of DH was only detectable in pSicA-DH when tetracycline was added at >4h. Surprisingly, the addition of tetracycline at 5 and 6 hours to pSicA-DH resulted in expression and secretion close to the untreated sample despite differences in $OD_{600}$ (Figure 6.4). This further lends support to the hypothesis that most of the protein of interest and T3SS machinery were already expressed by the end of late exponential phase. No further protein expression from 5 to 8 hours was needed for translocation of proteins nor was it necessary.

Keeping the cells within the growth phase space of an $OD_{600}$ between 2.5 and 2.8 is likely to maintain a highly secreting phenotype. The exact $OD_{600}$ would be dependent on the growth condition as different media and $O_2$ levels can shift the growth phase of *S. enterica*. This provided the basis for increasing secretion titer by extending the expression time of the SPI-1 system and prolonging activation of its associated promoters. However, no suitable candidate genes for bulk selection have been identified in this work.

## *Determining secretion titer using fluorescence proteins*

As developing the bulk selection was not successfully, I turned towards developing a high-throughput screen. Using secreted fluorescence protein is ideal as it is easy to assay and does not require costly reagents as compared to enzyme assays. The fluorescence levels of a fluorescent protein are also less sensitive to salts, pH and other factors when compared to enzymes.

We and others have shown that GFP cannot be secreted by the SPI-1 injectisome. However, it is possible to secrete spilt GFP as well as a GFP without a chromophore[132,201]. This provided evidence that the highly stable chromophore was the main barrier to the secretion of GFP. The use of a GFP with its chromophore mutated is no longer fluorescent and thereby ineffectual as a screen. The split GFP system required production of a complementary protein and might not fully associate with its partner in the supernatant. The low dynamic range and high background fluorescence reduced its effectiveness.

Given that the matured chromophore prevents secretion, I chose a slow maturing fluorescent protein instead to secrete. If the secretion rate of the protein was much faster than the maturation of the chromophore, then the fluorescent protein might be secreted. This strategy has precedence: the folding kinetics of the secreted substrate is important for other secretion

systems[202–204]. For this work, mRuby3 was identified as a good candidate as it takes 168 minutes to mature[161].
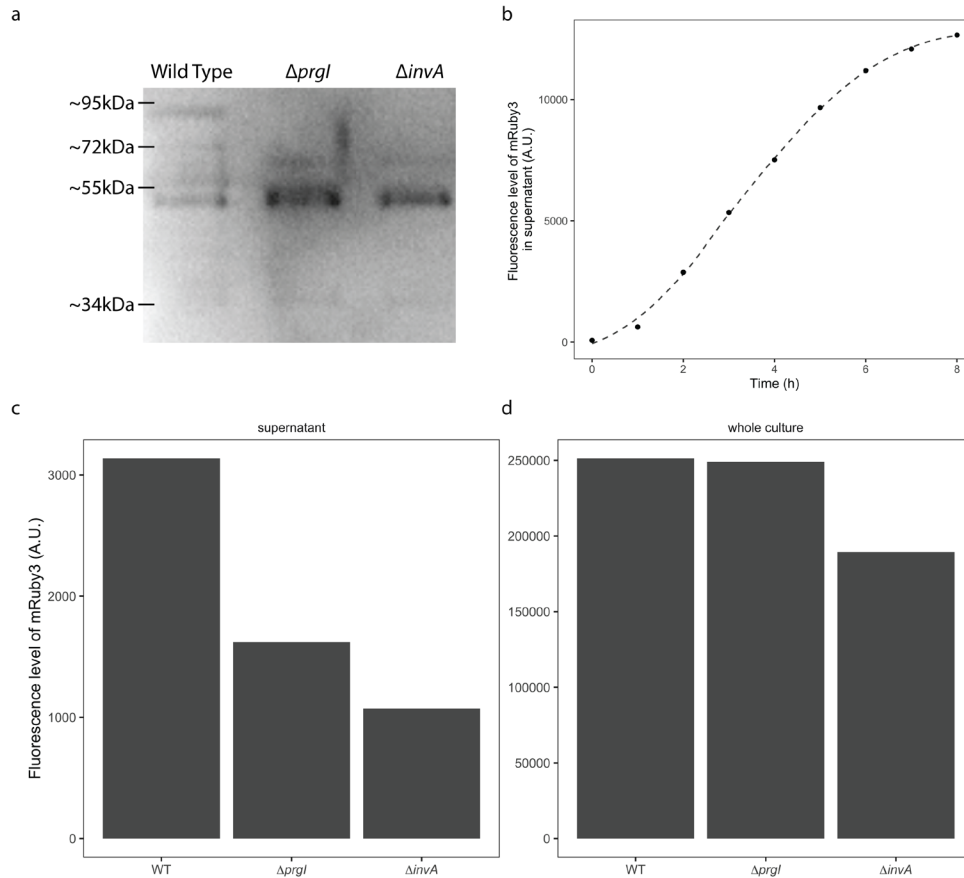


**Figure 6.5 a) Western blot against the FLAG tagged mRuby3 was carried out on the supernatant. b) mRuby3 fluorescence was monitored over time. Fluorescence in the supernatant (c) and whole culture (d) were measured by the plate reader after 16 hours from harvest to allow full maturation of mRuby3. A small increase in $OD_{600}$ was observed from harvest to measurement on the plate reader after an overnight incubation (<0.2).**

To test this hypothesis, I cloned an export plasmid harboring a FLAG-tagged mRuby3.I was able to detect the presence of mRuby3 in the supernatant by western blotting (Figure 6.5 a). However, mRuby3 was also present in the supernatant in the $\Delta prgI$ and $\Delta invA$ strains but to a lesser extent. Both $\Delta prgI$ and $\Delta invA$ strains served as negative controls as both strains are not known to have a functional T3SS, thus we did not expect to detect the presence of mRuby3 in the supernatant. The secretion of mRuby3 was detected in all three strains with a higher molecule weight band is seen only in the wild-type strain (Figure 6.5a).

To attain the maximum signal from mRuby3, we needed to determine the maturation kinetics of mRuby3 in the supernatant. We monitored the changes in fluorescence over 8 hours using a plate reader. The highest fluorescence detected took 8 hours to achieve (Figure 6.5b). Going forward, we left the supernatant and whole culture sample in the dark overnight to allow for complete maturation of mRuby3.

Using a plate reader, we can rapidly assay the fluorescence level of mRuby3 in the whole culture and supernatant as a measurement of total mRuby3 expression and mRuby3 secretion respectively. This was carried out using the three strains used for western blotting. As

expected,there was fluorescence detected in all three strains (Figure 6.5c) with the highest fluorescence detected in wild-type, which was almost double that of Δ*invA* (Figure 6.5c).

However, in certain replicates (not shown here), the whole culture fluorescence in Δ*prgI* was double that of wild type and this resulted in greater fluorescence in the supernatant. The ratio of fluorescence in the supernatant over the whole culture remained the same even in those experiments. Thus, Δ*invA* was be a more suitable negative control than Δ*prgI* for this secretion assay. Before adopting this screen for studies of the T3SS, additional studies are needed, including how mRuby3 is detectable in the supernatant of non-secreting strains.

## Conclusion

As a proof of concept, I showed that it was possible to achieve heterologous secretion when growing *S. enterica* in a continuous culture format. Although it was not possible to find a semi-toxic gene that would allow the development of bulk selection for increased heterologous secretion, the foundation for future studies has been laid out. Another possible protein choice could be a biomaterial-forming protein as high expression of such proteins often affects growth when retained intracellularly.

## Future prospects of SPI-1 mediated heterologous secretion

The genetic engineering of *E. coli* has opened up its capabilities as a host for heterologous protein production. The production strain of *E. coli* had undergone numerous engineering to shift its metabolism for better protein production. Similar work can be carried out in *S. enterica* to achieve similar impacts in titer. Using transcriptomics data, it may be possible to do this in a more directed manner. Freeing up cellular resources such as removing highly expressed non-essential genes would also help improve growth and titers[205,206]. There are numerous resources available such as the *S.* Typhimurium Biochemical, Genetic and Genomic (BiGG) knowledge base and Online GEne Essentiality (OGEE) database that would be useful to such endeavours[207,208]. I believe that the system biology can help guide both strain and process engineering[209].

A huge barrier to the adoption of *S. enterica* is its pathogenicity. Although the current production strain, DW01, has become attenuated through serial lab passaging, there is still a need to further cripple the production strain's infectivity. This can be achieved by removing other known pathogenicity factors. Removing all the amber stop codons would be another way to further insulate the production strain.

Various approaches to better balance cellular resources for growth, protein expression and secretion will be needed. Using insights from metabolic engineering can direct carbon flux towards heterologous protein expression[209]. As different POI can have drastically different amino acid composition, the energetic cost to the cell can also drastically be altered[210] and this might require engineering of amino acid production pathways.

Lastly, more tools such as different expression plasmids and different purification tags can further its adoption. As the volume of supernatant is greater than that of the cell pellet, current purification techniques can result in greater downstream processing time. In addition, the presence of different salts and ions in the media can interfere with the column matrix, further affecting its performance. New purification tags such as the elastin-like polypeptide might be useful for initial protein enrichment while self-cleavable tags such as the intein system can

remove the need for a protease[211,212]. The various issues associated with plasmid use in this thesis also further the need to shift towards genomic engineering for a stable production strain.

## *Materials and methods*

### *Cloning of pSicA DH-tetA*

The dual operon was cloned using traditional cloning. 10 µl of the export plasmid, pSicA DH-2XFLAG in NEB CutSmart buffer was digested with 1 µl XbaI for 2 hours at 37°C, followed by dephosphorylation with 1 µl of CIP for 1 hour. The resulting mixture was then cleaned up using the Wizard SV PCR Clean-up System. The *tetA* gene was cloned out from the T-SACK *E. coli* strain using the primers tetA XbaI F and tetA XbaI (Table 6.2). After PCR clean up, the product was digested with 1 µl XbaI for 2 hours at 37°C and then heat inactivated at 65°C for 20 mins. 50ng of each part was combined in a 10 µl ligation reaction with 0.5 µl T4 DNA ligase for 2 hours at 37°C and then heat inactivated at 65°C for 10 mins. 5 µl of the ligation reaction was then used to transform chemically competent *E. coli* DH10B cells by heat shock.

pSicA mRuby3 was cloned using the Golden Gate type II endonuclease strategy. Gene inserts were prepared using PCR with the BsaI restriction site flanking the 5' and 3' ends using primers found in Table 6.2. Inserts were then thermocycled with Golden Gate-compatible vectors, T4 DNA ligase (NEB), T4 DNA ligase buffer and BsaI (NEB) using standard procedures as described previously. 5 µl of the Golden Gate reaction was then used to transform chemically competent *E. coli* DH10B cells by heat shock.

The chemically competent cells were subjected to 20 minutes on ice, followed by 60 seconds at 42°C, 2 minutes on ice and then recovery for 1 hour at 37°C with 350 µl of SOC media. 50 µl of the transformed cells were plated on LB agar plates with the respective antibiotics. All genes were sequenced-verified by Quintara Inc. (Boston, MA).

| Cloning of pSicA DH-*tetA* | |
|---|---|
| tetA XbaI F | AGTCTAGATTAAAGAGGAGAAAGGTCATGAATAGTTCGACAAAGATCGCATTGGT |
| tetA XbaI R | AGTCTAGAAGCACTTGTCTCCTGTTTACTCC |
| Cloning of pSicA mRuby3 | |
| mRuby3 GG F | A GGTCTCAGCTTAGTAAAGGAGAAGAACTGATCAAGGAAAATATGC |
| mRuby3 GG R | AGGTCTCACGCTTTTGTATAGTTCATCCATGCCACCACC |

**Table 6.2 Primers used for cloning.**

### *Strains, media, growth and harvest of bacteria*

*Salmonella enterica* strains were grown overnight in LB Lennox (Dot Scientific Catalog no. DSL24066-500) with the appropriate antibiotics as per Metcalf 2014 at 37°C and 225 rpm in 24-well blocks. The strains used harbored the pLac *hilA* plasmid and the relevant export plasmids. The overnights were sub-cultured to an OD of 0.05 in 24-well blocks for all experiments (Axygen Catalog no. PDW10ML24C).

For secretion assay, secretion was induced by 0.1 mM IPTG (Dot Scientific Catalog no. DSI56000-5) at the point of subculture for 8 hours at 37°C and 225 rpm. Antibiotics were added as needed. At the point of harvest, 20 µL of the culture was added to 40 µL of 4X Laemilli buffer

for the whole culture lysate (WCL) samples. The blocks were then spun at 4000 g for 10 minutes to collect the supernatant (Sup) fractions. 40 μL of the supernatant was added to 16 μL of 4X Laemilli buffer. Cultures were grown in LB Lennox supplemented with NaCl to 15 g/L (LB-IM) and grown at 37$^\circ$C and 120 rpm in 24-well blocks to stimulate native inducing conditions.

## Turbidostat

Calibration of the IR measurements to $OD_{600}$ was carried out by adding a cell concentrate slowly, allowing the IR measurements to stabilize between each addition. The $OD_{600}$ was measured after each addition of cell concentrate. For growth, 2.5 mL of overnight was added to 250 mL of media supplemented with the appropriate antibiotics and 0.1mM IPTG. The cultures in the turbidostat were allowed to outgrow for 3 hours before the pumps were engaged to maintain a fixed IR.

## Adaptive lab evolution of T4L*

4 separate lineages of DW01 pLac *hilA* pSicA T4 lysozyme* were passaged every 24 hours for 3 days. Mutant 9 of a previously generated T4 lysozyme library in the lab was used. For each subculture, the overnight culture was diluted 1:100. The strains were grown in LB-L in 24-well blocks at 37$^\circ$C and 225 rpm with appropriate antibiotics.

## Western blotting

Samples were run on a 12.5% SDS PAGE gel at 150 V for 60 minutes. 2 μL and 4 μL of samples of WCL and supernatant were loaded respectively. The gels were then equilibrated in Towbin buffer with 20% methanol for 15 minutes. The samples were then transferred to a PVDF membrane (Millipore Catalog no. IPVH00010) using the Owl HEP-1 blotter at 0.3 A for 40 minutes. After which, the membrane was blocked in 5% milk in TBST for 1 hour. The membrane was then decorated with anti-FLAG M2 antibodies from mice (Sigma Catalog no. F3165-1MG) in 1% milk (1:6666) overnight, followed by three 5 minutes washes in TBST (0.1% Tween-20). The membrane was then decorated with anti-mouse secondary antibody from goat conjugated with HRP (Thermo Fisher Catalog no. 32430) in TBST (1:1000) for 1 hour, followed by three 5 minutes washes in TBST. Blots were then imaged using SuperSignal West Pico (Thermo Scientific Catalog no. 34080) on the Bio-Rad Chemidoc. The secretion titer was then quantified by densitometry using the Image Lab software v5.2.1 provided by Bio-Rad. The values were then normalized by $OD_{600}$ and to pLac *hilA* induced with 100 μM IPTG.

For GroEL blots, anti-GroEL antibodies from rabbit (Sigma Catalog no. G6532-.5ML) in 1% milk (1:10000) and anti-rabbit secondary from goat conjugated with HRP (Thermo Fisher catalog no. 32460) were used instead.

## $OD_{600}$ and fluorescence using plate reader

200 μL of culture or supernatant were placed in a 96 well black bottom plate for fluorescence reading or clear flat bottom plates for $OD_{600}$. Growth curve for the different strains with anti-microbial peptides were generated using the plate reader. The $OD_{600}$ from the plate reader was corrected for pathlength as well as multiplied by a scaling factor in order to be comparable to the measurements from the Nanodrop 100c.

For mRuby3 fluorescence measurement, an excitation of 560/40 and a filter of 620/15 was used with a gain of 120. The samples were left in the dark at room temperature for 16 hours before measurement.

## References

1.  Goeddel, D. V. *et al.* Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proc. Natl. Acad. Sci.* **76**, 106–110 (1979).
2.  Williams, D. C., Frank, R. V., Muth, W. L. & Burnett, J. P. Cytoplasmic inclusion bodies in *Escherichia coli* producing biosynthetic human insulin proteins. *Science* **215**, 687–689 (1982).
3.  Baeshen, N. A. *et al.* Cell factories for insulin production. *Microb. Cell Factories* **13**, 141 (2014).
4.  Sanchez-Garcia, L. *et al.* Recombinant pharmaceuticals from microbial cells: a 2015 update. *Microb. Cell Factories* **15**, 33 (2016).
5.  Kesik-Brodacka, M. Progress in biopharmaceutical development. *Biotechnol. Appl. Biochem.* **65**, 306–322 (2018).
6.  Top 15 Best-Selling Drugs of 2018. *GEN - Genetic Engineering and Biotechnology News* https://www.genengnews.com/a-lists/top-15-best-selling-drugs-of-2018/ (2019).
7.  Westers, L., Westers, H. & Quax, W. J. *Bacillus subtilis* as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism. *Biochim. Biophys. Acta* **1694**, 299–310 (2004).
8.  Thomas, P. & Smart, T. G. HEK293 cell line: a vehicle for the expression of recombinant proteins. *J. Pharmacol. Toxicol. Methods* **51**, 187–200 (2005).
9.  Macauley-Patrick, S., Fazenda, M. L., McNeil, B. & Harvey, L. M. Heterologous protein production using the *Pichia pastoris* expression system. *Yeast Chichester Engl.* **22**, 249–270 (2005).
10. Kim, J. Y., Kim, Y.-G. & Lee, G. M. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Appl. Microbiol. Biotechnol.* **93**, 917–930 (2012).
11. Anné, J., Maldonado, B., Van Impe, J., Van Mellaert, L. & Bernaerts, K. Recombinant protein production and *streptomycetes*. *J. Biotechnol.* **158**, 159–167 (2012).
12. Ÿztürk, S., Ÿalık, P. & Ÿzdamar, T. H. Fed-Batch Biomolecule Production by *Bacillus subtilis*: A State of the Art Review. *Trends Biotechnol.* **34**, 329–345 (2016).
13. Irons, S. L., Chambers, A. C., Lissina, O., King, L. A. & Possee, R. D. Protein Production Using the Baculovirus Expression System. *Curr. Protoc. Protein Sci.* **91**, 5.5.1-5.5.22 (2018).
14. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
15. Meyer, Hans-Peter & Schmidhalter, Diego R. Microbial Expression Systems and Manufacturing from a Market and Economic Perspective. in *Innovations in Biotechnology* 211–246 (BoD – Books on Demand, 2012).
16. Ferreira, R. da G., Azzoni, A. R. & Freitas, S. Techno-economic analysis of the industrial production of a low-cost enzyme using E. coli: the case of recombinant β-glucosidase. *Biotechnol. Biofuels* **11**, 81 (2018).
17. Vermasvuori, R. *et al.* Production of recombinant HIV-1 *nef* protein using different expression host systems: A techno-economical comparison. *Biotechnol. Prog.* **25**, 95–102 (2009).
18. Balasundaram, B., Harrison, S. & Bracewell, D. G. Advances in product release strategies and impact on bioprocess design. *Trends Biotechnol.* **27**, 477–485 (2009).

19. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front. Microbiol.* **5**, (2014).

20. Burdette, L. A., Leach, S. A., Wong, H. T. & Tullman-Ercek, D. Developing Gram-negative bacteria for the secretion of heterologous proteins. *Microb. Cell Factories* **17**, (2018).

21. Costa, T. R. D. *et al.* Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat. Rev. Microbiol.* **13**, 343–359 (2015).

22. Baneyx, F. & Mujacic, M. Recombinant protein folding and misfolding in Escherichia coli. *Nat. Biotechnol.* **22**, 1399–1408 (2004).

23. Missiakas, D. & Raina, S. Protein folding in the bacterial periplasm. *J. Bacteriol.* **179**, 2465–2471 (1997).

24. Galán, J. E. & Wolf-Watz, H. Protein delivery into eukaryotic cells by type III secretion machines. *Nature* **444**, 567–573 (2006).

25. Erhardt, M., Namba, K. & Hughes, K. T. Bacterial Nanomachines: The Flagellum and Type III Injectisome. *Cold Spring Harb. Perspect. Biol.* **2**, a000299–a000299 (2010).

26. Cornelis, G. R. The type III secretion injectisome. *Nat. Rev. Microbiol.* **4**, 811–825 (2006).

27. Galán, J. E., Lara-Tejero, M., Marlovits, T. C. & Wagner, S. Bacterial Type III Secretion Systems: Specialized Nanomachines for Protein Delivery into Target Cells. *Annu. Rev. Microbiol.* **68**, 415–438 (2014).

28. Affolter, M., Parent-Vaugeois, C. & Anderson, A. Curing and induction of the Fels 1 and Fels 2 prophages in the Ames mutagen tester strains of *Salmonella typhimurium*. *Mutat. Res. Mol. Mech. Mutagen.* **110**, 243–262 (1983).

29. Knuth, K., Niesalla, H., Hueck, C. J. & Fuchs, T. M. Large-scale identification of essential *Salmonella* genes by trapping lethal insertions. *Mol. Microbiol.* **51**, 1729–1744 (2004).

30. Zinder, N. D. & Lederberg, J. Genetic exchange in *Salmonella*. *J. Bacteriol.* **64**, 679–699 (1952).

31. Lorenz, M. G. & Wackernagel, W. Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol. Rev.* **58**, 563–602 (1994).

32. Collazo, C. M. & Galan, J. E. Requirement for exported proteins in secretion through the invasion-associated type III system of *Salmonella typhimurium*. *Infect. Immun.* **64**, 3524–3531 (1996).

33. Hansen-Wester, I. & Hensel, M. *Salmonella* pathogenicity islands encoding type III secretion systems. *Microbes Infect.* **3**, 549–559 (2001).

34. Lostroh, C. P. & Lee, C. A. The Salmonella pathogenicity island-1 type III secretion system. *Microbes Infect.* **3**, 1281–1291 (2001).

35. Lostroh, C. P., Bajaj, V. & Lee, C. A. The cis requirements for transcriptional activation by HilA, a virulence determinant encoded on SPI-1. *Mol. Microbiol.* **37**, 300–315 (2000).

36. Fahlen, T. F., Mathur, N. & Jones, B. D. Identification and characterization of mutants with increased expression of *hilA*, the invasion gene transcriptional activator of *Salmonella typhimurium*. *FEMS Immunol. Med. Microbiol.* **28**, 25–35 (2000).

37. Klein, J. R., Fahlen, T. F. & Jones, B. D. Transcriptional Organization and Function of Invasion Genes within *Salmonella enterica* Serovar Typhimurium Pathogenicity Island 1, Including the *prgH*, *prgI*, *prgJ*, *prgK*, *orgA*, *orgB,* and *orgC* Genes. *Infect. Immun.* **68**, 3368–3376 (2000).

38. Renault, T. T., Guse, A. & Erhardt, M. Export Mechanisms and Energy Transduction in Type-III Secretion Machines. 1–17 (2019) doi:10.1007/82_2019_166.

39. Lee, P.-C. & Rietsch, A. Fueling type III secretion. *Trends Microbiol.* **23**, 296–300 (2015).

40. Kubori, T. *et al.* Supramolecular Structure of the *Salmonella typhimurium* Type III Protein Secretion System. *Science* **280**, 602–605 (1998).

41. Burkinshaw, B. J. & Strynadka, N. C. J. Assembly and structure of the T3SS. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1843**, 1649–1663 (2014).

42. Azam, A., Li, C., Metcalf, K. J. & Tullman-Ercek, D. Type III secretion as a generalizable strategy for the production of full-length biopolymer-forming proteins. *Biotechnol. Bioeng.* n/a-n/a (2015) doi:10.1002/bit.25656.

43. Metcalf, K. J., Finnerty, C., Azam, A., Valdivia, E. & Tullman-Ercek, D. Using Transcriptional Control To Increase Titers of Secreted Heterologous Proteins by the Type III Secretion System. *Appl. Environ. Microbiol.* **80**, 5927–5934 (2014).

44. Song, M. *et al.* Control of type III protein secretion using a minimal genetic system. *Nat. Commun.* **8**, 1–9 (2017).

45. Akbar, S., Schechter, L. M., Lostroh, C. P. & Lee, C. A. AraC/XylS family members, HilD and HilC, directly activate virulence gene expression independently of HilA in *Salmonella typhimurium*. *Mol. Microbiol.* **47**, 715–728 (2003).

46. Ellermeier, C. D., Ellermeier, J. R. & Slauch, J. M. HilD, HilC and RtsA constitute a feed forward loop that controls expression of the SPI1 type three secretion system regulator hilA in Salmonella enterica serovar Typhimurium. *Mol. Microbiol.* **57**, 691–705 (2005).

47. Ellermeier, C. D. & Slauch, J. M. RtsA and RtsB Coordinately Regulate Expression of the Invasion and Flagellar Genes in *Salmonella enterica* Serovar Typhimurium. *J. Bacteriol.* **185**, 5096–5108 (2003).

48. Ellermeier, C. D. & Slauch, J. M. RtsA Coordinately Regulates DsbA and the *Salmonella* Pathogenicity Island 1 Type III Secretion System. *J. Bacteriol.* **186**, 68–79 (2004).

49. Kaniga, K., Trollinger, D. & Galan, J. E. Identification of two targets of the type III protein secretion system encoded by the inv and spa loci of *Salmonella typhimurium* that have homology to the Shigella IpaD and IpaA proteins. *J. Bacteriol.* **177**, 7078–7085 (1995).

50. Sukhan, A., Kubori, T., Wilson, J. & Galan, J. E. Genetic Analysis of Assembly of the *Salmonella enterica* Serovar Typhimurium Type III Secretion-Associated Needle Complex. *J. Bacteriol.* **183**, 1159–1167 (2001).

51. Darwin, K. H. & Miller, V. L. Type III secretion chaperone-dependent regulation: activation of virulence genes by SicA and InvF in *Salmonella typhimurium*. *EMBO J.* **20**, 1850–1862 (2001).

52. Heran Darwin, K. & Miller, V. L. InvF Is Required for Expression of Genes Encoding Proteins Secreted by the SPI1 Type III Secretion Apparatus in *Salmonella typhimurium*. *J. Bacteriol.* **181**, 4949–4954 (1999).

53. Tucker, S. C. & Galán, J. E. Complex function for SicA, a *Salmonella enterica* serovar typhimurium type III secretion-associated chaperone. *J. Bacteriol.* **182**, 2262–2268 (2000).

54. Kubori, T., Sukhan, A., Aizawa, S.-I. & Galán, J. E. Molecular characterization and assembly of the needle complex of the *Salmonella typhimurium* type III protein secretion system. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 10225–10230 (2000).

55. Deane, J. E., Abrusci, P., Johnson, S. & Lea, S. M. Timing is everything: the regulation of type III secretion. *Cell. Mol. Life Sci. CMLS* **67**, 1065–1075 (2010).

56. Hautefort, I., Proença, M. J. & Hinton, J. C. D. Single-Copy Green Fluorescent Protein Gene Fusions Allow Accurate Measurement of *Salmonella* Gene Expression In Vitro and during Infection of Mammalian Cells. *Appl. Environ. Microbiol.* **69**, 7480–7491 (2003).

57. Sturm, A. *et al.* The Cost of Virulence: Retarded Growth of *Salmonella Typhimurium* Cells Expressing Type III Secretion System 1. *PLoS Pathog* **7**, e1002143 (2011).

58. Sánchez-Romero, M. A. & Casadesús, J. Contribution of SPI-1 bistability to *Salmonella enterica* cooperative virulence: insights from single cell analysis. *Sci. Rep.* **8**, 1–11 (2018).

59. Smith, C., Stringer, A. M., Mao, C., Palumbo, M. J. & Wade, J. T. Mapping the Regulatory Network for *Salmonella enterica* Serovar Typhimurium Invasion. *mBio* **7**, (2016).

60. Temme, K. *et al.* Induction and Relaxation Dynamics of the Regulatory Network Controlling the Type III Secretion System Encoded within *Salmonella* Pathogenicity Island 1. *J. Mol. Biol.* **377**, 47–61 (2008).

61. Saini, S., Ellermeier, J. R., Slauch, J. M. & Rao, C. V. The Role of Coupled Positive Feedback in the Expression of the SPI1 Type Three Secretion System in *Salmonella*. *PLOS Pathog.* **6**, e1001025 (2010).

62. Widmaier, D. M. *et al.* Engineering the *Salmonella* type III secretion system to export spider silk monomers. *Mol. Syst. Biol.* **5**, (2009).

63. Glasgow, A. A., Wong, H. T. & Tullman-Ercek, D. A Secretion-Amplification Role for *Salmonella enterica* Translocon Protein SipD. *ACS Synth. Biol.* **6**, 1006–1015 (2017).

64. Golubeva, Y. A., Sadik, A. Y., Ellermeier, J. R. & Slauch, J. M. Integrating Global Regulatory Input Into the *Salmonella* Pathogenicity Island 1 Type III Secretion System. *Genetics* **190**, 79–90 (2012).

65. Kroger, C. *et al.* The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proc. Natl. Acad. Sci.* **109**, E1277–E1286 (2012).

66. Ellermeier, J. R. & Slauch, J. M. Adaptation to the host environment: regulation of the SPI1 type III secretion system in *Salmonella enterica* serovar Typhimurium. *Curr. Opin. Microbiol.* **10**, 24–29 (2007).

67. Kelly, A. *et al.* A global role for Fis in the transcriptional control of metabolism and type III secretion in *Salmonella enterica* serovar Typhimurium. *Microbiology* **150**, 2037–2053 (2004).

68. Schechter, L. M., Jain, S., Akbar, S. & Lee, C. A. The Small Nucleoid-Binding Proteins H-NS, HU, and Fis Affect *hilA* Expression in *Salmonella enterica* Serovar Typhimurium. *Infect. Immun.* **71**, 5432–5435 (2003).

69. Kim, K., Palmer, A. D., Vanderpool, C. K. & Slauch, J. M. The Small RNA PinT Contributes to PhoP-Mediated Regulation of the *Salmonella* Pathogenicity Island 1 Type III Secretion System in *Salmonella enterica* Serovar Typhimurium. *J. Bacteriol.* **201**, (2019).

70. Aguirre, A. *et al.* PhoP-Induced Genes within *Salmonella* Pathogenicity Island 1. *J. Bacteriol.* **188**, 6889–6898 (2006).

71. Palmer, A. D., Kim, K. & Slauch, J. M. PhoP-Mediated Repression of the SPI1 Type 3 Secretion System in *Salmonella enterica* Serovar Typhimurium. *J. Bacteriol.* **201**, (2019).

72. Troxell, B. *et al.* Fur Negatively Regulates hns and Is Required for the Expression of HilA and Virulence in *Salmonella enterica* Serovar Typhimurium. *J. Bacteriol.* **193**, 497–505 (2011).

73. Ellermeier, J. R. & Slauch, J. M. Fur Regulates Expression of the *Salmonella* Pathogenicity Island 1 Type III Secretion System through HilD. *J. Bacteriol.* **190**, 476–486 (2008).

74. Rice, C. J., Ramachandran, V. K., Shearer, N. & Thompson, A. Transcriptional and Post-Transcriptional Modulation of SPI1 and SPI2 Expression by ppGpp, RpoS and DksA in *Salmonella enterica* sv Typhimurium. *PLOS ONE* **10**, e0127523 (2015).

75. Volk, M., Vollmer, I., Heroven, A. K. & Dersch, P. Transcriptional and Post-transcriptional Regulatory Mechanisms Controlling Type III Secretion. 1–23 (2019) doi:10.1007/82_2019_168.

76. Matsui, M., Takaya, A. & Yamamoto, T. σ$^{32}$-Mediated Negative Regulation of Salmonella Pathogenicity Island 1 Expression. *J. Bacteriol.* **190**, 6636–6645 (2008).

77. Westermann, A. J. *et al.* The Major RNA-Binding Protein ProQ Impacts Virulence Gene Expression in *Salmonella enterica* Serovar Typhimurium. *mBio* **10**, (2019).

78. Takaya, A., Kubota, Y., Isogai, E. & Yamamoto, T. Degradation of the HilC and HilD regulator proteins by ATP-dependent Lon protease leads to downregulation of *Salmonella* pathogenicity island 1 gene expression. *Mol. Microbiol.* **55**, 839–852 (2004).

79. Baxter, M. A. & Jones, B. D. Two-Component Regulators Control *hilA* Expression by Controlling *fimZ* and *hilE* Expression within *Salmonella enterica* Serovar Typhimurium. *Infect. Immun.* **83**, 978–985 (2015).

80. Galán, J. E. & Collmer, A. Type III Secretion Machines: Bacterial Devices for Protein Delivery into Host Cells. *Science* **284**, 1322–1328 (1999).

81. Burkinshaw, B. J. *et al.* Structural analysis of a specialized type III secretion system peptidoglycan-cleaving enzyme. *J. Biol. Chem.* **290**, 10406–10417 (2015).

82. Danese, P. N. & Silhavy, T. J. CpxP, a Stress-Combative Member of the Cpx Regulon. *J. Bacteriol.* **180**, 831–839 (1998).

83. Lin, D., Rao, C. V. & Slauch, J. M. The *Salmonella* SPI1 Type Three Secretion System Responds to Periplasmic Disulfide Bond Status via the Flagellar Apparatus and the RcsCDB System. *J. Bacteriol.* **190**, 87–97 (2008).

84. Subramaniam, S. *et al.* Contribution of the Cpx envelope stress system to metabolism and virulence regulation in *Salmonella enterica* serovar Typhimurium. *PloS One* **14**, e0211584 (2019).

85. Vicente, M., Chater, K. F. & Lorenzo, V. D. Bacterial transcription factors involved in global regulation. *Mol. Microbiol.* **33**, 8–17 (1999).

86. Mahalik, S., Sharma, A. K. & Mukherjee, K. J. Genome engineering for improved recombinant protein expression in *Escherichia coli*. *Microb. Cell Factories* **13**, 177 (2014).

87. Branchu, P., Bawn, M. & Kingsley, R. A. Genome Variation and Molecular Epidemiology of *Salmonella enterica* Serovar Typhimurium Pathovariants. *Infect. Immun.* **86**, (2018).

88. Pang, S. *et al.* Genomic diversity and adaptation of *Salmonella enterica* serovar Typhimurium from analysis of six genomes of different phage types. *BMC Genomics* **14**, 718 (2013).

89. Fookes, M. *et al. Salmonella bongori* Provides Insights into the Evolution of the *Salmonellae*. *PLOS Pathog.* **7**, e1002191 (2011).

90. Eswarappa, S. M., Karnam, G., Nagarajan, A. G., Chakraborty, S. & Chakravortty, D. lac Repressor Is an Antivirulence Factor of *Salmonella enterica*: Its Role in the Evolution of Virulence in *Salmonella*. *PLOS ONE* **4**, e5789 (2009).

91. Myeni, S. K., Wang, L. & Zhou, D. SipB-SipC Complex Is Essential for Translocon Formation. *PLoS ONE* **8**, e60499 (2013).

92. Wang, Y., Zhang, L., Picking, W. L., Picking, W. D. & Guzman, R. N. D. Structural dissection of the extracellular moieties of the type III secretion apparatus. *Mol. Biosyst.* **4**, 1176–1180 (2008).

93. Chatterjee, S. *et al.* The crystal structures of the *Salmonella* type III secretion system tip protein SipD in complex with deoxycholate and chenodeoxycholate. *Protein Sci.* **20**, 75–86 (2011).

94. Lunelli, M., Hurwitz, R., Lambers, J. & Kolbe, M. Crystal Structure of PrgI-SipD: Insight into a Secretion Competent State of the Type Three Secretion System Needle Tip and its Interaction with Host Ligands. *PLoS Pathog.* **7**, e1002163 (2011).

95. Rathinavelan, T. *et al.* NMR Model of PrgI–SipD Interaction and Its Implications in the Needle-Tip Assembly of the *Salmonella* Type III Secretion System. *J. Mol. Biol.* **426**, 2958–2969 (2014).

96. Veenendaal, A. K. J. *et al.* The type III secretion system needle tip complex mediates host cell sensing and translocon insertion. *Mol. Microbiol.* **63**, 1719–1730 (2007).

97. Hayward, R. D. *et al.* Cholesterol binding by the bacterial type III translocon is essential for virulence effector delivery into mammalian cells. *Mol. Microbiol.* **56**, 590–603 (2005).

98. Shivcharan, S., Yadav, J. & Qadri, A. Host lipid sensing promotes invasion of cells with pathogenic *Salmonella*. *Sci. Rep.* **8**, 1–11 (2018).

99. Bahrani, F. K., Sansonetti, P. J. & Parsot, C. Secretion of Ipa proteins by *Shigella flexneri*: inducer molecules and kinetics of activation. *Infect. Immun.* **65**, 4005–4010 (1997).

100. Fujii, T. *et al.* Structure of a type III secretion needle at 7-Å resolution provides insights into its assembly and signaling mechanisms. *Proc. Natl. Acad. Sci.* **109**, 4461–4466 (2012).

101. Nans, A., Kudryashev, M., Saibil, H. R. & Hayward, R. D. Structure of a bacterial type III secretion system in contact with a host membrane in situ. *Nat. Commun.* **6**, 10114 (2015).

102. López, E. & Blázquez, J. Effect of subinhibitory concentrations of antibiotics on intrachromosomal homologous recombination in *Escherichia coli*. *Antimicrob. Agents Chemother.* **53**, 3411–3415 (2009).

103. Johnson, S. *et al.* Self-chaperoning of the Type III Secretion System Needle Tip Proteins IpaD and BipD. *J. Biol. Chem.* **282**, 4035–4044 (2007).

104. Kaur, K., Chatterjee, S. & De Guzman, R. N. Characterization of the *Shigella* and *Salmonella* Type III Secretion System Tip–Translocon Protein–Protein Interaction by Paramagnetic Relaxation Enhancement. *ChemBioChem* **17**, 745–752 (2016).

105. Chen, Y. & Anderson, D. M. Expression hierarchy in the *Yersinia* type III secretion system established through YopD recognition of RNA: RNA recognition of target genes by *Yersinia* YopD. *Mol. Microbiol.* **80**, 966–980 (2011).

106. Matson, J. S. & Nilles, M. L. LcrG-LcrV interaction is required for control of Yops secretion in *Yersinia pestis*. *J. Bacteriol.* **183**, 5082–5091 (2001).

107. Williams, A. W. & Straley, S. C. YopD of Yersinia pestis plays a role in negative regulation of the low-calcium response in addition to its role in translocation of Yops. *J. Bacteriol.* **180**, 350–358 (1998).

108. Bröms, J. E., Francis, M. S. & Forsberg, Å. Diminished LcrV Secretion Attenuates *Yersinia pseudotuberculosis* Virulence. *J. Bacteriol.* **189**, 8417–8429 (2007).

109. Nanao, M. *et al.* Type III secretion proteins PcrV and PcrG from *Pseudomonas aeruginosa* form a 1:1 complex through high affinity interactions. *BMC Microbiol.* **3**, 21 (2003).

110. Roehrich, A. D., Guillossou, E., Blocker, A. J. & Martinez-Argudo, I. *Shigella* IpaD has a dual role: signal transduction from the type III secretion system needle tip and intracellular secretion regulation. *Mol. Microbiol.* **87**, 690–706 (2013).

111. Stebbins, C. E. & Galán, J. E. Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion. *Nature* **414**, 77–81 (2001).

112. Engler, C., Kandzia, R. & Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLOS ONE* **3**, e3647 (2008).

113. Laemmli, U. K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227**, 680–685 (1970).

114. Towbin, H., Staehelin, T. & Gordon, J. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 4350–4354 (1979).

115. Riedel, G. *et al.* An Extended ΔCT-Method Facilitating Normalisation with Multiple Reference Genes Suited for Quantitative RT-PCR Analyses of Human Hepatocyte-Like Cells. *PLOS ONE* **9**, e93031 (2014).

116. Pornsukarom, S., van Vliet, A. H. M. & Thakur, S. Whole genome sequencing analysis of multiple *Salmonella* serovars provides insights into phylogenetic relatedness, antimicrobial resistance, and virulence markers across humans, food animals and agriculture environmental sources. *BMC Genomics* **19**, 801 (2018).

117. Leekitcharoenphon, P., Lukjancenko, O., Friis, C., Aarestrup, F. M. & Ussery, D. W. Genomic variation in *Salmonella enterica* core genes for epidemiological typing. *BMC Genomics* **13**, 88 (2012).

118. McClelland, M. *et al.* Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* **413**, 852–856 (2001).

119. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **46**, D8–D13 (2018).

120. Figueroa-Bossi, N., Uzzau, S., Maloriol, D. & Bossi, L. Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in *Salmonella*. *Mol. Microbiol.* **39**, 260–271 (2001).

121. Boddicker, J. D. & Jones, B. D. Lon Protease Activity Causes Down-Regulation of *Salmonella* Pathogenicity Island 1 Invasion Gene Expression after Infection of Epithelial Cells. *Infect. Immun.* **72**, 2002–2013 (2004).

122. Takaya, A. *et al.* Lon, a Stress-Induced ATP-Dependent Protease, Is Critically Important for Systemic *Salmonella enterica* Serovar Typhimurium Infection of Mice. *Infect. Immun.* **71**, 690–696 (2003).

123. Kage, H., Takaya, A., Ohya, M. & Yamamoto, T. Coordinated Regulation of Expression of *Salmonella* Pathogenicity Island 1 and Flagellar Type III Secretion Systems by ATP-Dependent ClpXP Protease. *J. Bacteriol.* **190**, 2470–2478 (2008).

124. Song, M. *et al.* ppGpp-dependent Stationary Phase Induction of Genes on *Salmonella* Pathogenicity Island 1. *J. Biol. Chem.* **279**, 34183–34190 (2004).

125. Xu, D. *et al.* Genome Sequence of *Salmonella enterica* Serovar Typhi Oral Vaccine Strain Ty21a. *Genome Announc.* **1**, (2013).

126. Chen, K. *et al.* BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat. Methods* **6**, 677–681 (2009).

127. Werren, J. H. Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc. Natl. Acad. Sci.* **108**, 10863–10870 (2011).

128. Werren, J. H., Nur, U. & Wu, C. I. Selfish genetic elements. *Trends Ecol. Evol.* **3**, 297–302 (1988).

129. Lynch, M. & Marinov, G. K. The bioenergetic costs of a gene. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 15690–15695 (2015).

130. Iranzo, J., Puigbò, P., Lobkovsky, A. E., Wolf, Y. I. & Koonin, E. V. Inevitability of Genetic Parasites. *Genome Biol. Evol.* **8**, 2856–2869 (2016).

131. Kitamura, K., Torii, Y., Matsuoka, C. & Yamamoto, K. DNA sequence changes in mutations in the *tonB* gene on the chromosome of *Escherichia coli* K12: insertion elements dominate the spontaneous spectra. *Idengaku Zasshi* **70**, 35–46 (1995).

132. Van Engelenburg, S. B. & Palmer, A. E. Imaging type-III secretion reveals dynamics and spatial segregation of *Salmonella* effectors. *Nat. Methods* **7**, 325–330 (2010).

133. Khetrapal, V. *et al.* A set of powerful negative selection systems for unmodified Enterobacteriaceae. *Nucleic Acids Res.* **43**, e83 (2015).

134. Kim, K., Golubeva, Y. A., Vanderpool, C. K. & Slauch, J. M. Oxygen-dependent regulation of SPI1 type three secretion system by small RNAs in *Salmonella enterica* serovar Typhimurium. *Mol. Microbiol.* **111**, 570–587 (2019).

135. Kolisnychenko, V. *et al.* Engineering a Reduced *Escherichia coli* Genome. *Genome Res.* **12**, 640–647 (2002).

136. Choi, J. W., Yim, S. S., Kim, M. J. & Jeong, K. J. Enhanced production of recombinant proteins with *Corynebacterium glutamicum* by deletion of insertion sequences (IS elements). *Microb. Cell Factories* **14**, 207 (2015).

137. Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).

138. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114 (2014).

139. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv13033997 Q-Bio* (2013).

140. Picard Tools - By Broad Institute. http://broadinstitute.github.io/picard/.

141. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *ArXiv12073907 Q-Bio* (2012).

142. Narasimhan, V. *et al.* BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* **32**, 1749–1751 (2016).

143. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)* **6**, 80–92 (2012).

144. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

145. Sharan, S. K., Thomason, L. C., Kuznetsov, S. G. & Court, D. L. Recombineering: A Homologous Recombination-Based Method of Genetic Engineering. *Nat. Protoc.* **4**, 206–223 (2009).

146. Ellis, H. M., Yu, D., DiTizio, T. & Court, D. L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 6742–6746 (2001).

147. Angiuoli, S. V. & Salzberg, S. L. Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics* **27**, 334–342 (2011).

148. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).

149. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).

150. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, (2014).

151. Petrone, B. L., Stringer, A. M. & Wade, J. T. Identification of HilD-Regulated Genes in *Salmonella enterica* Serovar Typhimurium. *J. Bacteriol.* **196**, 1094–1101 (2014).

152. Hung, C.-C. *et al. Salmonella* invasion is controlled through the secondary structure of the hilD transcript. *PLOS Pathog.* **15**, e1007700 (2019).

153. López-Garrido, J., Puerta-Fernández, E. & Casadesús, J. A eukaryotic-like 3' untranslated region in *Salmonella enterica hilD* mRNA. *Nucleic Acids Res.* **42**, 5894–5906 (2014).

154. Grenz, J. R., Chubiz, J. E. C., Thaprawat, P. & Slauch, J. M. HilE Regulates HilD by Blocking DNA Binding in *Salmonella enterica* Serovar Typhimurium. *J. Bacteriol.* **200**, (2018).

155. Chubiz, J. E. C., Golubeva, Y. A., Lin, D., Miller, L. D. & Slauch, J. M. FliZ Regulates Expression of the Salmonella Pathogenicity Island 1 Invasion Locus by Controlling HilD Protein Activity in *Salmonella enterica* Serovar Typhimurium. *J. Bacteriol.* **192**, 6261–6270 (2010).

156. Mouali, Y. E. *et al.* CRP-cAMP mediates silencing of *Salmonella* virulence at the post-transcriptional level. *PLOS Genet.* **14**, e1007401 (2018).

157. Martínez, L. C. *et al.* Integration of a complex regulatory cascade involving the SirA/BarA and Csr global regulatory systems that controls expression of the *Salmonella* SPI-1 and SPI-2 virulence regulons through HilD. *Mol. Microbiol.* **80**, 1637–1656 (2011).

158. Fortune, D. R., Suyemoto, M. & Altier, C. Identification of CsrC and Characterization of Its Role in Epithelial Cell Invasion in *Salmonella enterica* Serovar Typhimurium. *Infect. Immun.* **74**, 331–339 (2006).

159. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

160. Biek, D. P. & Cohen, S. N. Identification and characterization of recD, a gene affecting plasmid maintenance and recombination in *Escherichia coli*. *J. Bacteriol.* **167**, 594–603 (1986).

161. Balleza, E., Kim, J. M. & Cluzel, P. Systematic characterization of maturation time of fluorescent proteins in living cells. *Nat. Methods* **15**, 47–51 (2018).

162. Tarazona, S. *et al.* Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res.* **43**, e140 (2015).

163. Teng, M. *et al.* A benchmark for RNA-seq quantification pipelines. *Genome Biol.* **17**, 74 (2016).

164. Blair, J. M. A., Richmond, G. E., Bailey, A. M., Ivens, A. & Piddock, L. J. V. Choice of Bacterial Growth Medium Alters the Transcriptome and Phenotype of *Salmonella enterica* Serovar Typhimurium. *PLoS ONE* **8**, e63912 (2013).

165. McClelland, M. *et al.* Comparison of the *Escherichia coli* K-12 genome with sampled genomes of a *Klebsiella pneumoniae* and three *Salmonella enterica* serovars, Typhimurium, Typhi and Paratyphi. *Nucleic Acids Res.* **28**, 4974–4986 (2000).

166. Surette, M. G., Miller, M. B. & Bassler, B. L. Quorum sensing in *Escherichia coli*, *Salmonella typhimurium*, and *Vibrio harveyi*: A new family of genes responsible for autoinducer production. *Proc. Natl. Acad. Sci.* **96**, 1639–1644 (1999).

167. Choi, J. *et al.* LsrR-Mediated Quorum Sensing Controls Invasiveness of *Salmonella typhimurium* by Regulating SPI-1 and Flagella Genes. *PLOS ONE* **7**, e37059 (2012).

168. Choi, J., Shin, D. & Ryu, S. Implication of quorum sensing in *Salmonella enterica* serovar typhimurium virulence: the *luxS* gene is necessary for expression of genes in pathogenicity island 1. *Infect. Immun.* **75**, 4885–4890 (2007).

169. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

170. Hu, Z., Mellor, J., Wu, J. & DeLisi, C. VisANT: an online visualization and analysis tool for biological interaction data. *BMC Bioinformatics* **5**, 17 (2004).

171. Cheng, S. *et al.* Identification of a Novel *Salmonella* Type III Effector by Quantitative Secretome Profiling. *Mol. Cell. Proteomics MCP* **16**, 2219–2228 (2017).

172. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

173. Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. & Tanabe, M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* **47**, D590–D595 (2019).

174. Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D. & Woolf, P. J. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* **10**, 161 (2009).

175. Luo, W. & Brouwer, C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinforma. Oxf. Engl.* **29**, 1830–1831 (2013).

176. Yu, N. Y. *et al.* PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**, 1608–1615 (2010).

177. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).

178. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).

179. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLOS ONE* **6**, e21800 (2011).

180. Shneiderman, B. Tree Visualization with Tree-maps: 2-d Space-filling Approach. *ACM Trans Graph* **11**, 92–99 (1992).

181. Tennekes, M. *treemap: Treemap Visualization*. (2017).

182. Dorman, C. J. H-NS: a universal regulator for a dynamic genome. *Nat. Rev. Microbiol.* **2**, 391–400 (2004).

183. Navarre, W. W. *et al.* Selective Silencing of Foreign DNA with Low GC Content by the H-NS Protein in *Salmonella*. *Science* **313**, 236–238 (2006).

184. Hommais, F. *et al.* Large-scale monitoring of pleiotropic regulation of gene expression by the prokaryotic nucleoid-associated protein, H-NS. *Mol. Microbiol.* **40**, 20–36 (2001).

185. Olekhnovich, I. N. & Kadner, R. J. Role of Nucleoid-Associated Proteins Hha and H-NS in Expression of *Salmonella enterica* Activators HilD, HilC, and RtsA Required for Cell Invasion. *J. Bacteriol.* **189**, 6882–6890 (2007).

186. Ali, S. S. *et al.* Silencing by H-NS Potentiated the Evolution of *Salmonella*. *PLoS Pathog* **10**, e1004500 (2014).

187. Owen-Hughes, T. A. *et al.* The chromatin-associated protein H-NS interacts with curved DNA to influence DNA topology and gene expression. *Cell* **71**, 255–265 (1992).

188. Hinton, J. C. D. *et al.* Expression and mutational analysis of the nucleoid-associated protein H-NS of *Salmonella typhimurium*. *Mol. Microbiol.* **6**, 2327–2337 (1992).

189. Ueguchi, C., Suzuki, T., Yoshida, T., Tanaka, K. & Mizuno, T. Systematic Mutational Analysis Revealing the Functional Domain Organization of *Escherichia coli* Nucleoid Protein H-NS. *J. Mol. Biol.* **263**, 149–162 (1996).

190. Troxell, B. & Hassan, H. M. Transcriptional regulation by Ferric Uptake Regulator (Fur) in pathogenic bacteria. *Front. Cell. Infect. Microbiol.* **3**, (2013).

191. Thompson, C. C. & Carabeo, R. An Optimal Method of Iron Starvation of the Obligate Intracellular Pathogen, *Chlamydia Trachomatis*. *Front. Microbiol.* **2**, (2011).

192. Liu, H. *et al.* Magic Pools: Parallel Assessment of Transposon Delivery Vectors in Bacteria. *mSystems* **3**, (2018).

193. Natarajan, A., Haitjema, C. H., Lee, R., Boock, J. T. & DeLisa, M. P. An Engineered Survival-Selection Assay for Extracellular Protein Expression Uncovers Hypersecretory Phenotypes in *Escherichia coli*. *ACS Synth. Biol.* **6**, 875–883 (2017).

194. Wong, B. G., Mancuso, C. P., Kiriakov, S., Bashor, C. J. & Khalil, A. S. Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER. *Nat. Biotechnol.* **36**, 614–623 (2018).

195. McGeachy, A. M., Meacham, Z. A. & Ingolia, N. T. An Accessible Continuous-Culture Turbidostat for Pooled Analysis of Complex Libraries. *ACS Synth. Biol.* **8**, 844–856 (2019).

196. Bryson, V. & Szybalski, W. Microbial Selection. *Science* **116**, 45–51 (1952).

197. Toprak, E. *et al.* Building a morbidostat: an automated continuous-culture device for studying bacterial drug resistance under dynamically sustained drug inhibition. *Nat. Protoc.* **8**, 555–567 (2013).

198. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A System for the Continuous Directed Evolution of Biomolecules. *Nature* **472**, 499 (2011).

199. Cellitti, J., Bernstein, R. & Marqusee, S. Exploring subdomain cooperativity in T4 lysozyme II: Uncovering the C-terminal subdomain as a hidden intermediate in the kinetic folding pathway. *Protein Sci.* **16**, 852–862 (2007).

200. Matsumura, M. & Matthews, B. W. Control of enzyme activity by an engineered disulfide bond. *Science* **243**, 792–794 (1989).

201. Cabantous, S. & Waldo, G. S. In vivo and in vitro protein solubility assays using split GFP. *Nat. Methods* **3**, 845–854 (2006).

202. Charrier, M. *et al.* Engineering the S-Layer of *Caulobacter crescentus* as a Foundation for Stable, High-Density, 2D Living Materials. *ACS Synth. Biol.* **8**, 181–190 (2019).

203. Bakkes, P. J., Jenewein, S., Smits, S. H. J., Holland, I. B. & Schmitt, L. The rate of folding dictates substrate secretion by the *Escherichia coli* hemolysin type 1 secretion system. *J. Biol. Chem.* **285**, 40573–40580 (2010).

204. Drobnak, I., Braselmann, E. & Clark, P. L. Multiple Driving Forces Required for Efficient Secretion of Autotransporter Virulence Proteins. *J. Biol. Chem.* **290**, 10104–10116 (2015).

205. Shahrezaei, V. & Marguerat, S. Connecting growth with gene expression: of noise and numbers. *Curr. Opin. Microbiol.* **25**, 127–135 (2015).

206. Gutierrez, J. M. *et al.* Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion. *bioRxiv* 351387 (2019) doi:10.1101/351387.

207. Thiele, I. *et al.* A community effort towards a knowledge-base and mathematical model of the human pathogen *Salmonella* Typhimurium LT2. *BMC Syst. Biol.* **5**, 8 (2011).

208. Chen, W.-H., Minguez, P., Lercher, M. J. & Bork, P. OGEE: an online gene essentiality database. *Nucleic Acids Res.* **40**, D901–D906 (2012).

209. Prabhu, A. A., Hariramani, K., Lakshmi, P. & Dasu, V. V. Systems Metabolic Engineering Approach for Recombinant Protein Production in Microbial Cell Factories. in *Horizons in Bioprocess Engineering* (ed. Pogaku, R.) 211–240 (Springer International Publishing, 2019). doi:10.1007/978-3-030-29069-6_12.

210. Akashi, H. & Gojobori, T. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc. Natl. Acad. Sci.* **99**, 3695–3700 (2002).

211. Meyer, D. E. & Chilkoti, A. Purification of recombinant proteins by fusion with thermally-responsive polypeptides. *Nat. Biotechnol.* **17**, 1112–1115 (1999).

212. Fong, B. A. & Wood, D. W. Expression and purification of ELP-intein-tagged target proteins in high cell density E. coli fermentation. *Microb. Cell Factories* **9**, 77 (2010).

*Appendix I List of Gene Ontology terms summarized by REVIGO*

| GO term | Description | No of genes in set | q value |
|---------|-------------|:------------------:|:-------:|
| **Biological Process GO terms for upregulated genes** | | | |
| **GO:0000003** | reproduction | 54 | 9.60E-03 |
| **GO:0006935** | chemotaxis | 124 | 7.73E-07 |
| **GO:0008104** | protein localization | 466 | 4.41E-36 |
| **GO:0009150** | purine ribonucleotide metabolic process | 332 | 2.37E-18 |
| **GO:0009653** | anatomical structure morphogenesis | 237 | 4.14E-04 |
| **GO:0023052** | signaling | 270 | 6.37E-06 |
| **GO:0032501** | multicellular organismal process | 13 | 2.61E-03 |
| **GO:0032502** | developmental process | 243 | 1.23E-04 |
| **GO:0032505** | reproduction of a single-celled organism | 54 | 9.60E-03 |
| **GO:0040007** | growth | 64 | 5.51E-15 |
| **GO:0040008** | regulation of growth | 64 | 5.51E-15 |
| **GO:0040011** | locomotion | 206 | 5.31E-03 |
| **GO:0044419** | interspecies interaction between organisms | 374 | 5.32E-54 |
| **GO:0051704** | multi-organism process | 442 | 3.24E-51 |
| **GO:0006458** | 'de novo' protein folding | 14 | 3.15E-04 |
| **GO:0007154** | cell communication | 396 | 9.60E-06 |
| **GO:0016999** | antibiotic metabolic process | 270 | 8.70E-05 |
| **GO:0043094** | cellular metabolic compound salvage | 86 | 1.28E-03 |
| **GO:0070647** | protein modification by small protein conjugation or removal | 27 | 2.70E-06 |
| **GO:0044036** | cell wall macromolecule metabolic process | 289 | 3.22E-03 |
| **GO:0046148** | pigment biosynthetic process | 87 | 3.61E-03 |
| **GO:0042440** | pigment metabolic process | 89 | 3.67E-03 |
| **GO:0001505** | regulation of neurotransmitter levels | 47 | 9.52E-03 |
| **GO:0001817** | regulation of cytokine production | 13 | 2.61E-03 |
| **GO:0012501** | programmed cell death | 15 | 5.62E-04 |
| **GO:0008219** | cell death | 15 | 5.62E-04 |
| **GO:0007059** | chromosome segregation | 61 | 6.41E-04 |
| **GO:0045333** | cellular respiration | 373 | 3.94E-14 |
| **GO:0022402** | cell cycle process | 83 | 9.54E-04 |
| **GO:0051301** | cell division | 318 | 1.24E-08 |
| **GO:0007049** | cell cycle | 293 | 5.29E-09 |
| **GO:0048518** | positive regulation of biological process | 130 | 5.39E-12 |
| **GO:0051338** | regulation of transferase activity | 11 | 2.03E-05 |
| **GO:0002682** | regulation of immune system process | 13 | 2.61E-03 |

| GO:0034660 | ncRNA metabolic process | 489 | 4.28E-11 |
|---|---|---|---|
| GO:0042726 | flavin-containing compound metabolic process | 27 | 5.94E-04 |
| GO:0015833 | peptide transport | 477 | 3.01E-31 |
| GO:0010646 | regulation of cell communication | 13 | 2.61E-03 |
| GO:0023051 | regulation of signaling | 21 | 9.60E-06 |
| GO:0031347 | regulation of defense response | 13 | 2.61E-03 |
| GO:0065009 | regulation of molecular function | 54 | 1.89E-04 |
| GO:0006353 | DNA-templated transcription, termination | 32 | 2.11E-03 |
| GO:0009266 | response to temperature stimulus | 42 | 1.73E-06 |
| GO:0032879 | regulation of localization | 78 | 5.25E-14 |
| GO:0006260 | DNA replication | 256 | 1.88E-07 |
| GO:0006354 | DNA-templated transcription, elongation | 22 | 9.65E-03 |
| GO:0048522 | positive regulation of cellular process | 100 | 8.39E-10 |
| GO:0032940 | secretion by cell | 226 | 1.88E-34 |
| GO:0009231 | riboflavin biosynthetic process | 27 | 5.94E-04 |
| GO:0051128 | regulation of cellular component organization | 286 | 1.39E-07 |
| GO:0035872 | nucleotide-binding domain, leucine rich repeat containing receptor signaling pathway | 13 | 2.61E-03 |
| GO:0048523 | negative regulation of cellular process | 170 | 4.14E-04 |
| GO:0048519 | negative regulation of biological process | 205 | 1.93E-04 |
| GO:0006730 | one-carbon metabolic process | 47 | 4.38E-03 |
| GO:0051603 | proteolysis involved in cellular protein catabolic process | 21 | 3.56E-04 |
| GO:0072350 | tricarboxylic acid metabolic process | 152 | 8.87E-08 |
| GO:0051641 | cellular localization | 124 | 1.16E-04 |
| GO:0006352 | DNA-templated transcription, initiation | 43 | 7.26E-04 |
| GO:0042133 | neurotransmitter metabolic process | 47 | 9.52E-03 |
| GO:0009605 | response to external stimulus | 302 | 5.13E-06 |
| GO:0072521 | purine-containing compound metabolic process | 387 | 8.58E-18 |
| GO:0033036 | macromolecule localization | 500 | 2.35E-36 |
| GO:0051174 | regulation of phosphorus metabolic process | 20 | 4.98E-04 |
| GO:0009628 | response to abiotic stimulus | 83 | 1.62E-03 |
| GO:0030522 | intracellular receptor signaling pathway | 13 | 2.61E-03 |
| GO:0034470 | ncRNA processing | 351 | 2.46E-07 |
| GO:0006412 | translation | 493 | 7.04E-13 |
| GO:0006885 | regulation of pH | 12 | 9.65E-03 |
| GO:0043038 | amino acid activation | 139 | 1.93E-06 |

| GO:0006396 | RNA processing | 372 | 1.69E-07 |
|---|---|---|---|
| GO:0001932 | regulation of protein phosphorylation | 11 | 2.03E-05 |
| GO:0016567 | protein ubiquitination | 27 | 2.70E-06 |
| GO:0009451 | RNA modification | 234 | 5.74E-04 |
| GO:0006090 | pyruvate metabolic process | 114 | 1.31E-02 |
| **Biological Process GO terms for downregulated genes** | | | |
| GO:0009243 | O antigen biosynthetic process | 80 | 2.39E-14 |
| GO:0097164 | ammonium ion metabolic process | 99 | 5.10E-09 |
| GO:1901615 | organic hydroxy compound metabolic process | 318 | 9.32E-05 |
| GO:0008643 | carbohydrate transport | 353 | 2.88E-06 |
| GO:0006307 | DNA dealkylation involved in DNA repair | 29 | 3.58E-02 |
| GO:0042439 | ethanolamine-containing compound metabolic process | 41 | 1.83E-12 |
| GO:0043711 | pilus organization | 105 | 2.01E-06 |
| GO:1901160 | primary amino compound metabolic process | 41 | 1.83E-12 |
| GO:0009437 | carnitine metabolic process | 31 | 2.53E-03 |
| GO:0009308 | amine metabolic process | 105 | 3.51E-06 |
| GO:0006577 | amino-acid betaine metabolic process | 31 | 2.53E-03 |
| GO:0019627 | urea metabolic process | 34 | 2.40E-02 |
| GO:1902777 | 6-sulfoquinovose(1-) catabolic process | 12 | 2.97E-03 |
| GO:0006525 | arginine metabolic process | 119 | 5.29E-06 |
| GO:0019673 | GDP-mannose metabolic process | 31 | 5.28E-05 |
| GO:0051167 | xylulose 5-phosphate metabolic process | 13 | 4.05E-02 |
| GO:0030001 | metal ion transport | 247 | 1.65E-03 |
| GO:1902776 | 6-sulfoquinovose(1-) metabolic process | 12 | 2.97E-03 |
| GO:0070689 | L-threonine catabolic process to propionate | 10 | 4.48E-04 |
| GO:1901565 | organonitrogen compound catabolic process | 394 | 7.07E-03 |
| GO:0030030 | cell projection organization | 208 | 1.84E-05 |
| GO:0009225 | nucleotide-sugar metabolic process | 143 | 2.14E-02 |
| GO:0019301 | rhamnose catabolic process | 14 | 8.55E-03 |
| GO:0015891 | siderophore transport | 25 | 4.11E-02 |
| GO:0036376 | sodium ion export from cell | 20 | 9.32E-04 |
| GO:1901678 | iron coordination entity transport | 36 | 3.67E-02 |
| GO:0009109 | coenzyme catabolic process | 23 | 4.17E-02 |
| GO:0046402 | O antigen metabolic process | 80 | 2.39E-14 |
| GO:0009111 | vitamin catabolic process | 23 | 4.17E-02 |
| GO:0051187 | cofactor catabolic process | 40 | 3.58E-02 |

| Cellular Component GO terms for upregulated genes | | | |
|---|---|---|---|
| GO:0005576 | extracellular region | 213 | 1.12E-28 |
| GO:0005840 | ribosome | 266 | 8.08E-13 |
| GO:0009295 | nucleoid | 80 | 1.30E-02 |
| GO:0033643 | host cell part | 122 | 1.65E-43 |
| GO:0044215 | other organism | 127 | 1.26E-45 |
| GO:0044421 | extracellular region part | 25 | 8.10E-05 |
| GO:0070469 | respiratory chain | 76 | 5.07E-07 |
| GO:0019867 | outer membrane | 329 | 3.69E-03 |
| GO:0032153 | cell division site | 33 | 3.33E-03 |
| GO:0009986 | cell surface | 32 | 2.06E-06 |
| GO:0030312 | external encapsulating structure | 363 | 3.40E-02 |
| GO:0045281 | succinate dehydrogenase complex | 14 | 2.61E-05 |
| GO:0030257 | type III protein secretion system complex | 32 | 5.62E-04 |
| GO:0045239 | tricarboxylic acid cycle enzyme complex | 23 | 1.28E-03 |
| GO:0048476 | Holliday junction resolvase complex | 13 | 4.86E-02 |
| GO:0032993 | protein-DNA complex | 84 | 9.28E-03 |
| GO:0044445 | cytosolic part | 218 | 1.28E-11 |
| GO:1990204 | oxidoreductase complex | 109 | 1.32E-02 |
| GO:0098796 | membrane protein complex | 384 | 4.17E-07 |
| Cellular Component GO terms for downregulated genes | | | |
| GO:0009350 | ethanolamine ammonia-lyase complex | 11 | 2.53E-03 |
| GO:0009289 | pilus | 64 | 3.42E-02 |
| GO:0009346 | citrate lyase complex | 21 | 1.26E-02 |
| Molecular Function GO terms for upregulated genes | | | |
| GO:0000988 | transcription factor activity, protein binding | 43 | 1.28E-03 |
| GO:0001216 | bacterial-type RNA polymerase transcriptional activator activity, sequence-specific DNA binding | 36 | 2.78E-02 |
| GO:0003735 | structural constituent of ribosome | 251 | 1.44E-12 |
| GO:0004888 | transmembrane signaling receptor activity | 44 | 1.41E-02 |
| GO:0005096 | GTPase activator activity | 13 | 1.41E-05 |
| GO:0005198 | structural molecule activity | 308 | 7.18E-09 |
| GO:0009055 | electron carrier activity | 370 | 6.52E-03 |
| GO:0016987 | sigma factor activity | 37 | 1.85E-03 |
| GO:0044769 | ATPase activity, coupled to transmembrane movement of ions, rotational mechanism | 47 | 1.68E-11 |
| GO:0051020 | GTPase binding | 14 | 3.32E-10 |
| GO:0098772 | molecular function regulator | 76 | 6.11E-03 |

| GO:0000104 | succinate dehydrogenase activity | 44 | 1.10E-07 |
|---|---|---|---|
| GO:0061783 | peptidoglycan muralytic activity | 60 | 7.79E-03 |
| GO:0003916 | DNA topoisomerase activity | 43 | 7.64E-03 |
| GO:0019239 | deaminase activity | 39 | 1.13E-02 |
| GO:0004842 | ubiquitin-protein transferase activity | 27 | 2.70E-06 |
| GO:0004812 | aminoacyl-tRNA ligase activity | 144 | 1.26E-05 |
| GO:0016874 | ligase activity | 475 | 8.70E-05 |
| GO:0019843 | rRNA binding | 204 | 2.32E-09 |
| GO:0072341 | modified amino acid binding | 15 | 3.87E-02 |
| GO:0043021 | ribonucleoprotein complex binding | 66 | 1.21E-02 |
| GO:0048038 | quinone binding | 86 | 9.92E-07 |
| GO:0033218 | amide binding | 77 | 4.75E-02 |
| GO:0051537 | 2 iron, 2 sulfur cluster binding | 146 | 4.43E-03 |
| GO:0044877 | macromolecular complex binding | 66 | 1.21E-02 |
| GO:0005515 | protein binding | 443 | 6.01E-09 |
| GO:0005525 | GTP binding | 167 | 1.16E-02 |
| GO:0005507 | copper ion binding | 40 | 1.42E-02 |
| GO:0019205 | nucleobase-containing compound kinase activity | 65 | 4.14E-05 |
| GO:0019787 | ubiquitin-like protein transferase activity | 27 | 2.70E-06 |
| GO:0016763 | transferase activity, transferring pentosyl groups | 113 | 1.40E-03 |
| GO:0003697 | single-stranded DNA binding | 34 | 1.84E-02 |
| GO:0008863 | formate dehydrogenase (NAD+) activity | 40 | 8.87E-03 |
| GO:0016682 | oxidoreductase activity, acting on diphenols and related substances as donors, oxygen as acceptor | 23 | 1.97E-07 |
| GO:0070011 | peptidase activity, acting on L-amino acid peptides | 291 | 4.83E-05 |
| GO:0016814 | hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in cyclic amidines | 62 | 6.52E-04 |
| GO:0016747 | transferase activity, transferring acyl groups other than amino-acyl groups | 236 | 2.62E-02 |
| GO:0008137 | NADH dehydrogenase (ubiquinone) activity | 66 | 1.70E-07 |
| GO:0016679 | oxidoreductase activity, acting on diphenols and related substances as donors | 23 | 1.97E-07 |
| GO:0016725 | oxidoreductase activity, acting on CH or CH2 groups | 29 | 1.19E-02 |
| GO:0016675 | oxidoreductase activity, acting on a heme group of donors | 21 | 2.03E-05 |
| GO:0015002 | heme-copper terminal oxidase activity | 23 | 1.97E-07 |

| GO:0017171 | serine hydrolase activity | 117 | 3.49E-02 |
|---|---|---|---|
| GO:0042578 | phosphoric ester hydrolase activity | 224 | 8.02E-04 |
| GO:0008728 | GTP diphosphokinase activity | 12 | 3.08E-02 |
| GO:0016746 | transferase activity, transferring acyl groups | 281 | 2.71E-02 |
| GO:0016627 | oxidoreductase activity, acting on the CH-CH group of donors | 144 | 3.83E-04 |
| GO:0008233 | peptidase activity | 334 | 9.05E-05 |
| GO:0043138 | 3'-5' DNA helicase activity | 20 | 2.47E-02 |
| GO:0016651 | oxidoreductase activity, acting on NAD(P)H | 243 | 1.91E-02 |
| GO:0019001 | guanyl nucleotide binding | 180 | 2.35E-02 |
| GO:0043621 | protein self-association | 14 | 9.26E-07 |
| GO:0004550 | nucleoside diphosphate kinase activity | 10 | 6.36E-03 |
| GO:0016776 | phosphotransferase activity, phosphate group as acceptor | 66 | 4.47E-04 |
| GO:0008893 | guanosine-3',5'-bis(diphosphate) 3'-diphosphatase activity | 12 | 3.08E-02 |
| GO:0016794 | diphosphoric monoester hydrolase activity | 12 | 3.08E-02 |
| GO:0070403 | NAD+ binding | 16 | 2.78E-02 |
| GO:0051087 | chaperone binding | 41 | 2.34E-09 |
| GO:0070181 | small ribosomal subunit rRNA binding | 11 | 8.73E-03 |
| GO:0016779 | nucleotidyltransferase activity | 311 | 3.80E-02 |
| GO:0003899 | DNA-directed 5'-3' RNA polymerase activity | 35 | 2.82E-02 |
| **Molecular Function GO terms for downregulated genes** | | | |
| GO:0015293 | symporter activity | 153 | 1.17E-04 |
| GO:0016830 | carbon-carbon lyase activity | 331 | 1.50E-06 |
| GO:0031419 | cobalamin binding | 28 | 4.62E-04 |
| GO:0004615 | phosphomannomutase activity | 11 | 6.65E-03 |
| GO:0004067 | asparaginase activity | 10 | 2.21E-02 |
| GO:0103111 | D-glucosamine PTS permease activity | 144 | 2.30E-03 |
| GO:0004022 | alcohol dehydrogenase (NAD) activity | 24 | 3.50E-02 |
| GO:0004585 | ornithine carbamoyltransferase activity | 11 | 1.11E-02 |
| GO:0008410 | CoA-transferase activity | 17 | 1.56E-02 |
| GO:0008378 | galactosyltransferase activity | 19 | 3.42E-02 |
| GO:0016597 | amino acid binding | 45 | 3.80E-03 |
| GO:0008905 | mannose-phosphate guanylyltransferase activity | 10 | 1.49E-02 |
| GO:0015473 | fimbrial usher porin activity | 45 | 1.33E-03 |
| GO:0019200 | carbohydrate kinase activity | 115 | 2.97E-03 |
| GO:0008662 | 1-phosphofructokinase activity | 10 | 3.58E-02 |

| GO:0016774 | phosphotransferase activity, carboxyl group as acceptor | 65 | 1.23E-02 |
|---|---|---|---|
| GO:0015101 | organic cation transmembrane transporter activity | 36 | 2.82E-02 |
| GO:0008742 | L-ribulose-phosphate 4-epimerase activity | 18 | 1.56E-02 |
| GO:0008815 | citrate (pro-3S)-lyase activity | 24 | 3.62E-03 |
| GO:0016743 | carboxyl- or carbamoyltransferase activity | 23 | 3.03E-02 |
| GO:0004368 | glycerol-3-phosphate dehydrogenase activity | 22 | 4.61E-02 |
| GO:0016861 | intramolecular oxidoreductase activity, interconverting aldoses and ketoses | 79 | 2.21E-02 |