Computationally Analyzing the Effect of Cannabidiol on Enterohemorrhagic *E. coli* Using Transcriptomic Data

---

A Thesis

Presented to

The Division of Mathematical and Natural Sciences

Reed College

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Arts

Sol Taylor-Brill

May 2020

Approved for the Division

(Biology)

Jay Mellies

Anna Ritz

Acknowledgments

I would like to begin by thanking my professors who have helped me grow as a student and a human being over the past four years. Anna Ritz and Jay Mellies, thank you for being the most incredible thesis advisors. I can't imagine having done this without your support and guidance. Thank you for your endless patience with my millions of drafts and small crises. You both inspire me so much as scientists and people. There was a point during my time at Reed where I felt like I wasn't cut out to be a scientist, but your genuine passion for your fields of research reminded me why I fell in love with science in the first place, and your confidence in me has kept me going even when I wasn't confident in myself. To my humanities professors, Jay Dickson and Steve Wasserstrom, thank you for challenging and engaging me, even in topics I knew little about. Your classes were truly a joy to take, and the skills that you taught me in logic and writing have helped me in every area of my life. To Kelly Chacón and Leia Harper, thank you for being on my orals board. I'm excited to discuss my thesis with you, because your classes were really influential on me, and I look up to you both very much.

To my incredible friends, thank you for taking this journey with me. You truly have been the greatest source of joy in my life. To the inhabitants of the B101 thesis office, Carter, Purna, Jesse, Maddox, Gabe, Olivia, Loralee, Claire, Jamie, Karl, and Aidan, thank you for the support, laughter, memes, commiseration, and late-night thesis parties. I feel so grateful to have gone through this process with such an incredible group of people, and I am so sad that I cannot celebrate with you all in person.

To the first and dearest friends I made at Reed, Kaitlyn, Janet, Guin, Isaac, and Kiara: It is impossible for me to describe just how much I love you all. Thank you for supporting me through every success, failure, heartbreak, and chaotic scheme. Never in my life have I felt so surrounded by love, and that has given me the ability to do things that I never believed possible. The person I am today would be unrecognizable to the lonely, anxious, insecure person I was when I came to Reed. Thank you for giving me the space, confidence, and love I needed to grow into myself. So much of each of you has become woven into me, that I can't even separate myself from the love I have for you all. You are all truly the smartest, kindest, funniest people I know, and I miss you so much already. Kaitlyn, I think I must be the luckiest person alive to have been randomly assigned to live with you freshman year. Thank you for being the most incredible roommate and friend. When I think of you I am filled with not only love, but respect and admiration for how incredibly strong, kind, intelligent, and resilient you are. Thank you for loving me and challenging me. I am, without a doubt, a better person for having known you.

To my parents: thank you for your endless support and patience. You have supported me through changed majors, changed career plans, and personal crises, and have never given up on me. Thank you for listening to my midnight rants, crazy ideas, and ever-changing ten-year plans. I am so grateful for all the sacrifices that you have made for me, and I love you both very much.

To everyone else who has been a part of my time at Reed: thank you for making this community what it is. I have such beautiful memories from my time at Reed that I will hold with me forever.

List of Abbreviations

A/E	Attaching and effacing
CBD	Cannabidiol
EHEC	Enterohemorrhagic *E. coli*
Epi/NE	Epinephrine and norepinephrine
Gb_3	Globotriaosylceramide
HUS	Hemolytic Uremic Syndrome
LEE	Locus of enterocyte effacement
RAJ	Recto-anal junction
Stx	Shiga Toxin
THC	Tetrahydrocannabinol
T3SS	Type III Secretion System
2-AG	2-arachidonyl glycerol

Table of Contents

List of Tables

List of Figures

Abstract

Enterohemorrhagic *E. coli* O157 (EHEC) is a type of foodborne bacteria that can cause severe symptoms in humans including kidney failure and even death. Unfortunately, there is currently no effective treatment for this type of bacteria because traditional antibiotics actually increase the severity of symptoms by triggering an increase in toxin secretion. Emerging research has indicated that 2-AG, an endocannabinoid produced in the human intestinal tract, reduces the virulence of EHEC. Based on its structural similarity to 2-AG, CBD, a non-psychotropic phytocannabinoid, has been investigated as a potential treatment for EHEC. Past research has indicated that CBD reduces virulence in a similar type of bacteria, *C. rodentium*, in mice. However, the molecular mechanism underlying this effect is unknown. The goal of this thesis was to clarify the mechanisms by which CBD downregulates virulence in EHEC by computationally analyzing RNA-Seq data by running a gene set enrichment analysis and constructing an interactome. I hypothesized that type III secretion system (T3SS) genes, which encode a molecular syringe that EHEC uses to attach to host cells and inject effector proteins, would be downregulated in response to CBD. However, I found that these genes were actually upregulated. Instead, I found that a number of other pathways, primarily those related to metabolism and translation, were differentially regulated. I also identified several stress pathways as candidates for further investigation and created qPCR primers for genes in these pathways to be used in future experiments. Further research is needed to untangle the roles of metabolic, translation, and stress pathways in moderating EHEC's virulence in response to CBD.

For my parents.

Without your support this wouldn't have been possible.

Introduction

1.1 Enterohemorrhagic *E. coli* (EHEC)

E. coli O157, also known as enterohemorrhagic *E. coli* (EHEC), is a zoonotic (transmitted from animals to humans), food-borne bacteria that can cause bloody diarrhea, abdominal cramps, a type of kidney failure called hemolytic uremic syndrome (HUS), or even death. Annually, EHEC causes around 70,000 illnesses and 60 deaths in the United States (Noris & Remuzzi, 2005) and is the leading cause of acute kidney failure in otherwise healthy children (Mayer et al., 2012).

Unfortunately, there is currently no effective treatment for EHEC, as traditional antibiotics can actually increase the severity of symptoms by increasing Shiga toxin production and secretion, which is the primary cause of mortality associated with EHEC (Pacheco & Sperandio, 2012). A meta-analysis of 19 studies found that antibiotics did not effectively reduce either the duration or severity of illness (Panos et al., 2006), and a separate study found that antibiotic treatment of children with *E. coli* O157 was associated with a 3x greater risk of developing HUS (Wong et al., 2012). Therefore, only supportive care measures such as intravenous fluids, blood cell transfusions, and dialysis, in the case of kidney failure, are currently given to patients suffering from the disease (Mayer et al., 2012). As such, alternative treatments such as anti-virulence compounds must be investigated to discover safe and effective interventions for this disease.

This thesis investigates cannabidiol (CBD) as a potential treatment for EHEC, based on past research showing its effectivity in reducing colonization by

a similar model bacteria (*Citrobacter rodentium*) in mice (Novy, 2019). CBD and other antivirulence treatments will be discussed in more detail later.

1.2 Shiga Toxin

Shiga toxin (Stx) is a cytotoxin that is encoded by an inducible, lysogenic, lamboid bacteriophage in EHEC's chromosome (Melton-Celsa, 2014). Stress signals increase the transcription and production of Stx by inducing the prophage encoding these genes and causing the phage to enter into its lytic cycle, resulting in cell lysis and the release of Stx into the extracellular environment (Bielazewska, 2012). The prophage is induced by the auto-cleavage of the *cI* repressor through a mechanism mediated by RecA, which is produced and activated during the SOS response, a highly conserved response to DNA damage (Mühldorfer et al., 1996). Antibiotics that interfere with DNA synthesis such as quinolones or mitomycins may be especially likely to trigger the SOS response and increase Stx production (McGannon et al., 2010; Zhang et al., 2000), and are likely particularly dangerous for patients with EHEC. Other environmental factors such as temperature, iron concentration, growth phase, and quorum sensing also play a role in regulating the expression of Stx in EHEC (Pacheco & Sperandio, 2012). Vascular damage to the colon and kidney by Stx is thought to be the cause of hemorrhagic colitis and HUS (Pruimboom-Brees et al., 2000), two of the more severe potential complications of EHEC infection in humans. Stx can also bind to receptors in the central nervous system, causing seizures, cerebral edema or coma (Kumar et al., 2019).

EHEC produces two highly related Stx toxins (Stx1 and Stx2) (Melton-Celsa, 2014). Although the two toxins are very similar in structure and mechanism of pathogenicity, Stx2 is associated with more severe symptoms

(Melton-Celsa, 2014) and, unlike Stx1, Stx2 is only produced during the phage's lytic cycle (Tyler et al., 2004). Stx damages tissue by inhibiting protein synthesis and inducing cell death (Pacheco & Sperandio, 2012). Both Stx toxins are protein complexes composed of five β-subunits and one α-subunit (Obrig et al., 1993). The β-subunit pentamer binds with high affinity to globotriaosylceramide (Gb_3), a receptor that is expressed on endothelial tissue in the colon and other organs in humans (Obrig et al., 1993). After binding to Gb_3, the toxin is internalized into affected cells within an endsome (Melton-Celsa, 2014).

Inside the cell, the endosome-contained Stx is first trafficked to the Golgi apparatus, then to the endoplasmic reticulum (Melton-Celsa, 2014). During this process, the α-subunit is asymmetrically cleaved into two subunits: A1 which contains the active site and A2 which is tethered to the β pentamer (Melton-Celsa, 2014). The A1 subunit is released independently into the cytosol from the endoplasmic reticulum (Melton-Celsa, 2014). In the cytosol, the activated A1-subunit then goes on to deactivate 60S ribosomal subunits in the host cell by removing an adenine residue from 28S rRNA by cleaving an N-glycosidic bond, leaving ribosomes unable to elongate peptides (Endo et al., 1988). This disruption of protein synthesis leads to apoptosis (Pacheco & Sperandio, 2012).

Higher levels of Gb_3 expression are associated with higher sensitivity to the toxin (Obrig et al., 1993). The microvasculature of the human renal cortex was found to have unusually high levels of Gb_3 expression (50x higher than the umbilical endothelium) (Obrig et al., 1993), which explains why HUS, which affects the kidneys, is a relatively common side effect of EHEC infection.

1.3 Hemolytic Uremic Syndrome

Hemolytic uremic syndrome (HUS) is a type of kidney failure that can be caused by EHEC infection. HUS is characterized by severe non-immune-mediated hemolytic anemia, which is when blood cells are destroyed more quickly than they are produced due to an abnormally high level of hemolysis (blood cell destruction); thrombotic microangiopathy, the formation of blood clots in small blood vessels; thrombocytopenia, low platelet count (<60,000/mm^3), and kidney injury or failure (Mayer et al., 2012). Around 20% of individuals infected with an epidemic form of EHEC develop HUS, but the risk of developing HUS from sporadic infections of EHEC is much lower (3-9%) (Noris & Remuzzi, 2005). Around 70% of cases of HUS are caused by infection with *E. coli* O157:H7; however, HUS can also be caused by other types of bacterial infections or from genetic mutations in complement regulatory pathways (Mayer et al., 2012). HUS caused by bacteria that produce Shiga toxin is called Stx-HUS whereas HUS that is not caused by Shiga toxin-producing bacteria is called non-Stx-HUS. Sometimes HUS that is caused by genetic mutations is also called aHUS (atypical-HUS).

Stx binds to the Shiga toxin receptor, Gb_3, on the surface of vascular endothelial cells causing them to become thrombogenic (clot producing) through a mechanism that is not fully understood but is thought to be related to activation of the complement system, a component of the innate immune system which is involved in inflammation (Zoja et al., 2010). Binding of the toxin to Gb_3 sets off a signaling pathway that ultimately results in the production of adhesion molecules which increase the adherence of platelets and leukocytes (Zoja et al., 2002). This contributes to the formation of clots in small blood vessels, which can block off blood flow leading to tissue ischemia and sometimes even necrosis or

permanent damage (Mayer et al., 2012). Inflammation of the endothelial cells is also a common symptom in patients with HUS (Mayer et al., 2012).

The kidneys are especially vulnerable to damage because, as mentioned previously, there is a very high level of expression of Gb_3 on the arterioles and capillaries of the renal cortex, the outer portion of the kidney which contains many small blood vessels which provide nutrients for the rest of the kidney (Obrig et al., 1993). However, the Gb_3 receptor is also expressed in central nervous system tissue and 25% of patients with HUS develop neurologic symptoms such as stroke, seizure or coma (Noris & Remuzzi, 2005).

Children under the age of five are at the highest risk of developing HUS after being infected with EHEC (Noris & Remuzzi, 2005). It has long been suggested that the reason children are more prone to HUS is that children might have a higher level of Gb_3 expression (Obrig et al., 1993); however recent evidence has indicated that Gb3 expression level remains constant in humans throughout the lifespan, indicating that other age-related factors are responsible for this phenomenon (Ergonul et al., 2003). HUS due to EHEC infection is the number one cause of acute kidney failure in otherwise healthy children (Mayer et al., 2012).

As mentioned previously, treatment with antibiotics may actually increase a child's risk of developing HUS by up to 300% due to increased Shiga toxin production (Wong et al., 2012). Antimotility agents (drugs that treat diarrhea) may also be associated with an increased risk of developing HUS (Beatty, 2004), which is unfortunate since diarrhea is one of the most common symptoms of EHEC. 55 to 70% of children who develop HUS will experience acute renal failure (as opposed to merely renal impairment) (Tönshoff, 1994; Banatavala, 2001; Noris & Remuzzi, 2005). However, around 70% of children who experience acute kidney failure eventually regain full kidney function (Kelles et al., 1994).

Among patients who develop HUS, 70% require red blood cell transfusions to treat hemolytic anemia and 50% require dialysis to treat renal impairment (Mead & Griffin, 1998; Noris & Remuzzi, 2005). 3-5% of patients with Stx-HUS in industrialized countries die during the acute stage of infection (Noris & Remuzzi, 2005).

1.4 Cattle as a Reservoir

Around 75% of EHEC infections are caused by the consumption of contaminated beef or dairy products (Nguyen & Sperandio, 2012). Cattle are common asymptomatic carriers of EHEC (Nguyen & Sperandio, 2012) and are the main reservoir of EHEC in industrialized countries (Chase-Topping et al., 2007). Cattle do not experience symptoms of EHEC because, unlike humans, they do not express the Shiga toxin receptor globotriaosylceramide (Gb_3) in their gastrointestinal tract (Pruimboom-Brees et al., 2000). The Gb_3 receptor is what causes host cells to internalize Shiga toxins (Stx) (Obrig et al., 1993), leading to necrosis and/or apoptosis (Pruimboom-Brees et al., 2000). Since Stx cannot bind to blood vessels in the colon of cattle due to the lack of Gb_3 receptors, it is not internalized or transported to other organs where it can cause severe damage (Pruimboom-Brees et al., 2000).

In cattle, EHEC colonizes the recto-anal junction (RAJ). In order to colonize the RAJ in cattle and the large intestine in humans, EHEC must first pass through the acidic environment of the stomach. EHEC has three acid resistance (AR) systems (glucose-repressed, glutamate-dependent, and arginine-dependent) which allow it to survive these environments (Tuttle et al., 1999). The bacteria's ability to survive the stomach may be partially responsible for its ability to infect hosts at a relatively low dose (<50 CFUs) (Tilden et al., 1996).

In cattle, bacteria that have colonized the RAJ can be shed in the feces of infected animals (Nguyen & Sperandio, 2012). A small number of cattle called "super-shedders" shed a much higher level of EHEC than other infected cattle (>10^4 CFU/g of feces) (Omisakin et al., 2003). Chase-Topping et al. (2007), found that the chance of an animal being a super-shedder was related more to the sub-group of *E. coli* that the animal was infected with, and the individual host than to the environment of the farm on which the animal lived. For example, they found that super-shedders were twice as likely to carry *E. coli* O157 that had the phage type (PT) 21/28 than non-super-shedders (Chase-Topping et al., 2007). This is especially concerning because PT 21/28 is associated with more severe symptoms in humans (Chase-Topping et al., 2007).

Although super-shedders make up only around 9% of infected cattle, they may be responsible for 96% of the EHEC released into the environment (Omisakin et al., 2003), and are associated with a higher prevalence of EHEC infection in populations of co-penned cattle (Cobbold et al., 2007). Therefore, super-shedders represent a major vector of infection and contamination. EHEC can remain viable in manure for up to 56 days at room temperature (Wang et al., 1996) and can go on to contaminate meat, dairy products, water, or vegetation (Nguyen & Sperandio, 2012) which may eventually be consumed by humans, resulting in infection.

1.5 LEE Pathogenicity Island and the Formation of A/E Lesions

When EHEC reaches the RAJ in cattle or the large intestine in humans, it forms attaching and effacing (A/E) lesions which allow it successfully colonize

the host. The formation of A/E lesions involves tight attachment by the bacteria to the intestinal epithelium, destruction of the microvilli, and accumulation of actin at the site of attachment, forming a pedestal (Hartland & Leong, 2013). The genes that regulate this process are encoded by the locus of enterocyte effacement (LEE), a chromosomal pathogenicity island (Elliott et al., 1998).

The LEE is 35-kb long (Elliott et al., 1998) and encodes approximately 41 genes divided into 5 polycistronic operons (Elliott et al., 1998). These genes encode intimin, an adhesin that binds to Tir (translocated intimin receptor) which is integrated into the host cell membrane by EHEC which helps the bacteria attach to enterocytes (Elliott et al., 1998; Hartland et al., 1999). LEE also encodes effector (Esp) proteins, chaperones (Ces proteins), regulators, and the type III secretion system (T3SS) (Esc, and Sept proteins) (Elliot et al., 2000; Wong et al., 2011; Sperandio et al., 2000).

The type III secretion system, which is encoded by the LEE, is a molecular syringe which many Gram-negative bacteria use to inject effector proteins into the cytoplasm of host cells in order to modify the cells' behavior. A diagram showing the structure of the T3SS is shown in Figure 1. The T3SS effector proteins target several critical processes in the host cell (Shenoy et al., 2018) and plays an essential role in the formation of A/E lesions, which are necessary for successful colonization of the host (Connolly et al., 2015). Thus, the T3SS is an important component of EHEC's virulence program.

The T3SS is embedded within both the inner and outer bacterial membranes and punctures the membrane of affected cells in order to inject virulence factors (Slater et al., 2018). The T3SS is assembled in several steps. First, the membrane rings and export apparatus are assembled. Then, the inner rod and needle are assembled. Finally, the filament and translocon are assembled. After assembly is complete, secretion can begin (Slater et al., 2018). There are at

least 39 effector proteins that are known to be secreted by the T3SS into host cells (Mellies et al., 2007). Most of these effector proteins are encoded either by the LEE pathogenicity island or by cryptic prophages in EHEC's genome (Tobe et al., 2006). The relatively low GC content (38%) of the genes encoding the T3SS compared to the rest of the genome is evidence that the genes were acquired via horizontal gene transfer (Slater et al., 2018).

Ler (LEE-encoded regulator) is the first gene encoded by LEE and is a transcriptional activator of LEE operons 2, 3 and 4 (Elliot et al., 2000). Elliot et al. (2000) found that through its role in regulating the expression of T3SS proteins, Esp proteins, and Tir, Ler is an essential regulator of A/E lesion formation. However, Ler is also involved in the expression of other, non-A/E-related, virulence proteins that are encoded either within or outside of the LEE pathogenicity island, indicating that Ler is a global virulence gene regulator in EHEC (Elliot et al., 2000). Although LEE is essential to A/E lesion formation, it is not sufficient on its own; non-LEE genes are also required (Elliot et al., 1999; Mellies et al., 2007). For example, TccP (Tir-cytoskeleton-coupling protein), a T3SS effector protein which is known to be necessary for the actin accumulation that characterizes A/E lesion formation, is encoded by a cryptic prophage in EHEC's genome (Mellies et al., 2007).

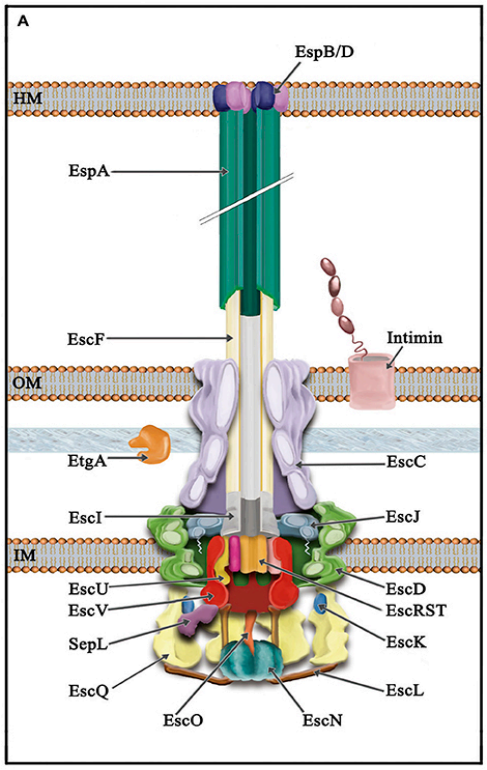

Figure 1. Structure of the Type III Secretion System that EHEC uses to inject LEE-encoded effector proteins into the cytoplasm of host cells (Gaytán et al., 2016). [IM = bacterial inner membrane; OM = bacterial outer membrane; HM = host membrane]

A number of factors play a role in regulating the formation of A/E lesions in addition to Ler. For example, bacteria are most likely to form A/E lesions when they are in the early exponential phase of growth and when they are in an environment that is around 37℃, which is approximately the temperature of the human body (Mellies et al., 2007). In addition to the role of environmental conditions, the transcription of LEE genes is negatively regulated by the LuxR homolog SdiA (Kanamaru et al., 2002). This mechanism is a type of cell density-dependent quorum sensing system which involves a Lux-R homolog transcriptional regulator and an autoinducer and is conserved in many types of Gram-negative bacteria (Kanamaru et al., 2002). SdiA senses the concentration of

acyl-homoserine lactones (AHLs) produced by other bacteria, preventing the bacteria from colonizing areas with large existing bacterial colonies (Kanamaru et al., 2002). Nguyen and Sperandio (2012) suggest that this occurs to prevent EHEC from colonizing the hostile bovine rumen where there is a high AHL concentration and to encourage colonization of the low-AHL RAJ (Nguyen & Sperandio, 2012).

1.6 Host-Bacteria Signaling

The gut contains many different types of small signaling molecules that are produced by the microbiome, the host, or introduced via dietary sources (Kumar et al., 2019). Commensal bacteria can produce oligopeptides that they use to defend their space from other strains of bacteria such as antibiotics or bacteriocins, metabolic compounds such as indole or short-chain fatty acids and quorum sensing signaling molecules (Kumar et al., 2019). Additionally, human hosts produce compounds such as hormones or metabolic by-products (Kumar et al., 2019). Many intestinal pathogens have evolved to be able to detect these small molecules and use information about the concentration of various compounds in the gut to regulate their growth and virulence programs (Kumar et al., 2019). Since expressing virulence genes is often energetically expensive for pathogens, it is important that they only express those genes under conditions in which successful colonization of the host might be possible. In the case of EHEC, the bacteria detect cues that they are in the correct location for A/E lesion formation which would be close to the epithelial surface of the large intestine in humans (Carlson-Banning & Sperandio, 2018).

These environmental signals are usually detected by bacterial sensor kinases (Moreira et al., 2016), and can activate the transcription of Ler, the master

regulator of the LEE pathogenicity island, which activates expression of the rest of the LEE genes, including the T3SS (Mellies et al., 1999; Slater et al., 2018). EHEC does not simply detect one type of small molecule, but a rather a variety that offer different information about the bacterium's location within the gastrointestinal tract (ie. whether it is in the stomach, small intestine, large intestine, etc) as well as the bacterium's proximity to the intestinal wall (Carlson-Banning & Sperandio, 2018). For example, EHEC is capable of detecting bicarbonate which is released from the pancreas into the intestine to neutralize the acidity of the partially digested food which is expelled from the stomach. This is useful to the bacteria because, in the absence of bicarbonate, the glutamate-dependent acid resistance (GDAR) pathway represses LEE expression, so that virulence activities will not begin in the stomach, but when the bacteria move into the pH neutral environment of the intestine, it detects bicarbonate which causes the GDAR pathway to "turn off", allowing expression of LEE once more (Carlson-Banning & Sperandio, 2018). This is only one example of the numerous signaling molecules that EHEC is capable of using to determine the location in the gastrointestinal tract.

One way that EHEC is able to detect proximity to the intestinal epithelium is through sensing oxygen concentration. The inner part of the intestine, the lumen, is mostly anaerobic (Albenberg et al., 2014). However, very near to the surface of the epithelium there is a small amount of oxygen because of diffusion from capillaries in the microvilli, the microscopic finger-like projections on the surface of intestinal epithelial cells which aid in absorption by increasing surface area (Albenberg et al., 2014). Because of this, the presence of oxygen is a cue to EHEC that it is close to the surface of the endothelium. Therefore, in the absence of oxygen, LEE is repressed by the proteins KdpE and FusR; whereas in the presence of oxygen, the proteins Cra (catabolite repressor activator) and KdpE

activate the expression of LEE, resulting in T3SS expression (Carlson-Banning & Sperandio, 2016). Damage to the host's microbiome might also increase the level of oxygen at the surface of the endothelium (Hughes et al., 2017). Therefore, antibiotics that disrupt the microbiome may also indirectly increase virulence signaling through increased oxygen levels. Detecting oxygen might also be beneficial to EHEC since it is capable of performing aerobic respiration, allowing it to outcompete the innate microbiome which is mostly composed of anaerobic bacteria (Lopez et al., 2016).

Another way epithelial proximity is detected is through adrenergic signaling. Host cells produce the neurotransmitters epinephrine (Epi) and norepinephrine (NE) and autoinducers 1 and 2 which can all be detected by bacterial pathogens (Slater et al., 2018; Kumar et al., 2019). Epi and NE form a chemical gradient in the gut with the concentration increasing with proximity to the host epithelium (Kumar et al, 2019). These hormones play a role in regulating blood flow and the absorption of electrolytes and oligopeptides in the intestines (Mittal et al., 2017). Epi and NE are sensed by the bacterial histidine kinase QseC, which is activated through auto-phosphorylation and is conserved across *Enterobacteriaceae* (Kumar et al., 2019). In EHEC, a second histidine kinase, QseE is activated by QseC (Carlson-Banning & Sperandio, 2018). The QseE and QseC pathways are visually summarized in the figure below (Fig 2). In the presence of Epi and/or NE, QseC and QseE phosphorylate the downstream regulators QseF and QseB respectively, and QseC activates the previously mentioned KdpE, which also plays a role in oxygen detection (Jung et al., 2012; Hughes et al., 2009; Reading et al., 2009) (Summarized in Carlson-Banning & Sperandio, 2018). QseC can also be activated by the autoinducer AI-3, which plays a role in EHEC quorum sensing, (Hughes et al., 2009) and QseE can also be activated by sulfate (Reading et al., 2009). QseB which is phosphorylated by QseE goes on to activate

the transcription of the *flhDC* promoter which regulates flagellar activity and thus, bacterial motility (Hughes et al., 2009). QseF activates genes involved in A/E lesion formation (Reading et al., 2007), and is involved in inducing entry to the lytic cycle by the Shiga toxin-producing prophages in EHEC's genome (Hughes et al., 2009). As was the case with oxygen-detection, KdpE and Cra activate expression of LEE (Carlson-Banning & Sperandio, 2016; Njoroge et al., 2012). This adrenergic pathway results in increased T3SS activity, A/E lesion formation, and more severe symptoms in the host (Kumar et al., 2019). One study which used *Citrobacterium rodentium*, a model bacteria that is very similar to EHEC, showed that *C. rodentiu*m had a diminished ability to colonize the gut of mice that did not secrete NE or Epi and that strains of C. rodentium which did not have functional *qseC* and/or *qseE* were not as effective as the wild type strain at colonizing the intestine of mice which did express Epi and NE (Moreira et al., 2016).

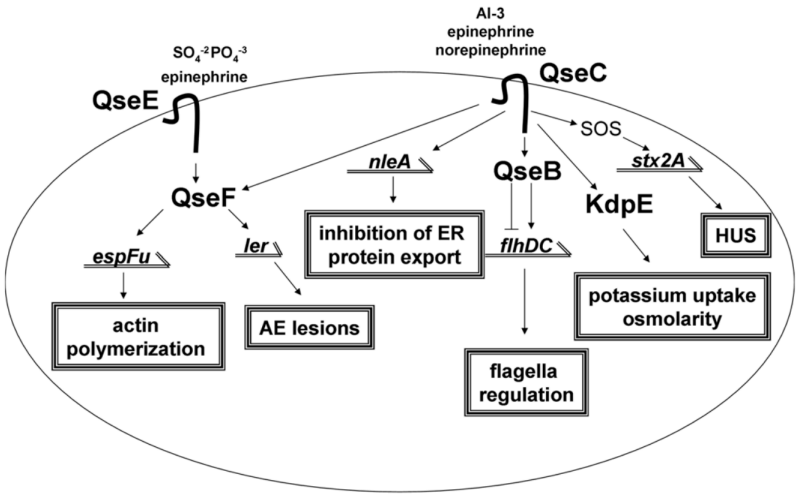

Figure 2. Activation of virulence genes through the QseE/QseC signaling cascade in response to epinephrine and/or norepinephrine in EHEC (Hughes et al., 2009).

These results demonstrate the importance of Epi/NE detection in regulating virulence in EHEC. Due to the role of these pathways play in

regulating virulence, disrupting adrenergic signaling represents a potential alternative to antibiotics for reducing the symptoms and colonization of certain gastrointestinal bacteria. This is potentially significant, not just in EHEC, but also in other enteropathogenic bacteria that are not effectively treated by antibiotics, such as Salmonella. In general, it is worthwhile to invest research effort into alternatives to antibiotics, sometimes called anti-infectives, as a strategy to combat antibiotic-resistant bacteria.

1.7 Cannabinoid Signaling Pathways

Endocannabinoids are lipid-based neurotransmitters that are endogenously produced by mammals, including humans, which bind to the cannabinoid receptors that are located on cells throughout the central and peripheral nervous system as well as other tissues and organs in the body. In humans, the main components involved in the endocannabinoid system are the endocannabinoids anandamide and 2-arachidonyl glycerol (2-AG), their respective inactivating enzymes, fatty acid amide hydrolase (FAAH) and monoacylglycerol lipase, and the two 7-pass transmembrane G protein-coupled receptors CB1 and CB2 (Camilleri et al., 2013). Several other endocannabinoids that have been identified, but currently, there is relatively little research on these compounds (Lee et al., 2016). As the name "endocannabinoid" suggests, these compounds behave similarly to the phytocannabinoids, tetrahydrocannabinol (THC) and cannabidiol (CBD), which are derived from the cannabis plant. The functions of the endocannabinoid system in the body is an area of active research; however, it is thought to be involved in appetite and metabolism (Mariscano et al., 2009), the suppression of neuroinflammation (Ceballos, 2015), immune function, neurogenesis, learning, and mood among many other

processes (Ronan et al., 2016). Dysfunction of the endocannabinoid system is thought to be linked to anorexia (Mariscano et al., 2009) and certain neurodegenerative diseases such as Huntington's or Parkinson's (Concannen et al., 2015).

Among its many potential functions, the endocannabinoid system is known to be involved in the regulation of gastrointestinal function (DiPatrizio, 2016). Endocannabinoid signaling plays a role in nausea, vomiting, gastric acid secretion, colonic motility, ion transport, gut inflammation, and gastroprotection (Izzo, 2010). It also plays a key role in regulating neurotransmission in the neurons of the GI tract, known collectively as the enteric nervous system (ENS) (Trautmann, 2015). Activation of cannabinoid receptors in the intestines leads to reduced peristalsis (contractive movement of the intestines) by inhibiting the excitatory signaling of cholinergic motor neurons (Camilleri, 2008), reduced acid secretion, and increased hunger (DiPatrizio & Piomelli, 2015). Due to the importance of the endocannabinoid system to gastrointestinal function, dysregulation of this system is thought to be involved in inflammatory bowel disease, a family of diseases that involve chronic intestinal inflammation, irritable bowel system, a disease characterized by abnormal intestinal contractions, abdominal pain, and diarrhea, and obesity (DiPatrizio, 2016). This idea is supported by the fact that single nucleotide polymorphisms in the genes that encode for FAAH, the enzyme that degrades the endocannabinoid anandamide, and the cannabinoid receptor CB1 are associated with irritable bowel syndrome and increased colon motility (Camilleri, 2008; Camilleri et al., 2013).

Given the importance of endocannabinoid signaling in the gastrointestinal tract and the documented ability of EHEC to detect other small molecules in the gut, such as epinephrine/norepinephrine, it has been hypothesized that EHEC

may be able to detect the presence of endocannabinoids. Sperandio et al investigated this hypothesis in their study on the role of the endocannabinoid 2-AG on the pathogenicity of *Citrobacter rodentium*, a previously mentioned model pathogen that is similar to EHEC. They found that mice in which the enzyme that inactivates 2-AG, monoacylglycerol lipase, had been inactivated experienced lower pathogen burdens after infection with *C. rodentium* than controls (Ellerman & Sperandio, Under Review). This indicates that higher levels of active 2-AG decreased the pathogen's ability to colonize the gut of mice. Sperandio et al. also treated EHEC with 2-AG and analyzed resulting changes in gene expression using RNA-Seq. They found that the expression of LEE-encoded genes decreased after exposure to 2-AG (Ellerman & Sperandio, Under Review). They confirmed these results using qPCR. Notably, expression of the genes encoding Shiga toxin did not increase after exposure to 2-AG (Ellerman & Sperandio, Under Review). Western blot analysis showed that treatment with 2-AG lowered EHEC's level of EspB, EspA, and Tir, previously mentioned proteins that play important roles in adhesion and type three secretion (Ellerman & Sperandio, Under Review)They also found that 2-AG decreased the expression of key pathogenicity genes in *Salmonella enterica* serovar *Typhimurium*, indicating that this pathway is conserved in other pathogens, and demonstrating the potential utility of endocannabinoids as a treatment for multiple types of enteric pathogens. They found that the downregulation of virulence in EHEC was mediated by the inhibition of QseC activity. In a *ΔqseC* deletion strain, 2-AG had no effect on the expression of LEE genes (Ellerman & Sperandio, Under Review). Using an auto-phosphorylation assay with wild type EHEC, they confirmed that QseC activity was lowered after the bacteria had been exposed to 2-AG. Additionally, they showed that when EHEC was treated with 2-AG in the presence of a CB1 antagonist, a compound which prevents 2-AG from binding to CB1 receptors, 2-

AG treatment no longer resulted in decreased virulence, indicating that downstream signaling from 2-AG's binding to CB1 was responsible for the inhibition of LEE expression (Ellerman & Sperandio, Under Review)Finally, they showed that the FadL transporter is required to transport 2-AG across the outer membrane so that it can interact with QseC which is located on the inner bacterial membrane (Ellerman & Sperandio, Under Review)

Based on Sperandio et al.'s research on 2-AG, Brandon Novy (Reed College Class of '19) decided to investigate cannabidiol (CBD), a phytocannabinoid that is produced by the cannabis plant, as a potential treatment for EHEC. Although over 100 types of phytocannabinoids have been isolated from the cannabis plant, the most prevalent bioactive compounds are CBD and tetrahydrocannabinol (THC) (Aizpurua-Olaizola et al., 2016). Unlike THC which is known to alter mood and perception, CBD is non-psychotropic and is considered non-addictive by the World Health Organization (WHO, 2017). Furthermore, CBD is considered low-risk and safe for recreational use (WHO, 2017) The 2018 Farm Act made the sale of CBD with less than 0.3% THC legal nationwide (USDA, 2018). However, the growth of cannabis and sale and use of CBD are still legally complicated given that cannabis plants that contain more than 0.3% THC are still considered a Schedule I controlled substance by the federal government (DEA). Currently, the only FDA-approved medical use of CBD is in the form of the drug "Epidiolex" which is used to treat seizures in children with Dravet or Lennox-Gastaut syndrome (FDA.gov). However, research indicates that CBD may also have utility in treating other conditions such as alcohol use disorders (Karoly et al., 2019), inflammation (Costiniuk et al., 2019), anxiety (Skelley et al., 2003), pain (Good et al., 2019), and schizophrenia (McGuire, et al., 2018).

In the body, CBD acts as a non-competitive negative allosteric modulator of the CB1 receptor, which means that it decreases the responsiveness of the receptor to agonists either by preventing CB1-agonists from binding or inhibiting the response (LaPrarie et al., 2015). Therefore, CBD may actually inhibit some of the effects of THC, which is a direct CB1 agonist (LaPrairie et al., 2015). Additionally, CBD has been found to antagonize the CB2 receptor (Pertwee, 2008). However, the exact nature of this interaction remains unclear. There is some evidence that CBD may interact with other receptors as well, including opioid and adrenergic receptors (Bih et al., 2015).

Given the fact that CBD is considered non-addictive and low-risk, and is now widely commercially available, CBD is an exciting prospective treatment for EHEC that would be relatively accessible and affordable for most people. Brandon Novy hypothesized that CBD would inhibit the expression of virulence genes in EHEC and that this effect would be mediated through interaction with QseC, as was the case with 2-AG (Ellerman/Sperandio, Unpublished) (Novy, 2019). While Novy found that CBD was effective at reducing intestinal colonization by *C. rodentium* in mice, this process turned out to be independent of QseC (Novy, 2019). This leaves the question: How does CBD inhibit virulence in EHEC? In order to answer this question, computational methods will be applied to analyze RNA-Seq data from EHEC treated with CBD and control samples to elucidate which signaling pathways control EHEC's response to CBD.

1.8 RNA-Seq

RNA-Sequencing (RNA-Seq) is a technique that quantifies the amounts of different RNA transcripts in a biological sample using next-generation sequencing. This information is like a snapshot of the transcriptome (all the

RNA in a cell) in a certain tissue at a single point in time. This data can be used to analyze the expression of different genes and can be used to compare gene expression under different conditions or in different tissue types. RNA-Seq data can also be applied to examine post-transcriptional modifications and alternative splicing, however, this is beyond the scope of this thesis. The process of isolating, sequencing and analyzing RNAseq data is represented in the simplified figure below (Fig 3).

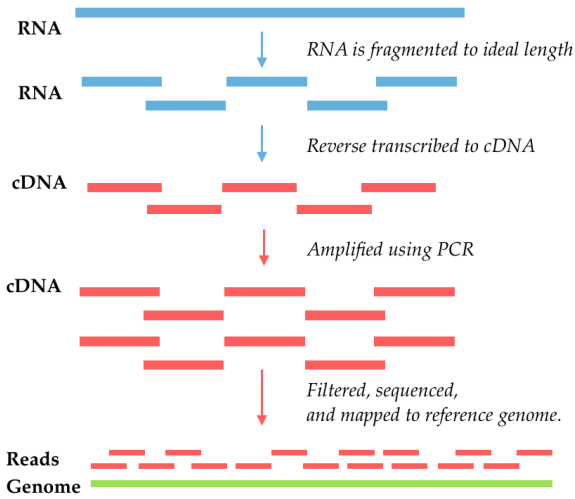

Figure 3. Overview of the process of isolating, preparing, sequencing, and mapping RNA from a cell to a reference genome.

Before sequencing, the RNA must first be isolated from the tissue. After the RNA is isolated, the ribosomal RNA (rRNA) a type of non-coding rRNA that is a component in ribosomes, is generally depleted since rRNA makes up the majority of the RNA in a cell. After this, the remaining RNA is then fragmented in order to obtain sequences that fall within the preferred length range of the sequencing machine. The fragmented RNA is reverse transcribed to cDNA (complementary DNA) since DNA is more stable and easier to work with. The cDNA is then amplified. Finally, there is a size selection process to remove very small pieces of RNA such as miRNA (micro-RNA).

The selected cDNA library is then sequenced using a next-generation sequencing method. In next-generation sequencing, the strands to be sequenced are separated and fluorescent-tagged nucleotides are added to each to form the complementary strand. This method produces a number of complementary fragments of different sizes with nucleotides that can be identified by their fluorescent tag. The information from the fluorescent-tagged nucleotides on strands of different sizes can be computationally combined to obtain the full sequence of the strand. Next-generation sequencing can sequence millions of fragments simultaneously, unlike the original sequencing method, Sanger sequencing, which can only sequence a single DNA fragment at a time. Since next-generation sequencing is a high throughput method, it significantly increases the speed at which sequencing can occur which has made RNA-sequencing possible.

After the fragments have all been sequenced, the sequence information is compiled in a FastQ file. The reads must then be processed and analyzed. Often the reads will first be "cleaned up" using a tool that trims low-quality data or removes adapter sequences. After the reads, have been trimmed, they must be mapped to a genome. This can be done in one of two ways. If the genome has not been fully sequenced, then it must be constructed using *de novo* assembly from the RNA-Seq data. More information on *de novo* assembly can be found in section 5.0 of the Appendix. If the genome has already been sequenced, no assembly is required.

Once a reference genome has been obtained, reads from the FastQ file are mapped to the reference genome. There are multiple programs available that will take a reference genome in the form of a FASTA file and the trimmed reads in the form of a FASTA or FastQ file and output a SAM or BAM alignments file that contains the location in the genome of all of the reads. The file containing

locations along with a gff (gene feature file) file containing genome location to gene name information can be used to determine how many reads correspond to each gene in the genome. In situations where the reference genome is not annotated, counts can be determined based solely on genome location rather than corresponding to a gene.

After counts have been determined for each locus, the counts can be normalized in various ways. This is done to control for certain variables which may affect the number of times a transcript corresponding to a gene occurs. The first step is often to normalize the counts based on sequencing depth (ie the total number of fragments that were sequenced). To do this the counts for each gene are converted into fragments per million mapped reads (FPM). After controlling for sequencing depth, additional normalizing measures can be incorporated as well. For example, longer genes will have more corresponding reads than shorter genes at the same expression level. To counter this effect, counts can be transformed into FPKM (Fragments Per Kilobase Million) which is the FPM of the gene divided by the length in kilobases of the gene. A similar measurement is RPKM (Reads Per Kilobase Million). The difference between the two is that FPKM is used for paired-end RNA-Seq data, in which fragments may be read twice, whereas RPKM is used for single-end RNA-Seq data. FPKM ignores duplicates of the same read. A different way to normalize RNA-Seq data is known as TPM (Transcripts Per Million) which, as the name suggests, represents how many times a fragment corresponding to a certain gene would appear out of a million. The process to obtain the TPM is similar to that of obtaining the RPKM or FPKM, only in reverse. First, the fragments per kilobase (FPK) is calculated for each gene by dividing the number of fragments that map to that gene by the length of the gene. Then, a scaling factor is determined by the sum of the FPK for

all genes divided by one million. The TPM of each gene is its FPK divided by the scaling factor.

Once the counts for each gene have been normalized, this data can be used for a number of analyses. For example, RNA-Seq data from different conditions can be compared to determine condition-based changes in the expression of various genes. There are many programs that are capable of analyzing RNA-Seq data to determine differential expression. Some commonly used tools include Cufflinks (Roberts et al., 2011), EdgeR (Robinson et al., 2010), Limma (Law et al., 2018), and DeSeq2 (Love et al., 2019). DeSeq2 was chosen for this project based on its compatibility with other software in the RNA-Seq pipeline. DeSeq2 takes multiple samples from each condition and tests for differential gene expression using a negative binomial distribution model (Love et al., 2016). DeSeq2 outputs text files containing the normalized gene counts for each sample and a differential expression results file as a tab-delineated spreadsheet containing a line for each gene which lists the gene name, base mean expression, the log2 fold change, the log fold change standard error, wald-statistic (log fold change/standard error), the p-value for the significance of the expression change, and the adjusted p-value. This information can be used to assess which genes experience the greatest changes in expression level under different conditions or to predict co-expression networks.

1.9 Protein-Protein Interaction Networks

In graph theory, a graph (G = V, E) is a network composed of vertices and edges where V is the set of all vertices (or nodes) which represents all of the elements in a graph and E is the set edges which represent the connections between any two elements (or an element and itself). Graphs can be directed,

meaning that the edges indicate the direction of interactions, which are usually represented as arrows, or undirected in which case the direction of the interaction is not specified and the edge is represented as a simple line. An example of an undirected graph is shown in Figure 4.

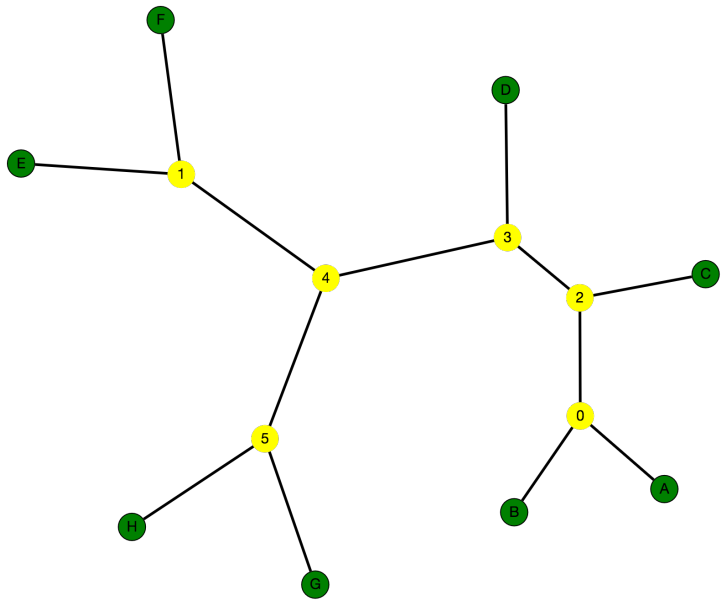

Figure 4. An example of a simple undirected graph (G = V, E). The vertices are all the points and the edges are the lines between them.

Graphs are useful tools which can be used to represent social networks, population dynamics, gene regulatory networks, food webs, and other biological and social networks. Gene expression information from RNA-Seq data can be used to create graphs called gene co-expression networks. Co-expression networks are undirected graphs in which vertices represent genes and edges between vertices represent a high level of co-expression between the two genes. Co-expression networks are constructed by searching for pairs of genes that show similar expression patterns across multiple samples. Genes that show similar expression patterns are often controlled by the same signaling pathway or transcription factor or else interact in some other way, such as forming

members of the same protein complex (Weirauch, 2011). This is sometimes known as the "guilt-by-association principle" (Grimes et al., 2019). Therefore, co-expression networks can be a valuable tool to predict interactions. The edges in a gene co-expression network do not necessarily represent a direct interaction between two genes -- only that the expression levels of those genes are correlated. Patterns within the graph can be used to make inferences and develop hypotheses that can be tested *in vitro* (Roy et al., 2013). For example, highly connected sub-graphs (a subset of all the nodes and edges in the graph in which many of the nodes are connected to each other) often indicate a group of genes which are controlled by a shared transcription factor or which are involved in a similar biological process (Roy et al., 2013).

In order to construct co-expression networks, several biological tests must first be performed in which the expression levels of all or a subset of genes are measured using RNA-Seq or micro-arrays for several samples or across different conditions (Weirauch, 2010). Then, the expression data can be used to calculate a co-expression measure between every pair of genes. There are many different ways to calculate co-expression between a pair of genes, including Euclidean distance, which measures the geometric distance between two gene vectors; Mutual information, which measures the mutual dependence between two variables; and Spearman's rank correlation coefficient, in which genes are ranked from highest to lowest based on their expression level and then the correlation between changes in ranks across samples is calculated for each pair of genes (Weirauch, 2010). One of the most commonly used measures is Pearson's correlation coefficient, which is a measure of the linear correlation between two variables (Weirauch, 2010). Since the changes in expression of each gene are relative to the average expression level of that gene it is useful for detecting synchronicity in trends of genes increasing or decreasing their expression levels

(Weirauch, 2010). Pearson's correlation coefficient (*r*) ranges from -1, indicating a perfect negative linear correlation, to +1, indicating a perfect positive linear correlation. A score of 0 indicates no correlation. The correlation indicates both the strength and the direction of a linear relationship, but does not indicate the slope of the line, nor is it able to capture other non-linear relationships between variables. The correlation is calculated for two variables, x and y, according to the following equation:

$$r(x,y) = \frac{\Sigma_{j=1}^{n}(x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\Sigma_{j=1}^{n}(x_j - \bar{x})^2 \Sigma_{j=1}^{n}(y_j - \bar{y})^2}}$$

Where n is the sample size, x_j and y_j are individual sample values, and $\bar{x}$ and $\bar{y}$ are the sample means for genes x and y.

Once a correlation measure has been calculated for every pair of genes, a scoring threshold is chosen to identify pairs of genes that are considered to have a high degree of co-expression. There are many ways to calculate a scoring threshold, but the simplest method is to pick a minimum Pearson's correlation coefficient (ie 0.8) and choose the pairs of genes which have an absolute value correlation measure that is equal to or greater than the threshold (Weirauch, 2010). Finally, a graph is constructed, in which all of the genes (or a chosen subset) are represented as nodes, and significant correlations are represented as edges between nodes (Weirauch, 2010). This entire process is illustrated in the following figure (Figure 5).

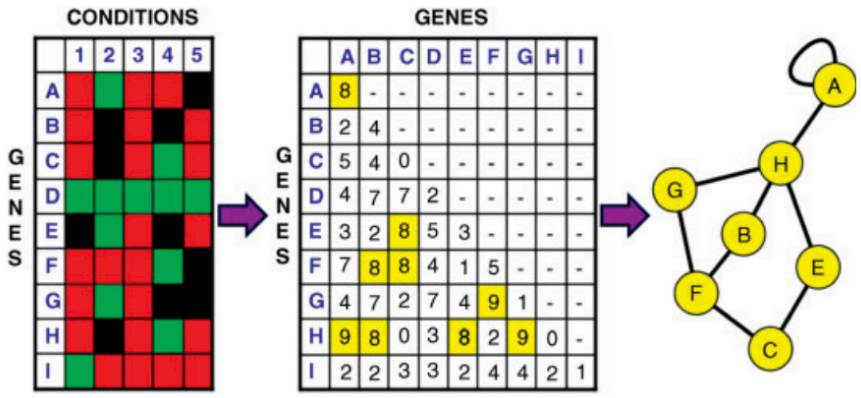

Figure 5. An illustration of the process of constructing a co-expression network from gene expression data (Weirauch, 2010).
In this example, the expression levels (represented as color) of genes A-I is measured across 5 different conditions. Then the expression levels from these experiments are used to create a co-expression matrix which is transformed into an undirected graph.

Interactomes can also be constructed using various experimental methods to identify interactions between proteins. One technique that is commonly used to identify a large number of protein interactions is the yeast two-hybrid system in which two proteins which are required to bind in order to activate transcription of a reporter gene are separately bound to either a “bait” or “prey” protein. If the bait and prey proteins interact, activation of transcription of the reporter gene can occur. The technique can be repeated with large numbers of different bait and prey protein combinations to identify hundreds to thousands of different protein-protein interactions. This information can then be used to create an interactome. Once an interactome has been constructed by one or more of these methods, it can be analyzed to identify genes or sets of genes that are involved in various biological processes.

1.10 Goals of this Thesis

In this thesis, information from RNA-Seq data and an *E. coli* interactome will be analyzed to identify genes and signaling pathways that may be involved in the inhibition of virulence signaling by CBD in EHEC. This is an important step toward understanding how this potential treatment interacts with its target, and would add to our knowledge base on the treatment of enteric pathogens in general.

Methods

2.1 RNA-Seq Analysis

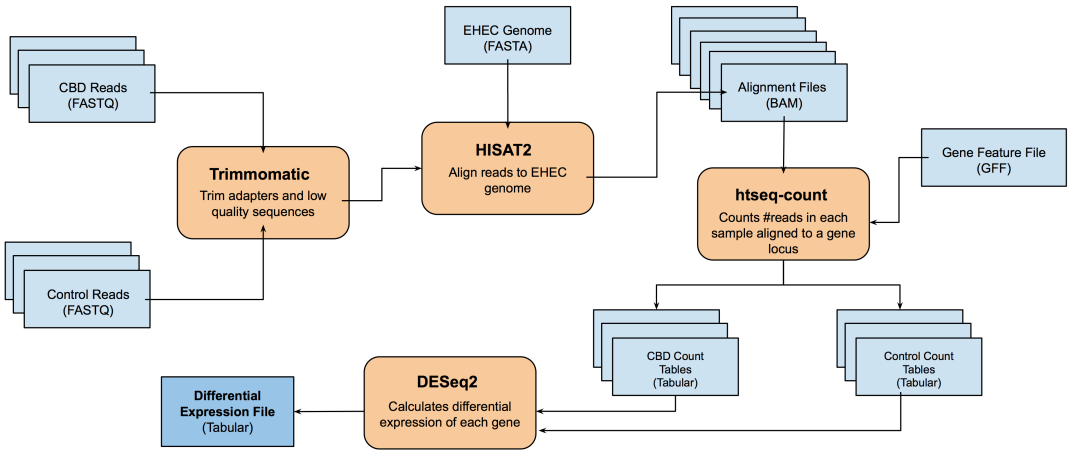

Figure 6. RNA-Seq analysis methods.
Orange boxes represent the programs used for the analysis and blue boxes represent the files that were either input or output during different steps in the process.

Six RNA-Seq samples (3 control samples and 3 CBD samples) were prepared by Brandon Novy (’19) and sequenced by the Center for Genome Research and Biocomputing at Oregon State University using an Illumina sequencer. The sequencing results were sent to us in the form of a FASTA file for each sample, containing the raw sequences of every read that was successfully sequenced from that sample.

After the raw data was received, the reads had to be trimmed, mapped, and quantified so that the data could be analyzed. An overview of this process is shown above in Figure 7. The first step in converting this raw data into a form that could be analyzed was to trim to adapters, which are short artificial

sequences of cDNA that are ligated to reads before they are sequenced, and low-quality data. This was done using the program *Trimmomatic* (Bolger et al., 2014) using the "SLIDINGWINDOW" setting, which clips a read once the average base quality of 4 bases falls below a certain threshold. After the reads were trimmed and cleaned up, each read had to be aligned to the EHEC genome, to determine which gene locus they corresponded to. This was done using the program *HISAT2* (Kim et al., 2015) using the default settings (single-end and unstranded). The inputs to *HISAT2* were the FastQ files containing the trimmed reads, and a FASTA file containing the complete sequenced genome of E. coli O157, including the chromosome the pO157 plasmid, and the pOSAKI plasmid. These sequences were obtained from *GenBank* (Hayashi et al., 2018).

After the reads had been mapped to the reference genome, the location of each read was output as a BAM (Binary Alignment Map) file which contained the sequence of each read and the indices of the sequence within the reference genome. These reads were then quantified using the program *htseq-count* (Anders et al., 2015), which counts the number of aligned reads that correspond to a specific gene locus. As input the program took a BAM file from *HISAT2* for each sample and a GFF (gene feature file) which contains an ID and genome index for each gene locus in *E. coli O157*. The GFF file was obtained from the *MicroScope* database. The output from *htseq-count* is a tabular file containing the gene ID and how many reads corresponded to that locus. Finally, the count files from all 6 samples (3 CBD and 3 Control) were input into *DESeq2* (Love et al., 2014), a program that calculates differential gene expression under different conditions. *DESeq2* was run using the default settings with a parametric fit type. The output of *DESeq2* was a text file containing for every gene an ID, its base mean expression level, the log2 fold change between conditions, the standard error, the wald-statistic (log fold change/standard error), the p-value, and the

adjusted p-value. All of these steps were done using the *Galaxy* web platform (Afgan et al., 2018). More detailed information on the concepts behind RNA-Seq analysis is available in section 1.8 of the Introduction.

2.2 Computational Methods

2.2.1 Name Mapping

In order to obtain the data in a form that could be easily understood, the locus IDs were converted into the common names. To do this, data on the common name, locus ID, protein description, and chromosome location for E. coli O157's genes were obtained from *NCBI* by searching for "txid386585[Organism:noexp]" under the Gene category and downloading all 5485 genes as a text file.

Then using a simple python program, the corresponding common name and protein description for each gene was added to the *DESeq2* output file which contained expression data. The text file containing gene IDs, expression data, common names, and protein descriptions was converted to an Excel file for ease of use. Unfortunately, not all genes in the *NCBI* database had common names available. Therefore, only 911 of the 5485 genes (16.6%) of the genes had common names at the end of the automated naming process. The names of 19 additional genes were identified by manually searching for them on the *Uniprot* and *KEGG* databases and by searching the literature. Ultimately 930 (17.0% of total) genes had been matched to their common name.

2.2.2 Gene Set Enrichment Analysis

A fast gene set enrichment analysis was performed using the *fgsea* package (Sergushichev, 2016) on *Galaxy* (Afgan et al., 2018) to identify the top up-regulated and down-regulated pathways. Of the 930 genes for which common names were known, 481 of these genes were differentially regulated with an adjusted p-value <0.001. These 481 genes were put into a text file along with their log2 fold change and were ranked from most upregulated to most downregulated. Information about the genes involved in known *E. coli* O157 pathways was obtained from the *BioCyc* database (Karp et al., 2019). A simple python script was used to reformat the pathway information from *BioCyc* into a "GMT" (Gene Matrix Transposed) file format in which each pathway is represented as a line with the pathway name in the first column, information about the pathway in the second column, and a list of the common names of all genes known to be involved in the pathway in the third column. Ultimately, this file contained information about 240 *E. coli* O157 pathways which were listed in order of the number of genes in the set from smallest to largest. The ranked list of genes from the RNA-Seq data and the GMT pathway file were input into *fgsea*, a program that calculates enrichment scores for each pathway based on the expression of the genes in that pathway (Sergushichev, 2016).

Fgsea takes as input an array of N genes, U, ranked using a statistic, S, for each gene which describes the expression change from most positive to most negative, and a list of M number of gene sets which correspond to pathways, $P = \{p_1, p_2 \ldots p_M\}$. Then each gene set is assigned an enrichment score, $s_r(p)$, which is calculated using the following procedure, as described in (Sergushichev, 2016).

First, an array is created containing enrichment score (ES) values for every gene in the total set of genes. The enrichment score is assigned according to the following formulas:

$$ES_i = \begin{cases} 0 & \text{if } i = 0, \\ ES_{i-1} + \frac{1}{NS}|S_i| & \text{if } 1 \geq i \geq N \text{ and } i \in p, \\ ES_{i-1} - \frac{1}{N-k} & \text{if } 1 \geq i \geq N \text{ and } i \notin p, \end{cases}$$

Where ES_i is the enrichment score of every gene, i is a gene in the set of all genes, p is the set of genes in a particular pathway, k is the number of genes in the pathway, N is the number of total genes for which there is information, S_i is the expression value for gene i, and NS is the sum of the absolute value of expression values (S) for every gene in the pathway p.

After an enrichment score has been assigned for every gene, the enrichment score for the pathway as a whole ($s_r(p)$) is assigned as the ES value from the array that was just constructed with the largest absolute value ie:

$$s_r(p) = \mathrm{ES}_{i^*}, \text{ where } i^* = \arg\max_i |\mathrm{ES}_i|$$

If the ES score with the largest absolute value is negative, then the $s_r(p)$ value is negative; if the largest ES score is positive then the $s_r(p)$ value is positive. Gene sets that have negative enrichment scores are enriched with downregulated genes, whereas those with positive scores are enriched with upregulated genes. To calculate the p-value for each pathway, a null distribution of ES scores for pathways with the same number of genes as the pathway in question is constructed using randomly selected sets of genes. A one-tailed p-value is then calculated using the null distribution. Unlike the original GSEA algorithm, *fgsea* calculates one null distribution for all pathways that have the same number of genes, making the process more efficient (Sergushichev, 2016).

2.2.3 Constructing and Analyzing an *E. coli* Interactome

Due to a lack of information about the *E. coli* O157 interactome, an interactome using data from *E. coli* K12 was constructed instead. This interactome relied on data from two studies (Hu et al, 2009) and (Rajagopala et al., 2014) that determined protein-protein interactions based on a literature review, experimental results using yeast-2-hybrid systems, and machine learning predictions (some of which were experimentally validated). In total, these studies gave information about 9587 protein-protein interactions involving 2606 unique proteins.

In order to examine genes of interest more closely, a subinteractome was constructed from the *E. coli* K12 interactome that contained only the nodes that were identified as candidates for qPCR (*ler*, *tir*, *espA*, *fadL*, *fadR*, *degP*, *yhaO*, *sdhC*, and *ftnB*) the nodes that were directly connected to candidates (neighbor nodes) and the nodes that were connected to the neighbor nodes. More information on why those specific genes were chosen as qPCR candidates is available in section 3.4 of the results. Unfortunately, several of the 8 candidate genes were not present in the *E. coli* K12 interactome because they are specific to the pathovar O157 and not found in the K12 genome. Therefore, the only candidate genes in the subinteractome were: *degP*, *fadL*, *fadR*, *sdhC*, and *ftnB*. In total, the subinteractome contained 261 edges and 223 nodes, of which 5 were candidates, 11 were neighbors of candidates, and 207 were neighbors of neighbors. Gene expression information was mapped onto the subinteractome by pairing the foldchange values from the RNASeq results with the common names of genes. Unfortunately, since *E. coli* K12 genes and *E. coli* O157 genes have different gene IDs, expression information could only be obtained for genes for which the *E. coli* O157 common names were known and were the same as the K12 common

names. As mentioned previously, less than 1/5 of the O157 genes for which gene expression was obtained were able to be mapped to common names. Out of the 223 nodes in the subinteractome, gene expression information was only able to be obtained for 66.

A second gene set enrichment analysis was run using the same methods as described above in section 2.2.2, but only including the 66 genes in the subinteractome for which expression value was available, in order to obtain information about differentially expressed pathways which were directly related to the 5 candidate nodes in the subinteractome.

In addition to the *fgsea* analysis, a cluster of genes in the subinteractome was chosen for further investigation based on the fact that a large portion of the genes were upregulated, and a higher percentage of genes in the cluster had expression value information available than other clusters in the graph. This cluster was investigated qualitatively by researching the function of each gene using information from the *Uniprot* database (The Uniprot Consortium, 2019).

2.3 Experimental Methods

2.3.1 Testing Growth

EHEC was quadrant streaked onto a lysogeny broth plate and incubated overnight at 37℃. A colony from the plate was then inoculated into 3 mL of liquid lysogeny broth and then incubated overnight at 37℃. The following day, the optical density at 600 nm of 10 μL of the liquid culture diluted in 1 mL of low glucose-DMEM was obtained. Then, 30 μL of the liquid overnight culture was added to six tubes which all contained 3 mL of pre-warmed LG-DMEM to achieve a 1/100 dilution. Three of the six tubes also contained 58 μL of 50 μM

Enjoi CBD solution. All of the tubes were incubated without shaking for 5 hours at 37℃. The cultures were then removed, and the optical densities at 600 nm were obtained using a UV spectrophotometer which had been blanked with plain LG-DMEM.

2.3.2 Testing qPCR Primer Efficacy

Primers for qPCR were designed using NCBI's Primer-BLAST tool for 5 of the 9 chosen candidate genes. The sequences for those primers are shown in Table 1 below. 10 μM primer stocks were prepared for all of the primers by suspending each in 10x the nmol of primer μl of TE buffer.

Table 1. Primer sequences designed for use in qPCR

Gene	Primer Sequences	TM (℃)	GC	Replicon Length (bp)
fadL	Forward: CGTCTCCATCTGGTCGTAGC Reverse: AATAGAAGCGCCCCAACCAA	59.97 59.96	60.0% 50.0%	113
fadR	Forward: AGTGCTGGCTACCGCTAATG Reverse: TGCGCGTATACAGTCCCTTC	60.18 59.90	55.0% 55.0%	146
yhaO	Forward: GCCGCACAGTTTATTAGCGG Reverse: TACGATCCCTTGCGTTGCTT	59.97 60.04	55.0% 50.0%	132
ftnB	Forward: GTGCGAAACTTGCGGGTATG Reverse: AGCTGATGCGACACGACATT	60.18 60.39	55.0% 50.0%	70
sdhC	Forward: GCGATAGCGTCCATTCTCCA Reverse: GCTCGAAACCTTCAGGGGAA	59.97 59.96	55.0% 55.0%	113

Five 50 μl PCR reaction tubes were set up for the five primer pairs I designed. First, a PCR master mix was created from 25.0 μl of 10x Tris buffer, 5.0 μl of 10 mM dNTPs, 1.25 μl of Taq polymerase, and 168.8 μl of nuclease-free water. Then 40 μl of the master mix was added to each of the five PCR tubes. Then 1 μl of the respective forward and reverse primers were added to each tube. In the BSL cabinet, 8 μl of a liquid culture of EHEC which had been incubated overnight in the shaking incubator at 37℃ was added to each tube.

Then, the tubes were placed in an MJ Mini Thermocycler, and the following program was run:

95℃ for 5 minutes (denaturing)

[30 cycles]

95℃ for 30 seconds (denaturing)

55℃ for 30 seconds (annealing)

72℃ for 30 seconds (extending)

72℃ for 2 minutes (final extend)

12.5 µl of each of the PCR products was mixed with 2.5 µl of gel loading mix and then loaded into a 0.8% agarose gel which had been stained with 2 µl of SybrSafe DNA dye. 11 µl of an Invitrogen 100 bp ladder was loaded into the left-most lane. The gel was then run at 110 V for around 20 minutes. The gel was then imaged using the UV setting on a gel imager.

Results

3.1 Differential Expression Analysis

To determine the overall gene expression similarity of the six samples (3 CBD and 3 control) to each other, the following figure was made (Fig 7) which depicts the Euclidean distance between all pairs of samples based on the $\log_2$-transformed expression level of every gene in each sample. To calculate the distance between samples, each sample is represented as a point in multidimensional space, with the expression level of each gene as a dimension and the $\log_2$-transformed expression of that gene as the value of that point for that dimension. Although difficult to visualize, the distance between two points in multidimensional space can be calculated using the following formula:

$$d(q,p) = \sqrt{\Sigma_{i=1}^{n}(q_i - p_i)^2}$$

Where d is the distance between two points, q & p, n is the number of dimensions, and q_i and p_i represent the expression value for the respective points for gene i.

DESeq2 automatically generates a heat map of Euclidean distances using a matrix file where each row is a gene and each column is the $\log_2$-transformed values for a sample (Fig 7).

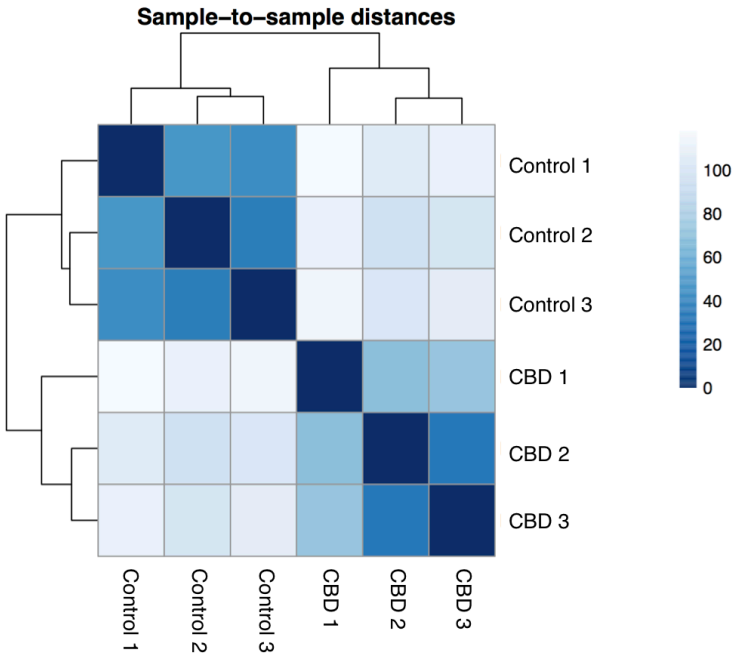

Figure 7. Heat map showing Euclidean distance between samples using log2 transformed counts of all genes.

The darker the color, the more similar two samples are to each other. The line of diagonal dark boxes represents each sample compared to itself which gives a distance of 0. The graph is symmetrical across this axis with each pair of samples (other than a sample by itself) being represented twice.

As can be seen in Figure 7, all three CBD samples were more similar to each other than they were to any of the control samples and all of the control samples were more similar to each other than any of the CBD samples. This is clear from the fact that all pairs of controls vs other controls and CBD samples vs other samples have darker boxes (less distance) than the controls vs. CBD pairs, indicating that the pairs within the groups have more similar gene expression profiles than pairs between groups. This indicates that treatment with CBD has a somewhat consistent impact on gene expression.

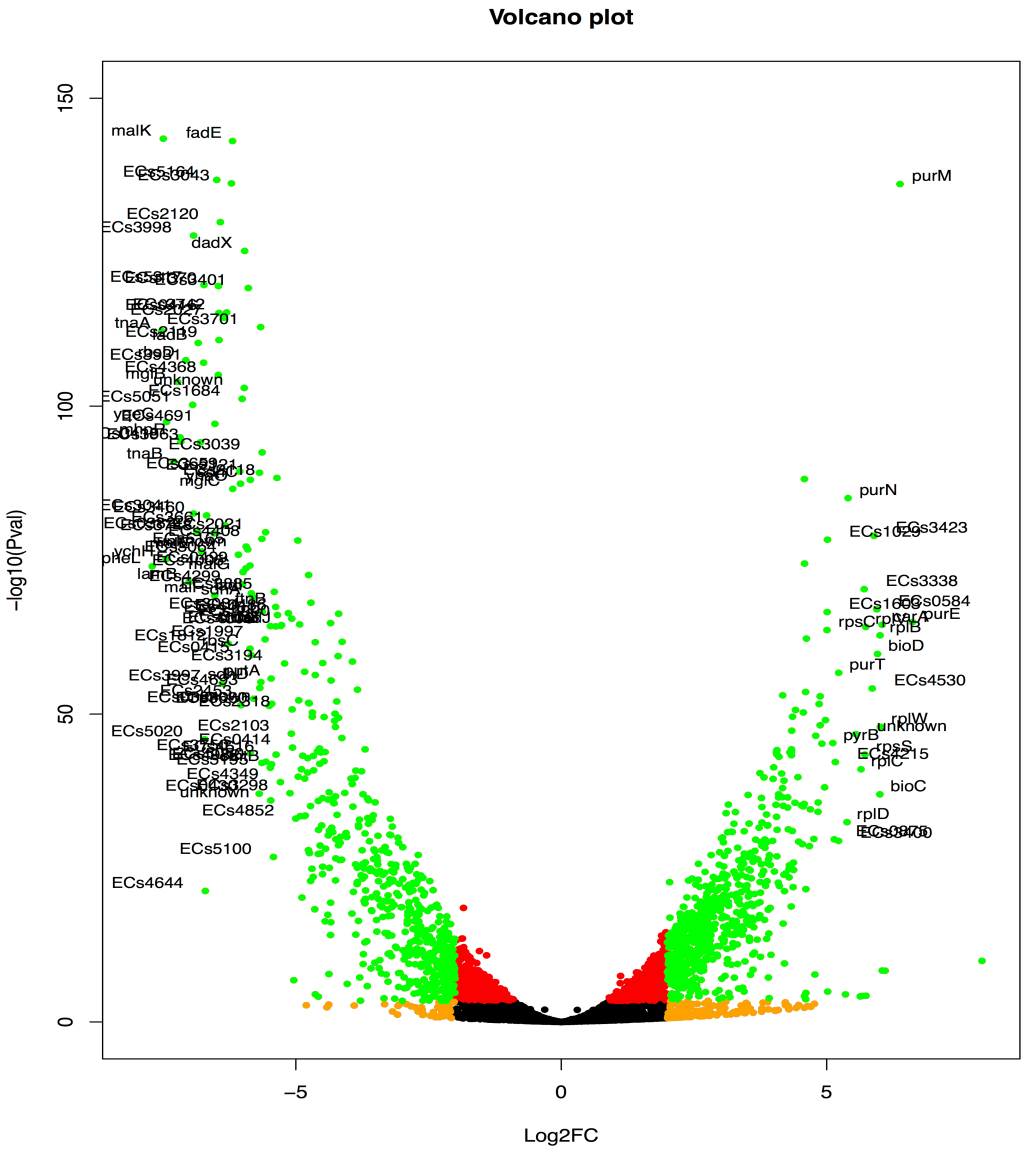

Figure 8. Volcano plot showing the statistical significance and differential expression ($\log_{10}$(pval) vs. $\log_2$(foldchange)) of 5305 genes in *E. coli* O157:H7 from RNA-Seq data.

Black points have an adj p-val of >0.001 and a $\log_2$FC of < 2. Red points have an adj p-val of <0.001 and a $\log_2$FC of < 2. Orange points have an adj p-val >0.001 and a $\log_2$FC of >2. Green points have an adj p-val <0.001 and a $\log_2$FC of >2.

In order to visualize the change in expression and p-value of all 5305 genes, a volcano plot was constructed from the RNA-Seq results (Figure 8). The

figure shows the distribution of log_2fold change and –log(pval) for all 5305 genes for which differential expression between the CBD and control conditions could be calculated (Figure 8). The shape of the plot shows the expected distribution for volcano plots with a high density of points with low foldchanges and low -log p-values (black points) and decreasing numbers of points as the foldchange and -log p-values increase. The fact that it is mostly symmetrical indicates that there were about an equal number of genes that were downregulated as were upregulated, rather than a universal increase or decrease in gene expression. It is the genes that show the highest expression changes with the greatest significance (green points) that are typically of interest. The labeled points have log_2 fold changes with an absolute value of >5 and an adjusted p-value of <1E-10, meaning that they are the most differentially regulated genes in the figure.

3.1.1 Top Differentially Regulated Genes

The following tables show the genes that had the most extreme expression changes in response to CBD. Table 2 shows the top 5 upregulated genes, and Table 3 shows the top 5 downregulated genes, all of which have adjusted p-values <0.001. Tables with the top 20 upregulated and downregulated genes are available in the appendix (Table S1 and Table S2).

Table 2. Top 5 Up-regulated genes.

Genes	**Log2 Fold Change**	**Adjusted p-value**	**Function**
pyrL	7.925466585	6.49E-10	pyrBI operon leader peptide
purE	6.617429713	8.23E-64	5-(carboxyamino)imidazole ribonucleotide mutase
purM	6.380872455	9.33E-134	phosphoribosylaminoimidazole synthetase
glnK	6.10150229	2.20E-8	nitrogen regulatory protein P-II 2

ECs4339	6.045428328	2.09E-8	beta-hydroxydecanoyl-ACP dehydrase

The top 3 upregulated genes: *pyrL*, *purE*, and *purM* (Table 2) are all involved in nucleotide biosynthesis (*The Uniprot Consortium*, 2019). This possibly indicates that the production of DNA and/or RNA increases in response to CBD. *glnK* is involved in transcriptional regulation and nitrogen regulation, and ECs4339 is thought to be involved in unsaturated fatty acid biosynthesis (*The Uniprot Consortium*, 2019).

Table 3. Top 5 Down-regulated genes.

Genes	**Log2 Fold Change**	**Adjusted p-value**	**Function**
pheL	-7.704981472	9.71E-73	leader peptide of chorismate mutase-P-prephenate dehydratase
tnaA	-7.527879548	2.15E-110	tryptophanase
malK	-7.4963831	2.00E-140	maltose/maltodextrin transporter ATP-binding protein
ychH	-7.446332907	7.46E-74	uncharacterized
yqeC	-7.433999302	7.66E-96	uncharacterized

The most downregulated gene, *pheL*, (Table 3) is involved in regulating the biosynthesis of phenylalanine, an amino acid (*The Uniprot Consortium*, 2019). *tnaA* is involved in the catabolism of tryptophan, and *malK* is involved in maltose transport (*The Uniprot Consortium*, 2019). It's worth noting that the CBD solution that was given to the bacteria contained a small amount of maltose syrup. It is possible that this gene was downregulated simply due to higher levels of maltose in the environment rather than the CBD itself. Future trials would have to be repeated with a different vehicle in order to determine the effect of pure CBD on *malK* expression. The functions of the 4th and 5th most downregulated genes (*ychH* and *yqeC*) are not fully known (*The Uniprot Consortium*, 2019).

3.1.2 Differential Expression of T3SS Genes

Originally, I suspected that CBD would cause decreased expression of T3SS genes in EHEC because Brandon Novy's qPCR results showed that *espA*, a gene that encodes a T3SS structural protein, was significantly downregulated in EHEC when it was treated with CBD (Novy, 2019). However, my RNA-Seq results did not support this hypothesis. In fact, many T3SS genes were actually upregulated in the presence of CBD (Table 4).

Table 4. Expression changes of several important T3SS genes

Gene Name	**Protein Function**	**Log2 Foldchange**	**Adjusted p-value**
espA	T3SS structural protein	2.2115	1.28E-5
ler	T3SS LEE master regulator	-1.2295	0.0008
tir	Translocated intimin receptor	0.6686	0.0707
escC	T3SS outer membrane ring protein	2.3050	2.44E-11
escU	T3SS structural protein	2.7459	2.74E-10
cesD	T3SS LEE chaperone	2.5090	3.01E-11
grlA	T3SS LEE transcriptional regulator	3.8979	5.75E-18

Note: *tir* did *not* have a statistically significant difference in expression between the CBD and control conditions. Additionally, *tir* is known to be controlled largely though post-transcriptional regulation (Pifer et al., 2018).

Although *ler*, the master regulator of the LEE pathogenicity island, was downregulated as I expected, many genes that encode structural proteins that make up the T3SS molecular syringe were actually upregulated. Surprisingly, even *espA*, which Brandon Novy found to be downregulated in two separate qPCR experiments, appeared to be upregulated. It's not clear why *espA* appears to have opposite expression changes in my results vs. Brandon Novy's. It's possible that the response of EHEC to CBD is variable. This is a topic that likely warrants more investigation in the future. However, based on my results alone, it appears that the T3SS genes are not only not downregulated, as I was expecting,

but that they are actually upregulated in response to CBD. Gene set enrichment results indicated that overall the set of T3SS genes were neither significantly upregulated nor significantly downregulated ($p = 0.3679$).

3.2 Gene Set Enrichment Analysis

In order to learn more about which pathways were being differentially expressed in response to CBD, a fast gene set enrichment analysis was run using the RNA-Seq results and information about which genes make up different pathways. Of the 240 pathways that were input into *fgsea*, only 19 were statistically differentially regulated (adjusted p-val ≤ 0.1).

Of the 9 pathways that were significantly upregulated, five are related to nucleotide biosynthesis (Table 5). The L-glutamate biosynthesis and L-glutamine degradation pathways are biologically linked because L-glutamine, an amino acid, can be converted into L-glutamate. Therefore, it makes sense the expression of the genes in these pathways are linked. Glutamate and glutamine play important roles in bacterial metabolism. L-glutamine is linked to amino acid metabolism and has been shown to reduce the catabolism of most amino acids in a dose-dependent manner (Dai et al., 2013). Amino acid metabolism and synthesis have been previously discovered to play an important role in the formation of persister cells, bacteria that enter a temporary state of inactivity under stress conditions in order to survive (Amato et al., 2014). Glutamate plays an important role in both carbon and nitrogen metabolism since it is synthesized by an intermediate in the TCA cycle and acts as an amino group donor, making it an essential for anabolic reactions (Commichau et al., 2008). The breakdown of glutamate by the enzyme GAD has been identified as a component of many stress pathways in bacteria and is considered a key part of the *E. coli*'s response

to acid stress (Feehily & Karatzas, 2012). Therefore, it's not that surprising that these pathways would be upregulated when bacteria are treated with CBD.

Another pathway that was upregulated was the amino-acyl tRNA charging path, which plays a role in translation. This paired with the upregulation of nucleotide biosynthesis pathways may indicate that translation is upregulated.

Table 5. All up-regulated pathways with an adjusted p-value ≤ 0.1.

Up-Regulated Pathways	**Enrichment Score**	**p-value**	**Adjusted p-value**
5-Aminoimidazole Ribonucleotide Biosynthesis	0.96638655	0.0019	0.0318
Inosine-5'-phosphate Biosynthesis	0.92050209	0.0094	0.0732
L-glutamate Biosynthesis	0.92050209	0.0094	0.0732
UMP Biosynthesis	0.88799799	0.0019	0.0318
L-glutamine Degradation	0.68333313	0.0058	0.0595
Purine Nucleotide *De Novo* Biosynthesis	0.66027521	0.0019	0.0318
Pyrimidine Ribonucleotide *De Novo* Biosynthesis	0.65618712	0.0077	0.0671
Amino-acyl tRNA Charging	0.37052892	0.0057	0.0595
Metabolic Clusters	0.37052892	0.0057	0.0595

Ranked from most to least upregulated.

All 10 of the pathways which are significantly downregulated in response to CBD are related to metabolism (Table 6). This makes sense since a decrease in energy expenditure is a common response to stress in bacteria and underlies the formation of persisters (Amato et al., 2014). By entering a period of metabolically-induced inactivity during stress, bacteria can often survive stressful conditions and then resume normal function when the stressor has been removed (Amato et al., 2014). This is effective because most antibiotics and

certain other stressful conditions target key cell functions and can't cause damage if the bacteria are not performing those functions.

The role of glycerol degradation (Table 6) in persister formation has been previously established however the results about the directionality of the effect are mixed. One study (Girgis et al., 2012) found that glycerol metabolism was positively linked to persister formation. Whereas a different study (Spoering & Lewis, 2006), found that high levels of glycerol degradation actually decreased persister formation. Therefore, it is difficult to say what role the downregulation of glycerol degradation plays in EHEC's response to CBD, but it is consistent with the trend of the downregulation of metabolic pathways that appears throughout the rest of the table.

Table 6. All down-regulated pathways with an adjusted p-value ≤ 0.1.

Down-Regulated Pathways	**Enrichment Score**	**p-value**	**Adjusted p-value**
L-tryptophan Degradation	-0.99791232	0.0020	0.0318
Aerobic Respiration	-0.94537815	0.0021	0.0318
Electron Transfer	-0.94537815	0.0021	0.0318
Fermentation	-0.92033543	0.0064	0.0595
Fermentation to Pyruvate	-0.91841004	0.0106	0.0748
Fermentation to Short-Chain Fatty Acids	-0.90167364	0.0149	0.1000
Fermentation to Lactate	-0.86344538	0.0064	0.0595
Fatty Acid and Lipid Degradation	-0.86344538	0.0064	0.0595
Glycerol Degradation	-0.83755274	0.0021	0.0318
TCA Cycle	-0.79663554	0.0021	0.0318

Ranked from most to least downregulated.

3.3 Effect of CBD on Growth

To check if CBD was inhibiting the growth of EHEC, which would explain the reduced colonization in mice, but would also probably mean that bacteria would be more likely to develop resistance due to selective pressure, the growth of bacteria over five hours was tested both with and without CBD. At the beginning of the five hours, a 1/100 dilution of the overnight culture of EHEC which was used to create new cultures had an optical density of 0.001 A at 600 nm. The optical density values of the six cultures which were created from the overnight culture and incubated for 5 hours in either LG-DMEM with 50 μM CBD or plain LG-DMEM are given in Table 7 below. All 6 cultures experienced at least a 200-fold increase in optical density from the starting culture.

Table 7. Optical density at 600 nm of cultures that were grown for 5 hours from a 1/100 dilution of a liquid overnight culture in DMEM either in the presence or absence of CBD.

Sample #	CBD Treatment OD^{600}	Control OD^{600}
1	0.214	0.371
2	0.369	0.320
3	0.352	0.356
Average	0.312	0.349

There was no significant difference in growth between the cultures that were treated with CBD and the controls ($t = -0.605$, $df = 2$, $p = 0.607$). This indicates that the observed effect of CBD on mice is not due to growth inhibition.

3.4 qPCR Candidates and Primer Efficacy

Eight genes were chosen as candidates for further exploration using qPCR. These genes were chosen because they are known to play a role in systems that I hypothesized may be involved in the reduced pathogenicity of EHEC when treated with CBD, and because they were significantly differentially regulated. The pathways in which these genes are involved will be discussed in more detail in the discussion. The genes that were chosen were: *ler*, *tir*, and *espA*, which are involved in the T3SS, *fadL*, *fadR* and *degP*, which are involved in envelope repair, *yhaO*, which is involved in fatty acid metabolism, *ftnB* which is involved in the Cpx envelope stress pathway, and *sdhC* which is involved in aerobic respiration. An aerobic gene was selected because of GSEA results that indicated that genes involved in aerobic respiration pathways were significantly downregulated. *yhaO*, *fadL,* and *fadR* have been previously implicated in virulence in EHEC (Pifer et al., 2018). *yhaO* and *ftnB* were also chosen because they were among the most highly downregulated genes with foldchanges of -6.042 and -5.380 respectively. In addition to the eight candidate genes, an additional gene, *rpoA,* was selected to be used as a control since it is often used as a "housekeeping gene".

The results from the test PCR run with the five primer pairs I designed myself are shown below (Fig 9).

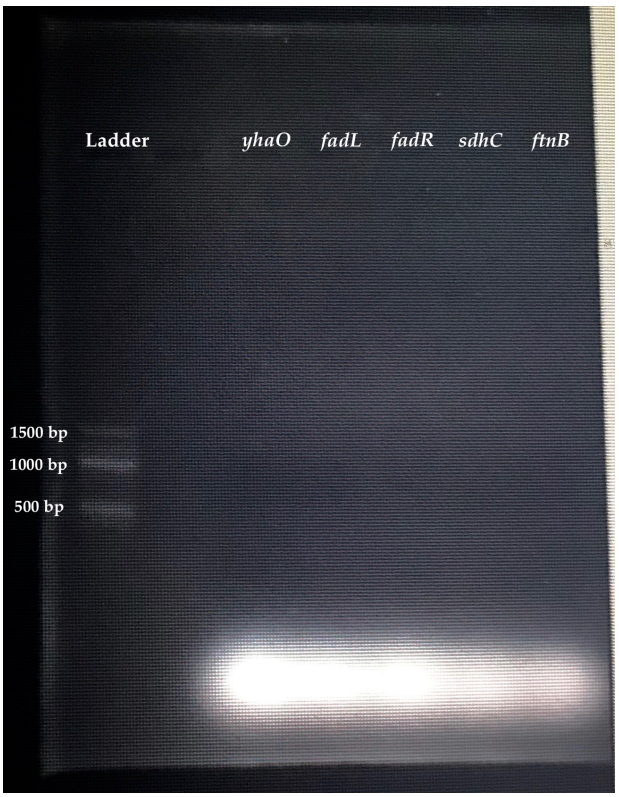

Figure 9. Gel electrophoresis results from products of PCR that was run with EHEC DNA and the five primer pairs that were designed.
From left-to-right the lanes contain: Invitrogen 100 bp ladder, and the *yhaO* (132 bp), *fadL* (113 bp), *fadR* (146 bp), *sdhC* (113 bp) and *ftnB* (70 bp) PCR products.
The image is low quality because I was unable to access the lab computer on which this image is saved and had to rely on a photo taken on an iPhone.

As can be seen in Figure 9, DNA fragments are visible for all five primer pairs that were tested. These fragments appear to be <300bp in size, which makes sense because all 5 primer pairs were meant to produce fragments between 70-146 bp in length (Table 1). Therefore, it seems likely that all of the primer pairs successfully amplified the correct regions of the genome.

I was unable to complete the rest of the qPCR experiments with the candidates I had chosen due to Covid-19. However, my primers may be able to be used by future students studying EHEC.

3.5 Interactome Analysis Results

The subinteractome which was constructed from candidate genes that were chosen for qPCR (section 3.4), the neighbors of candidates, and the neighbors of neighbors is shown below (Figure 10).

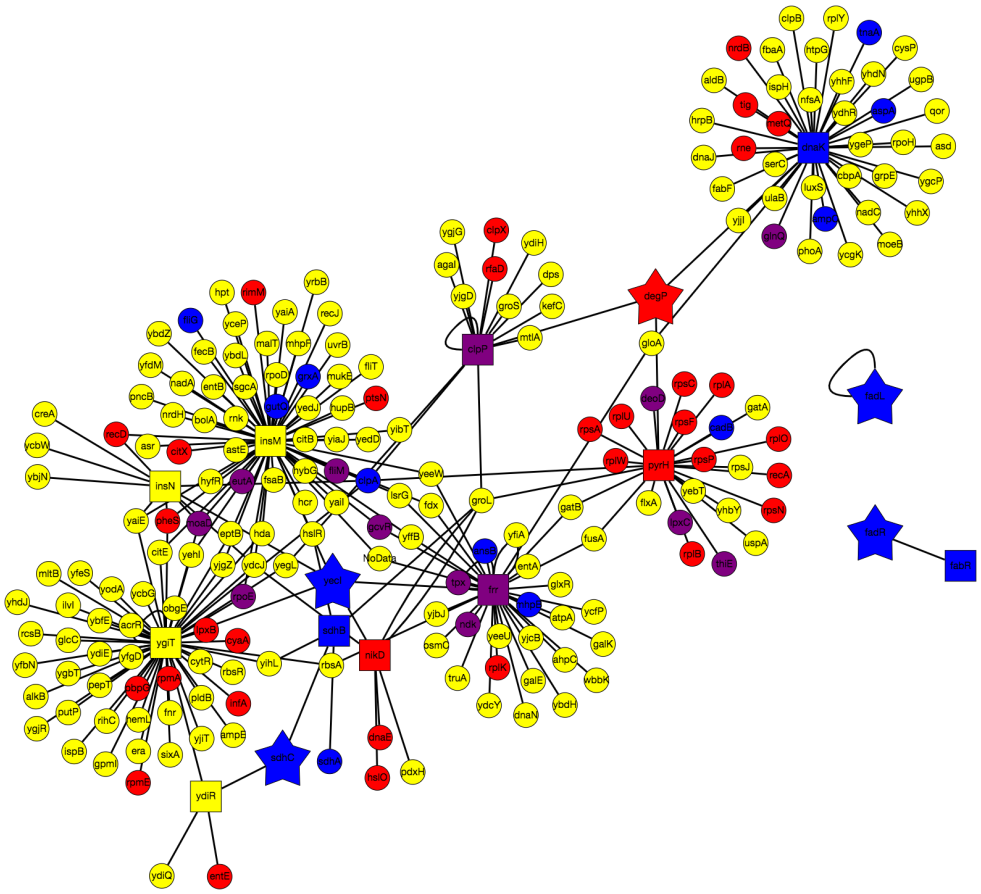

Figure 10. *E. coli* K12 sub-interactome.

Candidate nodes are represented as stars, immediate neighbors of candidate nodes are represented as squares, neighbors of neighbors are represented as circles. Red nodes are upregulated (foldchange > 0.5), blue nodes are downregulated (foldchange < 0.5), purple nodes have foldchanges between -0.5 and 0.5. Nodes that were unable to be mapped to gene regulation data are represented in yellow. Link to interactive graph: http://graphspace.org/graphs/28361?user_layout=13056

As can be seen in the figure, most of the candidate genes had relatively few neighbors whereas their neighbors had a higher number of connections. This potentially indicates that the selected candidate genes primarily regulate processes through their limited connections to hub nodes. However, it is also possible that there is just less information that is known about the interactions of these genes for whatever reason. A table listing the foldchanges and proteins functions of the immediate neighbors of candidate nodes is available below (Table 8).

Table 8. Immediate neighbors of candidate nodes in *E. coli* K12 subinteractome

Gene Name	**Protein**	**Expression Foldchange**
clpP	ATP-dependent Clp protease proteolytic subunit	-0.006993
dnaK	Chaperone protein DnaK	-0.711098
fabR	HTH-type transcriptional repressor FabR	-0.689206
frr	Ribosome-recycling factor	0.133232
insM	IS600 ORF1-like protein	No Data
insN	Putative transposase InsN for insertion sequence element IS911A	No Data
nikD	Nickel import ATP-binding protein NikD	3.037433
pyrH	Uridylate kinase	2.714403
sdhB	Succinate dehydrogenase iron-sulfur subunit	-5.479551
ydiR	Putative electron transfer flavoprotein subunit YdiR	No Data
ygiT	Antitoxin MqsA	No Data

Of the 11 proteins in Table 8, seven were related to pathways that I previously identified as pathways of interest. The genes *clpP*, *dnaK*, and *ygiT* are all known to be involved in stress response pathways in EHEC. *frr* is involved in translation and ribosome efficiency. *fabR* represses fatty acid synthesis and is involved in the maintenance of the bacterial membrane. *pyrH* is involved in pyrimidine biosynthesis and thus plays a role in the synthesis of RNA and DNA.

sdhB plays a role in the TCA cycle, which was previously indicated by *fgsea* analysis to be downregulated (Table 6). The fact that these genes play a role in pathways of interest is to be expected since they interact with candidate genes that were chosen partly based on their roles in these pathways. Of these neighbor genes of interest, only 2, *sdhB* and *pyrH*, showed large increases or decreases in gene expression when the bacteria were exposed to CBD. Although it is not clearly involved in a pathway of interest, *nikD*, which is involved in ATP binding is also relatively highly upregulated.

Using all the genes in the subinteractome for which gene expression information was available, including the candidate genes, an *fgsea* analysis was run (described in Methods section 2.2.3). Table 9 below shows the pathways that were significantly differentially regulated. Although other pathways were found to be differentially regulated, all of them had very high p-values.

Table 9. Significantly differentially regulated pathways including only nodes from the subinteractome

Differentially-Regulated Pathways	**Enrichment Score**	**p-value**	**Adjusted p-value**
Electron Transfer	-0.98412670	0.0024	0.0238
TCA Cycle	-0.98412670	0.0024	0.0238
Aerobic Respiration	-0.98412670	0.0024	0.0238

Although there were other pathways that were up or down-regulated, only the above three had adjusted p-values < 0.1. All three pathways share the same enrichment scores and p-values because they include the same subset of genes.

The fact that metabolic pathways were very down-regulated in the subinteractome is unsurprising given that *sdhC*, a gene that is known to be involved in aerobic respiration, was one of the five candidate nodes. *sdhC* was directly connected to *sdhB*, which in turn was connected to *sdhA*. Both *sdhB* and *sdhA* are involved in aerobic respiration and, like *sdhC* (foldchange = -5.8563), are

very down-regulated (foldchanges of -5.4796 and -5.8076 respectively), accounting for the changes in the metabolic pathways. Metabolic pathways were also found to be highly downregulated in the whole genome *fgsea* analysis (see Table 6), which is why an aerobic respiration gene was chosen as a candidate. The downregulation of metabolic pathways makes sense since bacteria often decrease their energy expenditure and active processes during times of stress (Amato et al., 2014).

Although it didn't result in any significantly differentially regulated pathways, a cluster of genes in the subinteractome (Figure 11) was chosen for further investigation based on the high number of differentially regulated genes in the cluster, most of which were upregulated.

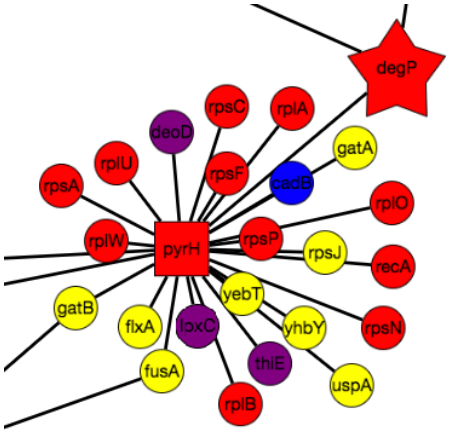

Figure 11. Close up of the cluster of genes connected to *pyrH*, a neighbor of the candidate node *degP*.
Red genes are upregulated, blue are downregulated, purple show little change in expression, and the yellow nodes are genes for which expression values were not able to be found. This cluster of genes was chosen for closer exploration due to the high proportion of differentially regulated genes.

This cluster of nodes surrounds *pyrH* which is a neighbor of the candidate node *degP* (Fig 11). Although one of the nodes connected to *pyrH* is downregulated (*cadB*, foldchange = -1.7795), most of *pyrH*'s neighbors are

upregulated. Two of the connected genes, *rplW* and *rplB,* even have foldchanges >6, making them among the most upregulated genes (Table S1). As mentioned previously, *pyrH* is involved in pyrimidine nucleotide biosynthesis and *de novo* CTP biosynthesis; the functions of its differentially regulated neighbors are represented in the following table (Table 10):

Table 10. Differentially regulated (foldchange >0.5 or <-0.5) neighbors of *pyrH*

Gene	Log_2 Foldchange	Protein	Pathway(s)
rplW	6.011158	50S ribosomal protein L23	Ribosomal large subunit assembly, Translation
rplB	6.004577	50S ribosomal protein L2	Cytoplasmic translation, ribosomal large subunit assembly
rpsC	5.006046	30S ribosomal protein S3	Ribosomal small subunit assembly, Translation
rplO	4.525727	50S ribosomal protein L15	Translation
rpsF	4.337633	30S ribosomal protein S6	Translation
rpsP	4.224719	30S ribosomal protein S16	Ribosomal small subunit assembly, Translation
rpsA	3.883043	30S ribosomal protein S1	Negative regulation of cytoplasmic translation, Ribosomal small subunit assembly
rplA	3.769461	50S ribosomal protein L1	Maturation of LSU-rRNA, Negative regulation of translation initiation, Ribosomal large subunit assembly
rpsN	3.578682	30S ribosomal protein S14	Ribosomal small subunit assembly, Translation
degP	2.480371	Periplasmic serine endoprotease DegP *(Candidate chosen for known role in envelope repair)*	Chaperone-mediated protein folding, Protein quality control for misfolded or incompletely synthesized proteins, Proteolysis, Response to heat, Response to oxidative stress
rplU	2.191056	50S ribosomal protein L21	Translation
recA	0.733179	Protein RecA	Cellular response to DNA damage, DNA recombination, DNA repair, Response to radiation, SOS response

cadB	-1.779484	Probable cadaverine/ lysine antiporter	Cadaverine transport, Cellular stress response to acidic pH, L-lysine transmembrane transport

Ranked from most upregulated to most downregulated. Includes candidate gene, *degP*. Genes that had were not differentially expressed or for which there was no expression data were not included. Protein information and pathways were obtained from the *Uniprot* database (The Uniprot Consortium, 2019).

As can be seen in Table 10, many of the genes in the cluster are related to translation – either through their involvement in the biogenesis of ribosomes, which are the essential machinery of translation, or such as is the case for *degP*, one of the five candidate genes, through chaperoning protein folding and protein quality control. However, this data is slightly complicated since some of the upregulated genes, such as *rpsA* and *rplA*, are actually believed to be involved in inhibiting translation. That being said, the majority of the upregulated genes in the cluster positively regulate translation, indicating that perhaps CBD is causing EHEC to increase its rate of translation.

In order to learn more about the role of genes related to ribosome biogenesis, I investigated the role of ribosomes in stress pathways and identified several important genes not included in the cluster in Figure 11. These genes which are essential to the ribosome hibernation response, in which ribosomes dimerize to prevent damage under stressful conditions, were selected based on information presented in (Matzov et al, 2019) and (Ueta et al., 2005) (Table 11).

Table 11. Foldchange values of genes involved in ribosome hibernation and dimerization

Gene Name	Log_2 Foldchange	Adjusted p-value	Protein Function
hpf	-0.9054	0.0006	Converts immature 90S ribosome dimers to mature 100S dimers
rmf	-2.8696	<1.0e-15	Induces dimerization of ribosomes into immature 90S dimer
ybhY	1.8464	1.06e-9	Stabilizes dimer formation
yfiA	-6.6814	<1.0e-15	Prevents ribosome dimer formation.

The above four genes were chosen based on a literature review of *E. coli* genes involved in ribosome hibernation during stress. The functions of *hpf* and *rmf* were obtained from (Matzov et al., 2019). Information about the functions of *ybhY* and *yfiA* was obtained from (Ueta et al., 2005).

The two primary genes involved in ribosome dimerization are *rmf* which causes two ribosomes to form an immature 90S dimer and *hpf* which converts the immature 90S dimer to a more stable 100S dimer (Matzov et al., 2019). The basic proposed pathway of ribosome dimer formation is shown in Figure 12.

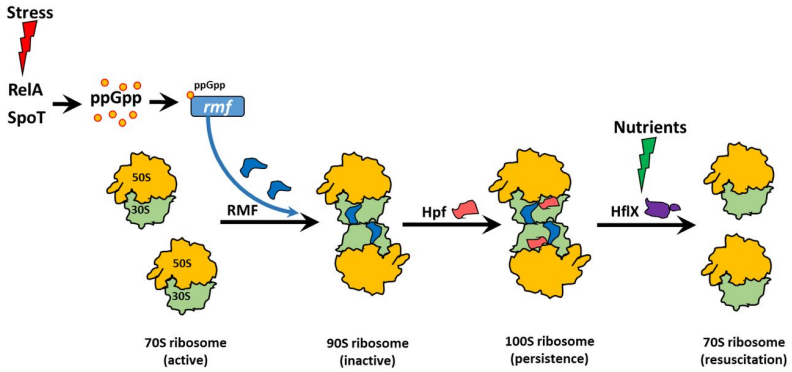

Figure 12. Ribosome dimer formation in *E. coli* in response to stress (in this example nutrient deprivation) (Song & Wood, 2019).

The other two genes in Table 9, *ybhY* and *yfiA,* have opposite functions; *ybhH* encourages dimer formation whereas *yfiA* prevents dimer formation (Ueta et al., 2005). Surprisingly, given that ribosome hibernation has been implicated as a crucial part of bacterial stress survival (Matzov et al., 2019), the primary genes involved in dimer formation (*rmf* and *hpf*) were actually slightly downregulated. However, *ybhH* which stabilizes dimer formation was slightly upregulated, and *yfiA* which prevents dimer formation was highly downregulated. In fact, *yfiA* was the 23rd most downregulated gene. These mixed results make it unclear whether ribosome hibernation is increasing, decreasing, or remaining unchanged when the bacteria are exposed to CBD. Although there are some contradictory trends in the data, the cluster of upregulated genes in Table 10 and the fact that

yfiA is one of the most highly downregulated genes seems to indicate that CBD is affecting translation, ribosome biogenesis, and ribosome hibernation in *E. coli* O157.

Discussion

The goal of this thesis was to use transcriptomic data to uncover the effect of CBD on gene expression in enterohemorrhagic *E. coli*. Previous research by Brandon Novy (2019) showed that exposure to CBD resulted in a decrease in virulence in *C. rodentium*, a murine model bacteria that is similar to EHEC. When treated with CBD, mice that had been infected with *C. rodentium* shed lower levels of bacteria in their feces (Novy, 2019). However, the molecular mechanism that underlies this effect was unclear. Therefore, I set out to use RNA-Seq data to elucidate the pathway by which CBD downregulates virulence in EHEC.

Originally, I hypothesized that CBD would result in downregulation of genes that play a role in the type III secretion system since the T3SS is a key component of EHEC's virulence program and because Novy observed significant downregulation of *espA*, a gene that encodes a key T3SS structural component, in response to CBD. However, the RNA-Seq data did not support this hypothesis. In fact, the majority of T3SS genes, including *espA*, actually appeared to be slightly upregulated in the presence of CBD (Table 4). Gene set enrichment results indicated that overall the T3SS pathway was neither significantly downregulated nor upregulated ($p = 0.3679$). Given that these results vary starkly from past data, it may be beneficial to do further experiments using qPCR on specific T3SS genes such as *espA* and *ler* to better understand the effect of CBD on the type III secretion system.

Gene set enrichment analysis on the RNA-seq results showed that several pathways related to nucleotide biosynthesis, amino acid biosynthesis, and the "amino-acyl tRNA charging" pathway, which is involved in translation, were

significantly upregulated (Table 5), whereas various metabolic pathways were significantly downregulated (Table 6). Interestingly, Brandon Novy had also done an initial RNA-Seq analysis (but with fewer replicates) and found that out of the 438 genes that were significantly differentially regulated, many were related to metabolism, amino acid biosynthesis, macromolecular biosynthesis, and the transportation of small molecules (Novy, 2019). Therefore, both of our RNA-Seq analyses showed that genes related to metabolism, amino acid biosynthesis, and macromolecular biosynthesis are consistently being differentially expressed in the presence of CBD. The fact that the three CBD samples and 3 control samples had much more similar gene expression profiles to other samples within the group than to each other (Figure 7) is also evidence that the changes in gene expression produced by CBD are somewhat consistent.

The upregulation in nucleotide biosynthesis and amino-acyl tRNA charging in my RNA-Seq data potentially indicates that either transcription or translation is being increased. The downregulation of metabolic pathways indicates that the bacteria are decreasing their energy production and expenditure. It is not uncommon for bacteria to decrease energy production under stress. The formation of persister cells, bacteria that enter an inactive state to survive stressful conditions, has been linked to downregulation of metabolic activity (Amato et al., 2014). However, persister formation should result in decreased growth, and results from a growth experiment indicated that CBD had no significant impact on EHEC's growth *in vitro* (Table 7). This means that despite the fact that CBD is seemingly causing EHEC to decrease energy production, the bacteria are somehow compensating for this and managing to sustain growth.

Since no change was observed in the T3SS genes, further research was done to investigate other pathways that might be significant in regulating

EHEC's response to CBD. There are at least three pathways that were identified as potentially significant in regulating EHEC's response to CBD: the Cpx envelope stress response, the σ^E envelope stress response, and amino acid and fatty acid metabolism.

E. coli uses the Cpx envelope stress pathway to respond to stressors that cause envelope protein misfolding such as low pH, copper, chloride ions, and overexpressed periplasmic proteins (Raivio et al., 2013). The Cpx response leads to an upregulation of genes encoding protein folding and degrading factors, and the downregulation of inter-envelope protein complexes such as pilli and secretion systems (Raivio et al., 2013). The Cpx pathway is believed to have an effect on phospholipid metabolism, adherence, pathogenesis, and other stress pathways (Raivio et al., 2013; MacRitchie et al., 2008). One study found that activation of the Cpx pathway resulted in the downregulation of several T3SS genes, including *espA* and *tir*, in EPEC (enteropathogenic *E. coli*) a type of bacteria that is very similar to EHEC (MacRitchie et al., 2008). The Cpx pathway begins with Cpx sensor proteins triggering increased synthesis of the proteins CpxP and DegP. Although my results showed no significant change in expression of *cpxP*, *degP* was upregulated 2-fold. Surprisingly, activation of the Cpx pathway was also associated with upregulation of the gene *ftnB*, which was highly downregulated ($\log_2$foldchange = -5.380). Therefore, *degP* and *ftnB* were chosen as candidates for further investigation.

A similar pathway of interest is the *rpoE* (σ^E) envelope stress pathway. The *rpoE* stress pathway is involved in responding to periplasmic protein stress, oxidative stress, and heat shock (Raivio, 2005). Like the Cpx pathway, the *rpoE* pathway plays an important role in virulence since virulence systems are embedded within the bacterial membrane (Raivio, 2005). The *rpoE* stress pathway also plays an essential role in outer membrane protein biogenesis

(Raivio, 2005). Like the Cpx pathway, the *rpoE* pathway also triggers *degP* expression (Raivio, 2005). In addition to *degP*, two other genes in the *rpoE* pathway, *fadL* and *fadR*, which were downregulated and play roles in fatty acid metabolism and transport were chosen as candidate genes for further investigation.

Finally, several genes related to amino acid and fatty acid metabolism were chosen based on evidence that these pathways affect the expression of LEE and T3SS genes (Pifer et al., 2018). Specifically, two transcriptional regulators, CutR and FadR, which play a role in regulating metabolism were shown to affect LEE expression (Pifer et al., 2018). FadR was shown to repress LEE expression whereas CutR activated LEE expression (Pifer et al., 2018). FadR and CutR interact with several genes, including the fatty acid pump, *fadL*, and the serine importer, *yhaO* (Pifer et al., 2018). CutR is believed to activate expression of *fadL*, which inhibits the activity of FadR, preventing it from inhibiting LEE expression (Pifer et al., 2018). CutR also interacts with YhaO which activates the transcription factor YhaJ leading to increased expression of LEE genes (Pifer et al., 2018). As discussed in section 1.5 of the introduction, LEE genes control virulence in EHEC. The RNA-Seq data indicated that *fadL*, *fadR*, *cutR,* and *yhaO* were all downregulated in the presence of CBD, with *yhaO* being the 43rd most downregulated gene. Therefore, given the strong downregulation of *yhaO* and the fact that pathways related to metabolism were very differentially regulated in general (Tables 5 & 6), this pathway seemed like a promising candidate for further exploration. *yhaO* was chosen as a candidate gene, and *fadL* and *fadR* had already been identified as candidates due to their role in the *rpoE* stress pathway.

The candidate genes from each of these pathways, as well as an aerobic respiration gene, *sdhC*, and three genes that are involved in the T3SS, *ler*, *tir*, and *espA*, were chosen for further investigation with qPCR (See section 3.4). Genes

were chosen based on both their importance in their respective pathway as well as how significantly their gene expression changed in response to CBD. I had hoped that qPCR would provide more accurate information about expression changes in these genes and would either confirm or reject my surprising findings about *espA* being upregulated. Unfortunately, I was not able to complete these experiments due to Covid-19. However, I did design primers that produced PCR products of the expected size (Figure 9) which can be used in the future to investigate further the expression of these genes in response to CBD.

Since these pathways could not be investigated experimentally, a computational approach was used instead. A subinteractome containing a subset of candidate genes and genes that were connected to them was constructed as described in section 2.2.3. A cluster of genes within the interactome (Figure 11) was investigated further because it contained a high number of differentially regulated genes. The cluster surrounded *pyrH*, a gene involved in nucleotide biosynthesis, which was connected to the candidate gene *degP*, which is involved in envelope stress pathways. The majority of differentially regulated genes in the cluster (Table 10) were upregulated and were related to ribosome biogenesis and translation. Although the results were not entirely clear since some of the upregulated genes were involved in inhibiting translation, the majority of the genes positively regulate translation and ribosome biogenesis. These results seem to indicate that translation and ribosome biogenesis increased in response to CBD. These results are surprising given that many of the metabolic pathways were downregulated (Table 6) and translation requires an enormous amount of energy (60-80% of energy use in the cell) (Zhu & Dai, 2020). Some researchers have suggested that maintaining a high rate of translation may be an important aspect of stress response since bacteria need to be able to produce proteins to respond to the stressor (Zhu & Dai, 2020). They found that maintaining

translation was most essential for bacterial in the period shortly after being initially exposed to the stressor and that inhibiting translation after the bacteria had had a chance to respond had less of an effect on bacterial survival (Zhu & Dai, 2020). This indicates that timing is an important variable to take into account about changes in translation in response to stress. The RNA for RNA-Seq was extracted after the bacteria had been growing in a culture containing CBD for five hours, during the late exponential phase of growth. Perhaps experimenting with timing by testing the expression of certain genes shortly after exposure and after the bacteria had been exposed to CBD overnight would result in drastically different expression values.

Based on evidence that ribosome biogenesis was increasing when the bacteria were exposed to CBD, I investigated the role of ribosomes in stress pathways further. Under many stressful conditions ribosomes form "hibernation dimers" which protect them from degradation, but also prevent them from being used for translation. There are four main genes involved in ribosome dimerization in *E. coli*, however the expression changes of these genes do not clearly indicate whether dimerization is increasing or not (Table 11). Although a gene involved in repressing dimerization, *yfiA*, was found to be extremely downregulated, the genes that are involved in initiating dimerization were also slightly downregulated (Table 11). Therefore, it's not completely clear whether ribsome dimerization is increasing, decreasing or remaining constant in response to CBD. However, given that *yfiA* was the 23rd most downregulated gene in the entire genome there does seem to be some effect on ribosome behavior. Is it possible that ribosome biogenesis is increasing in order to maintain the translation rate while a subset of ribosome go into hibernation? In the nutrient starvation stress pathway, some ribosome go into hibernation, while others remain active in order to produce stress-defense proteins (Zhu & Dai, 2020). I

think it's possible that a similar process might occur in response to CBD. The role of ribosomes in responding to CBD is likely a process that warrants further investigation.

Ultimately, the RNA-Seq data seems to indicate that metabolic pathways, nucleotide biosynthesis, translation, and ribsome biogenesis are processes that are being altered in some way in response to CBD. Currently the mechanism by which these pathways are altered and the overall effect on virulence remains unclear. One aspect of this thesis that needs to be taken into account is that RNA-Seq data only quantifies RNA transcripts, which gives information about transcription, but not about post-transcriptional modifications. Certain genes that have been studied in this thesis including *rpoE*, *espA*, and *tir*, are all believed to undergo significant post-transcriptional modification (Raivio et al., 2013). Additionally, there is data to suggest that ribosomes preferentially translate certain RNA transcripts over others (Zhu & Dai, 2020), so simply quantifying mRNA does not give the full picture of which genes are actually being translated. Future research may address these weaknesses by using ribosome profiling, a high throughput method similar to RNA-Seq which only sequences RNA which is protected by an attached ribosome. This method is useful because it gives information about which genes are being actively translated in the cell. Additionally, the quantities of specific candidate proteins could be tested under different conditions using mass-spectrometry. Future researchers might use these techniques to further investigate changes in the expression of proteins related to metabolism, translation, ribosome biogenesis, nucleotide biosynthesis and the proposed stress pathways (Cpx envelope stress, σ^E envelope stress, and fatty acid/amino acid metabolism) mentioned previously.

I suspect that the mechanism that regulates EHEC's response to CBD is much more complex than my initial hypothesis, which was that virulence would

decrease due to downregulated T3SS components, and instead that CBD has wide ranging effects on translation, metabolism, and virulence. Despite the fact that the effect of CBD on EHEC remains unclear, given the demonstrated efficacy in CBD in mice trials, I believe that this subject warrants further investigation. Gaining a better understanding of the way CBD affects EHEC would be useful, not only in assessing its utility as a potential treatment, but also in increasing our knowledge of the way EHEC and similar pathogens respond to different stressors and small signaling molecules in the gut. As the incidence of antibiotic resistant bacteria increases, it is important to explore non-bactericidal alternatives, which are sometimes called anti-infectives. Investigation into CBD, a seemingly non-bacteriostatic non-lethal treatment for EHEC, would contribute to our growing knowledge about anti-infective treatments. I hope that thesis will serve as a starting point for further research into the molecular effects of CBD on EHEC.

Appendix

5.0 Explanation of *de novo* assembly

This can be done using a De Bruijn algorithm (Illumina, 2009). The De Bruijn algorithm breaks the reads into smaller fragments, of length *k* base pairs, known as k-mers, where the value of k is automatically determined based on the length of the reads and the error rate (Illumina, 2009). The De Bruijn algorithm then finds sequence overlaps of length *k*-1 between the various k-mers (Illumina, 2009). These overlaps are represented as edges (connections) in a graph (graphs are explained in more detail in the "co-expression networks" section 1.9) where the k-mers represent nodes. The graph is then used to attempt to assemble the genome. There are many programs that can be used to perform de novo assembly from RNA-Seq reads (Illumina, 2009).

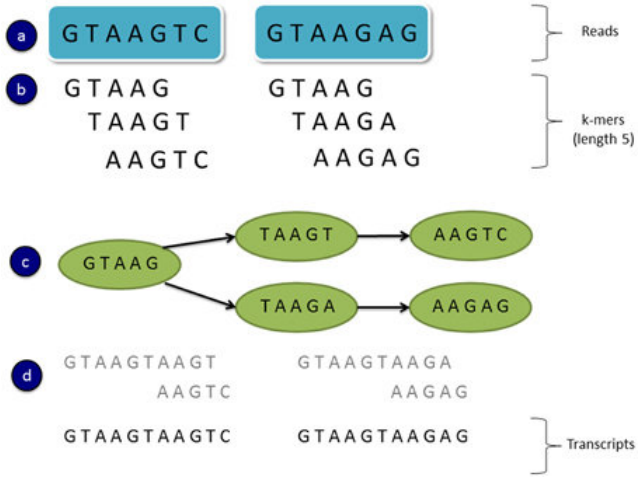

Figure S1. An example of *de novo* assembly using a De Bruijn graph (Moreton et al., 2015).

The process of going from reads to a genome is represented in the figure above (Figure S1).

5.1 Supplementary Tables and Figures

Table S1. Top 20 up-regulated genes (that have an adjusted p-value <0.001)

Genes	**Log2 Fold Change**	**Adjusted p-value**	**Function**
pyrL	7.925466585	6.49E-10	pyrBI operon leader peptide
purE	6.617429713	8.23E-64	5-(carboxyamino)imidazole ribonucleotide mutase
purM	6.380872455	9.33E-134	phosphoribosylaminoimidazole synthetase
glnK	6.10150229	2.20E-8	nitrogen regulatory protein P-II 2
ECs4339	6.045428328	2.09E-8	beta-hydroxydecanoyl-ACP dehydrase
carA	6.042950677	2.15E-63	carbamoyl phosphate synthase small subunit
rplW	6.011158357	4.75E-47	50S ribosomal protein L23
rplB	6.004577256	1.03E-61	50S ribosomal protein L2
bioC	6.001359793	2.97E-36	biotin biosynthesis protein BioC
bioD	5.9575488	1.01E-58	dithiobiotin synthetase
purK	5.944074215	7.75E-66	5-(carboxyamino)imidazole ribonucleotide synthase
purL	5.935361396	1.35E-77	phosphoribosylformylglycinamidine synthase
xanP	5.858100166	3.52E-53	transporter
citD	5.740390746	0.000162	citrate lyase subunit gamma
rplV	5.736415519	4.53E-63	50S ribosomal protein L22
rpsS	5.72125339	1.28E-42	30S ribosomal protein S19
purC	5.705645374	4.50E-69	phosphoribosylaminoimidazole-succinocarboxamide synthase
rrfF	5.673793155	0.000176	5S ribosomal RNA
rplC	5.646835997	2.92E-40	50S ribosomal protein L3
hyfA	5.631751213	0.000205	hydrogenase 4 membrane subunit

Table S2. Top 20 down-regulated genes (that have an adjusted p-value <0.001)

Genes	**Log2 Fold Change**	**Adjusted p-value**	**Function**
pheL	-7.704981472	9.71E-73	leader peptide of chorismate mutase-P-prephenate dehydratase
tnaA	-7.527879548	2.15E-110	tryptophanase
malK	-7.4963831	2.00E-140	maltose/maltodextrin transporter ATP-binding protein
ychH	-7.446332907	7.46E-74	uncharacterized
yqeC	-7.433999302	7.66E-96	uncharacterized
tnaB	-7.30203813	1.78E-89	tryptophan permease TnaB
mglB	-7.22741238	2.93E-102	galactose-binding transport protein
mhpR	-7.180736287	2.26E-93	MhpR family transcriptional regulator
ECs0418	-7.164920955	1.04E-92	transcriptional regulator
rbsD	-7.072241763	9.89E-106	D-ribose pyranase
lamB	-7.008387041	3.09E-70	maltoporin
acs	-6.942984921	1.31E-98	acetyl-CoA synthetase
mglA	-6.92715623	3.7E-81	galactose/methyl galaxtoside transporter ATP-binding protein
trmJ	-6.925803669	1.53E-125	TdcA family transcriptional regulator
yahN	-6.858092698	2.84E-78	cytochrome subunit of dehydrogenase
fadB	-6.838766902	1.65E-108	multifunctional fatty acid oxidation complex subunit alpha
fadH	-6.795616118	1.35E-92	NADPH dehydrogenase
malE	-6.774767206	3.97E-75	maltose ABC transporter periplasmic protein
glgS	-6.73739298	2.39E-105	glycogen synthesis protein GlgS
lgoR	-6.730261524	1.26E-117	transcriptional activator for L-galactonate catabolism

Bibliography

Acheson, D. W., Moore, R., De Breucker, S., Lincicome, L., Jacewicz, M., Skutelsky, E., & Keusch, G. T. (1996). Translocation of Shiga toxin across polarized intestinal cells in tissue culture. *Infection and Immunity*, *64*(8), 3294–3300.

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B. A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., & Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, *46*(W1), W537–W544. https://doi.org/10.1093/nar/gky379

Aizpurua-Olaizola, O., Soydaner, U., Öztürk, E., Schibano, D., Simsir, Y., Navarro, P., Etxebarria, N., & Usobiaga, A. (2016). Evolution of the Cannabinoid and Terpene Content during the Growth of Cannabis sativa Plants from Different Chemotypes. *Journal of Natural Products*, *79*(2), 324–331. https://doi.org/10.1021/acs.jnatprod.5b00949

Albenberg, L., Esipova, T., Judge, C., Bittinger, K., Chen, J., Laughlin, A., Grunberg, S., Baldassano, R., Lewis, J., Li, H., Thom, S., Bushman, F., Vinogradov, S., & Wu, G. (2014). Correlation Between Intraluminal Oxygen Gradient and Radial Partitioning of Intestinal Microbiota in Humans and Mice. *Gastroenterology*, *147*(5), 1055-1063.e8. https://doi.org/10.1053/j.gastro.2014.07.020

Alexeyenko, A., Wassenberg, D. M., Lobenhofer, E. K., Yen, J., Linney, E., Sonnhammer, E. L. L., & Meyer, J. N. (2010). Dynamic Zebrafish Interactome Reveals Transcriptional Mechanisms of Dioxin Toxicity. *PLOS ONE*, *5*(5), e10465. https://doi.org/10.1371/journal.pone.0010465

Amato, S. M., Fazen, C. H., Henry, T. C., Mok, W. W. K., Orman, M. A., Sandvik, E. L., Volzing, K. G., & Brynildsen, M. P. (2014). The role of metabolism in

bacterial persistence. *Frontiers in Microbiology*, *5*. https://doi.org/10.3389/fmicb.2014.00070

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Ashton, J. C., Dowie, M. J., & Glass, M. (2017). Chapter 5 - The Endocannabinoid System and Human Brain Functions: Insight From Memory, Motor, and Mood Pathologies. In E. Murillo-Rodríguez (Ed.), *The Endocannabinoid System* (pp. 115–186). Academic Press. https://doi.org/10.1016/B978-0-12-809666-6.00005-8

Bäumler, A. J., & Sperandio, V. (2016). Interactions between the microbiota and pathogenic bacteria in the gut. *Nature*, *535*(7610), 85–93. https://doi.org/10.1038/nature18849

Bielaszewska, M., Idelevich, E. A., Zhang, W., Bauwens, A., Schaumburg, F., Mellmann, A., Peters, G., & Karch, H. (2012). Effects of Antibiotics on Shiga Toxin 2 Production and Bacteriophage Induction by Epidemic Escherichia coli O104:H4 Strain. *Antimicrobial Agents and Chemotherapy*, *56*(6), 3277–3282. https://doi.org/10.1128/AAC.06315-11

Bolger, A.M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics, btu170.

Cameron, D. R., Shan, Y., Zalis, E. A., Isabella, V., & Lewis, K. (2018). A Genetic Determinant of Persister Cell Formation in Bacterial Pathogens. *Journal of Bacteriology*, *200*(17), e00303-18. https://doi.org/10.1128/JB.00303-18

Camilleri, M., Carlson, P., McKinzie, S., Grudell, A., Busciglio, I., Burton, D., Baxter, K., Ryks, M., & Zinsmeister, A. R. (2008). Genetic variation in endocannabinoid metabolism, gastrointestinal motility, and sensation. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, *294*(1), G13–G19. https://doi.org/10.1152/ajpgi.00371.2007

Carlson-Banning, K. M., & Sperandio, V. (2016). Catabolite and Oxygen Regulation of Enterohemorrhagic Escherichia coli Virulence. *MBio*, *7*(6). https://doi.org/10.1128/mBio.01852-16

Carlson-Banning, K. M., & Sperandio, V. (2018). Enterohemorrhagic Escherichia coli outwits hosts through sensing small molecules. *Current Opinion in Microbiology*, *41*, 83–88. https://doi.org/10.1016/j.mib.2017.12.002

Chase-Topping, M. E., McKendrick, I. J., Pearce, M. C., MacDonald, P., Matthews, L., Halliday, J., Allison, L., Fenlon, D., Low, J. C., Gunn, G., & Woolhouse, M. E. J. (2007). Risk Factors for the Presence of High-Level Shedders of Escherichia coli O157 on Scottish Farms. *Journal of Clinical Microbiology*, *45*(5), 1594–1603. https://doi.org/10.1128/JCM.01690-06

Cobbold, R. N., Hancock, D. D., Rice, D. H., Berg, J., Stilborn, R., Hovde, C. J., & Besser, T. E. (2007). Rectoanal Junction Colonization of Feedlot Cattle by Escherichia coli O157:H7 and Its Association with Supershedders and Excretion Dynamics. *Applied and Environmental Microbiology*, *73*(5), 1563–1568. https://doi.org/10.1128/AEM.01742-06

Coburn, B., Sekirov, I., & Finlay, B. B. (2007). Type III Secretion Systems and Disease. *Clinical Microbiology Reviews*, *20*(4), 535–549. https://doi.org/10.1128/CMR.00013-07

Commichau, F. M., Gunka, K., Landmann, J. J., & Stülke, J. (2008). Glutamate Metabolism in Bacillus subtilis: Gene Expression and Enzyme Activities Evolved to Avoid Futile Cycles and to Allow Rapid Responses to Perturbations of the System. *Journal of Bacteriology*, *190*(10), 3557–3564. https://doi.org/10.1128/JB.00099-08

Commissioner, O. of the. (2019, September 10). *FDA approves first drug comprised of an active ingredient derived from marijuana to treat rare, severe forms of epilepsy*. FDA. http://www.fda.gov/news-events/press-announcements/fda-approves-first-drug-comprised-active-ingredient-derived-marijuana-treat-rare-severe-forms

Concannon, R., Finn, D. P., & Dowd, E. (2015). Chapter 3—Cannabinoids in Parkinson's disease. In L. Fattore (Ed.), *Cannabinoids in Neurologic and Mental Disease* (pp. 35–59). Academic Press. https://doi.org/10.1016/B978-0-12-417041-4.00003-5

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., &

Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, *17*(1), 13. https://doi.org/10.1186/s13059-016-0881-8

Connolly, J. P. R., Finlay, B. B., & Roe, A. J. (2015). From ingestion to colonization: The influence of the host environment on regulation of the LEE encoded type III secretion system in enterohaemorrhagic Escherichia coli. *Frontiers in Microbiology*, *6*. https://doi.org/10.3389/fmicb.2015.00568

Costiniuk, C. T., & Jenabian, M.-A. (2019). Cannabinoids and inflammation: Implications for people living with HIV. *AIDS (London, England)*, *33*(15), 2273–2288. https://doi.org/10.1097/QAD.0000000000002345

Dai, Z.-L., Li, X.-L., Xi, P.-B., Zhang, J., Wu, G., & Zhu, W.-Y. (2013). L-Glutamine regulates amino acid utilization by intestinal bacteria. *Amino Acids*, *45*(3), 501–512. https://doi.org/10.1007/s00726-012-1264-4

de Ceballos, M. L. (2015). Chapter 1—Cannabinoids for the treatment of neuroinflammation. In L. Fattore (Ed.), *Cannabinoids in Neurologic and Mental Disease* (pp. 3–14). Academic Press. https://doi.org/10.1016/B978-0-12-417041-4.00001-1

De Novo Assembly Using Illumina Reads. (2009) Illumina. 8. https://www.illumina.com/Documents/products/technotes/technote_denovo_assembly_ecoli.pdf

Defoirdt, T. (2018). Quorum-Sensing Systems as Targets for Antivirulence Therapy. *Trends in Microbiology*, *26*(4), 313–328. https://doi.org/10.1016/j.tim.2017.10.005

Deng, W., Marshall, N. C., Rowland, J. L., McCoy, J. M., Worrall, L. J., Santos, A. S., Strynadka, N. C. J., & Finlay, B. B. (2017). Assembly, structure, function and regulation of type III secretion systems. *Nature Reviews Microbiology*, *15*(6), 323–337. https://doi.org/10.1038/nrmicro.2017.20

DiPatrizio, N. V. (2016). Endocannabinoids in the Gut. *Cannabis and Cannabinoid Research*, *1*(1), 67–77. https://doi.org/10.1089/can.2016.0001

DiPatrizio, N. V., & Piomelli, D. (2015). Intestinal lipid–derived signals that sense dietary fat. *The Journal of Clinical Investigation*, *125*(3), 891–898. https://doi.org/10.1172/JCI76302

Dundar, W. F., Skrabanek, L., & Zumbo, P. (n.d.). *Introduction to differential gene expression analysis using RNA-seq*. 97.

Elliott, S. J., Hutcheson, S. W., Dubois, M. S., Mellies, J. L., Wainwright, L. A., Batchelor, M., Frankel, G., Knutton, S., & Kaper, J. B. (1999). Identification of CesT, a chaperone for the type III secretion of Tir in enteropathogenic Escherichia coli. *Molecular Microbiology*, *33*(6), 1176–1189. https://doi.org/10.1046/j.1365-2958.1999.01559.x

Elliott, Simon J., Sperandio, V., Girón, J. A., Shin, S., Mellies, J. L., Wainwright, L., Hutcheson, S. W., McDaniel, T. K., & Kaper, J. B. (2000). The Locus of Enterocyte Effacement (LEE)-Encoded Regulator Controls Expression of Both LEE- and Non-LEE-Encoded Virulence Factors in Enteropathogenic and Enterohemorrhagic Escherichia *coli*. *Infection and Immunity*, *68*(11), 6115–6126.

Elliott, Simon J., Wainwright, L. A., McDaniel, T. K., Jarvis, K. G., Deng, Y., Lai, L.-C., McNamara, B. P., Donnenberg, M. S., & Kaper, J. B. (1998). The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69. *Molecular Microbiology*, *28*(1), 1–4. https://doi.org/10.1046/j.1365-2958.1998.00783.x

Endo, Y., Tsurugi, K., Yutsudo, T., Takeda, Y., Ogasawara, T., & Igarashi, K. (1988). Site of action of a Vero toxin (VT2) from Escherichia coli O157:H7 and of Shiga toxin on eukaryotic ribosomes. RNA N-glycosidase activity of the toxins. *European Journal of Biochemistry*, *171*(1–2), 45–50. https://doi.org/10.1111/j.1432-1033.1988.tb13756.x

Ergonul, Z., Clayton, F., Fogo, A. B., & Kohan, D. E. (2003). Shigatoxin-1 binding and receptor expression in human kidneys do not change with age. *Pediatric Nephrology*, *18*(3), 246–253. https://doi.org/10.1007/s00467-002-1025-9

Feehily, C., & Karatzas, K. a. G. (2013). Role of glutamate metabolism in bacterial responses towards acid and other stresses. *Journal of Applied Microbiology*, *114*(1), 11–24. https://doi.org/10.1111/j.1365-2672.2012.05434.x

Gaytán, M. O., Martínez-Santos, V. I., Soto, E., & González-Pedrajo, B. (2016). Type Three Secretion System in Attaching and Effacing Pathogens.

Frontiers in Cellular and Infection Microbiology, *6*. https://doi.org/10.3389/fcimb.2016.00129

Girgis, H. S., Harris, K., & Tavazoie, S. (2012). Large mutational target size for rapid emergence of bacterial persistence. *Proceedings of the National Academy of Sciences*, *109*(31), 12740–12745. https://doi.org/10.1073/pnas.1205124109

Good, P., Haywood, A., Gogna, G., Martin, J., Yates, P., Greer, R., & Hardy, J. (2019). Oral medicinal cannabinoids to relieve symptom burden in the palliative care of patients with advanced cancer: A double-blind, placebo controlled, randomised clinical trial of efficacy and safety of cannabidiol (CBD). *BMC Palliative Care*, *18*(1), 110. https://doi.org/10.1186/s12904-019-0494-6

Grimes, T., Potter, S. S., & Datta, S. (2019). Integrating gene regulatory pathways into differential network analysis of gene expression data. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-41918-3

Hao, K., Lin, B., Nian, F., Gao, X., Wei, Z., Luo, G., Lu, Y., Lan, M., Yang, J., & Wu, G. (2019). RNA-seq analysis of the response of plant-pathogenic oomycete Phytophthora parasitica to the fungicide dimethomorph. *Revista Argentina de Microbiología*, *51*(3), 268–277. https://doi.org/10.1016/j.ram.2018.08.007

Hartland, E. L., Batchelor, M., Delahay, R. M., Hale, C., Matthews, S., Dougan, G., Knutton, S., Connerton, I., & Frankel, G. (1999). Binding of intimin from enteropathogenic Escherichia coli to Tir and to host cells. *Molecular Microbiology*, *32*(1), 151–158. https://doi.org/10.1046/j.1365-2958.1999.01338.x

Hartland, E. L., & Leong, J. M. (2013). Enteropathogenic and enterohemorrhagic E. coli: Ecology, pathogenesis, and evolution. *Frontiers in Cellular and Infection Microbiology*, *3*. https://doi.org/10.3389/fcimb.2013.00015

Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., & Ishii, K. (2018). *Escherichia coli O157:H7 str. Sakai DNA, complete genome* (1447699251). NCBI Nucleotide Database. http://www.ncbi.nlm.nih.gov/nuccore/NC_002695.2

Hayden, J. D., & Ades, S. E. (2008). The extracytoplasmic stress factor, sigmaE, is required to maintain cell envelope integrity in Escherichia coli. *PloS One*, *3*(2), e1573. https://doi.org/10.1371/journal.pone.0001573

Hughes, D. T., Clarke, M. B., Yamamoto, K., Rasko, D. A., & Sperandio, V. (2009). The QseC Adrenergic Signaling Cascade in Enterohemorrhagic E. coli (EHEC). *PLOS Pathogens*, *5*(8), e1000553. https://doi.org/10.1371/journal.ppat.1000553

Hughes, D. T., & Sperandio, V. (2008). Inter-kingdom signalling: Communication between bacteria and their hosts. *Nature Reviews. Microbiology*, *6*(2), 111–120. https://doi.org/10.1038/nrmicro1836

Hughes, E. R., Winter, M. G., Duerkop, B. A., Spiga, L., Carvalho, T. F. de, Zhu, W., Gillis, C. C., Büttner, L., Smoot, M. P., Behrendt, C. L., Cherry, S., Santos, R. L., Hooper, L. V., & Winter, S. E. (2017). Microbial Respiration and Formate Oxidation as Metabolic Signatures of Inflammation-Associated Dysbiosis. *Cell Host & Microbe*, *21*(2), 208–219. https://doi.org/10.1016/j.chom.2017.01.005

Ibeas Bih, C., Chen, T., Nunn, A. V. W., Bazelot, M., Dallas, M., & Whalley, B. J. (2015). Molecular Targets of Cannabidiol in Neurological Disorders. *Neurotherapeutics*, *12*(4), 699–730. https://doi.org/10.1007/s13311-015-0377-3

Imamovic, L., Ballesté, E., Martínez-Castillo, A., García-Aljaro, C., & Muniesa, M. (2016). Heterogeneity in phage induction enables the survival of the lysogenic population. *Environmental Microbiology*, *18*(3), 957–969. https://doi.org/10.1111/1462-2920.13151

Izzo, A. A., & Sharkey, K. A. (2010). Cannabinoids and the gut: New developments and emerging concepts. *Pharmacology & Therapeutics*, *126*(1), 21–38. https://doi.org/10.1016/j.pharmthera.2009.12.005

Jaskowiak, P. A., Costa, I. G., & Campello, R. J. G. B. (2018). Clustering of RNA-Seq samples: Comparison study on cancer data. *Methods*, *132*, 42–49. https://doi.org/10.1016/j.ymeth.2017.07.023

Jung, K., Fried, L., Behr, S., & Heermann, R. (2012). Histidine kinases and response regulators in networks. *Current Opinion in Microbiology*, *15*(2), 118–124. https://doi.org/10.1016/j.mib.2011.11.009

Kanamaru, K., Kanamaru, K., Tatsuno, I., Tobe, T., & Sasakawa, C. (2000). SdiA, an Escherichia coli homologue of quorum-sensing regulators, controls the expression of virulence factors in enterohaemorrhagic Escherichia coli O157:H7. *Molecular Microbiology*, *38*(4), 805–816. https://doi.org/10.1046/j.1365-2958.2000.02171.x

Karoly, H. C., Mueller, R. L., Bidwell, L. C., & Hutchison, K. E. (2019). Cannabinoids and the Microbiota-Gut-Brain Axis: Emerging Effects of Cannabidiol and Potential Applications to Alcohol Use Disorders. *Alcoholism, Clinical and Experimental Research*. https://doi.org/10.1111/acer.14256

Kim, Daehwan and Langmead, Ben and Salzberg, Steven L (2015). HISAT: a fast spliced aligner with low memory requirements. In *Nature Methods, 12 (4), pp. 357-360.* [doi:10.1038/nmeth.3317][Link]

Kumar, A., Ellermann, M., & Sperandio, V. (2019). Taming the Beast: Interplay between Gut Small Molecules and Enteric Pathogens. *Infection and Immunity*, *87*(9). https://doi.org/10.1128/IAI.00131-19

Laprairie, R. B., Bagher, A. M., Kelly, M. E. M., & Denovan-Wright, E. M. (2015). Cannabidiol is a negative allosteric modulator of the cannabinoid CB1 receptor. *British Journal of Pharmacology*, *172*(20), 4790–4805. https://doi.org/10.1111/bph.13250

Law, C. W., Alhamdoosh, M., Su, S., Dong, X., Tian, L., Smyth, G. K., & Ritchie, M. E. (2018). RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Research*, *5*. https://doi.org/10.12688/f1000research.9005.3

Lee, Y., Jo, J., Chung, H. Y., Pothoulakis, C., & Im, E. (2016). Endocannabinoids in the gastrointestinal tract. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, *311*(4), G655–G666. https://doi.org/10.1152/ajpgi.00294.2015

Lopez, C. A., Miller, B. M., Rivera-Chavez, F., Velazquez, E. M., Byndloss, M. X., Chavez-Arroyo, A., Lokken, K. L., Tsolis, R. M., Winter, S. E., & Baumler, A. J. (2016). Virulence factors enhance Citrobacter rodentium expansion through aerobic respiration. *Science*, *353*(6305), 1249–1253. https://doi.org/10.1126/science.aag3042

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

MacRitchie, D. M., Ward, J. D., Nevesinjac, A. Z., & Raivio, T. L. (2008). Activation of the Cpx Envelope Stress Response Down-Regulates Expression of Several Locus of Enterocyte Effacement-Encoded Genes in Enteropathogenic Escherichia coli. *Infection and Immunity*, *76*(4), 1465–1475. https://doi.org/10.1128/IAI.01265-07

Marsicano, G., Pagotto, U., Cervino, C., & Pasquali, R. (2009). Energy Homeostasis: Endocannabinoid System. In L. R. Squire (Ed.), *Encyclopedia of Neuroscience* (pp. 1021–1028). Academic Press. https://doi.org/10.1016/B978-008045046-9.00459-9

Matzov, D., Bashan, A., Yap, M.-N. F., & Yonath, A. (2019). Stress response as implemented by hibernating ribosomes: A structural overview. *The FEBS Journal*, *286*(18), 3558–3565. https://doi.org/10.1111/febs.14968

Mayer, C. L., Leibowitz, C. S., Kurosawa, S., & Stearns-Kurosawa, D. J. (2012). Shiga Toxins and the Pathophysiology of Hemolytic Uremic Syndrome in Humans and Animals. *Toxins*, *4*(11), 1261–1287. https://doi.org/10.3390/toxins4111261

McDaniel, T. K., Jarvis, K. G., Donnenberg, M. S., & Kaper, J. B. (1995). A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proceedings of the National Academy of Sciences*, *92*(5), 1664–1668. https://doi.org/10.1073/pnas.92.5.1664

McGannon, C. M., Fuller, C. A., & Weiss, A. A. (2010). Different Classes of Antibiotics Differentially Influence Shiga Toxin Production. *Antimicrobial Agents and Chemotherapy*, *54*(9), 3790–3798. https://doi.org/10.1128/AAC.01783-09

McGuire, P., Robson, P., Cubala, W. J., Vasile, D., Morrison, P. D., Barron, R., Taylor, A., & Wright, S. (2018). Cannabidiol (CBD) as an Adjunctive Therapy in Schizophrenia: A Multicenter Randomized Controlled Trial. *The American Journal of Psychiatry*, *175*(3), 225–231. https://doi.org/10.1176/appi.ajp.2017.17030325

McPartland, J. M., Duncan, M., Di Marzo, V., & Pertwee, R. G. (2015). Are cannabidiol and Δ9-tetrahydrocannabivarin negative modulators of the endocannabinoid system? A systematic review. *British Journal of Pharmacology*, *172*(3), 737–753. https://doi.org/10.1111/bph.12944

Mellies, J. L., Barron, A. M. S., & Carmona, A. M. (2007). Enteropathogenic and Enterohemorrhagic Escherichia coli Virulence Gene Regulation. *Infection and Immunity*, *75*(9), 4199–4210. https://doi.org/10.1128/IAI.01927-06

Mellies, J. L., Elliott, S. J., Sperandio, V., Donnenberg, M. S., & Kaper, J. B. (1999). The Per regulon of enteropathogenic Escherichia coli: Identification of a regulatory cascade and a novel transcriptional activator, the locus of enterocyte effacement (LEE)-encoded regulator (Ler). *Molecular Microbiology*, *33*(2), 296–306. https://doi.org/10.1046/j.1365-2958.1999.01473.x

Melton-Celsa, A. R. (2014). Shiga Toxin (Stx) Classification, Structure, and Function. *Microbiology Spectrum*, *2*(2). https://doi.org/10.1128/microbiolspec.EHEC-0024-2013

Milne-Davies, B., Helbig, C., Wimmi, S., Cheng, D. W. C., Paczia, N., & Diepold, A. (2019). Life After Secretion—Yersinia enterocolitica Rapidly Toggles Effector Secretion and Can Resume Cell Division in Response to Changing External Conditions. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.02128

Mittal, R., Debs, L. H., Patel, A. P., Nguyen, D., Patel, K., O'Connor, G., Grati, M., Mittal, J., Yan, D., Eshraghi, A. A., Deo, S. K., Daunert, S., & Liu, X. Z. (2017). Neurotransmitters: The critical modulators regulating gut-brain axis. *Journal of Cellular Physiology*, *232*(9), 2359–2372. https://doi.org/10.1002/jcp.25518

Moreira, C. G., Russell, R., Mishra, A. A., Narayanan, S., Ritchie, J. M., Waldor, M. K., Curtis, M. M., Winter, S. E., Weinshenker, D., & Sperandio, V. (2016). Bacterial Adrenergic Sensors Regulate Virulence of Enteric Pathogens in the Gut. *MBio*, *7*(3). https://doi.org/10.1128/mBio.00826-16

Mühldorfer, I., Hacker, J., Keusch, G. T., Acheson, D. W., Tschäpe, H., Kane, A. V., Ritter, A., Olschläger, T., & Donohue-Rolfe, A. (1996). Regulation of the

Shiga-like toxin II operon in Escherichia coli. *Infection and Immunity*, *64*(2), 495–502.

Müller, D., Benz, I., Liebchen, A., Gallitz, I., Karch, H., & Schmidt, M. A. (2009). Comparative Analysis of the Locus of Enterocyte Effacement and Its Flanking Regions. *Infection and Immunity*, *77*(8), 3501–3513. https://doi.org/10.1128/IAI.00090-09

Nguyen, Y., & Sperandio, V. (2012). Enterohemorrhagic E. coli (EHEC) pathogenesis. *Frontiers in Cellular and Infection Microbiology*, *2*, 90. https://doi.org/10.3389/fcimb.2012.00090

Nisco, N. J. D., Rivera-Cancel, G., & Orth, K. (2018). The Biochemistry of Sensing: Enteric Pathogens Regulate Type III Secretion in Response to Environmental and Host Cues. *MBio*, *9*(1), e02122-17. https://doi.org/10.1128/mBio.02122-17

Njoroge, J. W., Nguyen, Y., Curtis, M. M., Moreira, C. G., & Sperandio, V. (2012). Virulence Meets Metabolism: Cra and KdpE Gene Regulation in Enterohemorrhagic Escherichia coli. *MBio*, *3*(5). https://doi.org/10.1128/mBio.00280-12

Noble, W. S. (2009). A Quick Guide to Organizing Computational Biology Projects. *PLOS Computational Biology*, *5*(7), e1000424. https://doi.org/10.1371/journal.pcbi.1000424

Noris, M., & Remuzzi, G. (2005). Hemolytic Uremic Syndrome. *Journal of the American Society of Nephrology*, *16*(4), 1035–1050. https://doi.org/10.1681/ASN.2004100861

Novy, B. (2019). *Mellow microbes: Cannabidiol downregulates pathogenesis in type-three system pathogens*. Reed College.

Obrig, T. G., Louise, C. B., Lingwood, C. A., Boyd, B., Barley-Maloney, L., & Daniel, T. O. (1993). Endothelial heterogeneity in Shiga toxin receptors and responses. *Journal of Biological Chemistry*, *268*(21), 15484–15488.

Omisakin, F., MacRae, M., Ogden, I. D., & Strachan, N. J. C. (2003). Concentration and Prevalence of Escherichia coli O157 in Cattle Feces at Slaughter. *Applied and Environmental Microbiology*, *69*(5), 2444–2447. https://doi.org/10.1128/AEM.69.5.2444-2447.2003

Pacheco, A. R., & Sperandio, V. (2012). Shiga toxin in enterohemorrhagic E.coli: Regulation and novel anti-virulence strategies. *Frontiers in Cellular and Infection Microbiology*, *2*. https://doi.org/10.3389/fcimb.2012.00081

Panos, G. Z., Betsi, G. I., & Falagas, M. E. (2006). Systematic review: Are antibiotics detrimental or beneficial for the treatment of patients with Escherichia coli O157:H7 infection? *Alimentary Pharmacology & Therapeutics*, *24*(5), 731–742. https://doi.org/10.1111/j.1365-2036.2006.03036.x

Peng, X., Emiliani, F., Smallwood, P. M., Rattner, A., Lei, H., Sabbagh, M. F., & Nathans, J. (2018). Affinity capture of polyribosomes followed by RNAseq (ACAPseq), a discovery platform for protein-protein interactions. *ELife*, *7*, e40982. https://doi.org/10.7554/eLife.40982

Pertwee, R. G. (2008). The diverse CB1 and CB2 receptor pharmacology of three plant cannabinoids: Δ9-tetrahydrocannabinol, cannabidiol and Δ9-tetrahydrocannabivarin. *British Journal of Pharmacology*, *153*(2), 199–215. https://doi.org/10.1038/sj.bjp.0707442

Pifer, R., Russell, R. M., Kumar, A., Curtis, M. M., & Sperandio, V. (2018). Redox, amino acid, and fatty acid metabolism intersect with bacterial virulence in the gut. *Proceedings of the National Academy of Sciences*, *115*(45), E10712–E10719. https://doi.org/10.1073/pnas.1813451115

Pruimboom-Brees, I. M., Morgan, T. W., Ackermann, M. R., Nystrom, E. D., Samuel, J. E., Cornick, N. A., & Moon, H. W. (2000). Cattle lack vascular receptors for Escherichia coli O157:H7 Shiga toxins. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(19), 10325–10329.

Raivio, T. L. (2005). MicroReview: Envelope stress responses and Gram-negative bacterial pathogenesis. *Molecular Microbiology*, *56*(5), 1119–1128. https://doi.org/10.1111/j.1365-2958.2005.04625.x

Raivio, T. L., Leblanc, S. K. D., & Price, N. L. (2013). The Escherichia coli Cpx Envelope Stress Response Regulates Genes of Diverse Function That Impact Antibiotic Resistance and Membrane Integrity. *Journal of Bacteriology*, *195*(12), 2755–2767. https://doi.org/10.1128/JB.00105-13

Reading, N. C., Rasko, D. A., Torres, A. G., & Sperandio, V. (2009). The two-component system QseEF and the membrane protein QseG link adrenergic and stress sensing to bacterial pathogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(14), 5889–5894. https://doi.org/10.1073/pnas.0811409106

Reitsma Pieter H., Versteeg Henri H., & Middeldorp Saskia. (2012). Mechanistic View of Risk Factors for Venous Thromboembolism. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *32*(3), 563–568. https://doi.org/10.1161/ATVBAHA.111.242818

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J. L., & Pachter, L. (2011). Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biology*, *12*(3), R22. https://doi.org/10.1186/gb-2011-12-3-r22

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Roy, S., Bhattacharyya, D. K., & Kalita, J. K. (2014). Reconstruction of gene co-expression network from microarray data using local expression patterns. *BMC Bioinformatics*, *15*(Suppl 7), S10. https://doi.org/10.1186/1471-2105-15-S7-S10

Sergushichev, A. A. (2016). An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *BioRxiv*, 060012. https://doi.org/10.1101/060012

Sharkey, K. A., & Wiley, J. W. (2016). The role of the endocannabinoid system in the brain-gut axis. *Gastroenterology*, *151*(2), 252–266. https://doi.org/10.1053/j.gastro.2016.04.015

Shenoy, A. R., Furniss, R. C. D., Goddard, P. J., & Clements, A. (2018). Modulation of Host Cell Processes by T3SS Effectors. In G. Frankel & E. Z. Ron (Eds.), *Escherichia coli, a Versatile Pathogen* (pp. 73–115). Springer International Publishing. https://doi.org/10.1007/82_2018_106

Skelley, J. W., Deas, C. M., Curren, Z., & Ennis, J. (2019). Use of cannabidiol in anxiety and anxiety-related disorders. *Journal of the American Pharmacists Association: JAPhA*. https://doi.org/10.1016/j.japh.2019.11.008

Slater, S. L., Sågfors, A. M., Pollard, D. J., Ruano-Gallego, D., & Frankel, G. (2018). The Type III Secretion System of Pathogenic Escherichia coli. In G. Frankel & E. Z. Ron (Eds.), *Escherichia coli, a Versatile Pathogen* (pp. 51–72). Springer International Publishing. https://doi.org/10.1007/82_2018_116

Slutsker, L., Ries, A. A., Greene, K. D., Wells, J. G., Hutwagner, L., & Griffin, P. M. (1997). Escherichia coli O157:H7 diarrhea in the United States: Clinical and epidemiologic features. *Annals of Internal Medicine*, *126*(7), 505–513. https://doi.org/10.7326/0003-4819-126-7-199704010-00002

Slutsker, L., Ries, A. A., Maloney, K., Wells, J. G., Greene, K. D., & Griffin, P. M. (1998). A nationwide case-control study of Escherichia coli O157:H7 infection in the United States. *The Journal of Infectious Diseases*, *177*(4), 962–966. https://doi.org/10.1086/515258

Song, S., & Wood, T. K. (2019). PpGpp Ribosome Dimerization Model for Bacterial Persister Formation and Resuscitation. *BioRxiv*, 663658. https://doi.org/10.1101/663658

Sperandio, V., Mellies, J. L., Delahay, R. M., Frankel, G., Crawford, J. A., Nguyen, W., & Kaper, J. B. (2000). Activation of enteropathogenic Escherichia coli (EPEC) LEE2 and LEE3 operons by Ler. *Molecular Microbiology*, *38*(4), 781–793. https://doi.org/10.1046/j.1365-2958.2000.02168.x

Sperandio, V., Torres, A. G., & Kaper, J. B. (2002). Quorum sensing Escherichia coli regulators B and C (QseBC): A novel two-component regulatory system involved in the regulation of flagella and motility by quorum sensing in E. coli. *Molecular Microbiology*, *43*(3), 809–821. https://doi.org/10.1046/j.1365-2958.2002.02803.x

Spoering, A. L., & Lewis, K. (2001). Biofilms and Planktonic Cells of Pseudomonas aeruginosa Have Similar Resistance to Killing by Antimicrobials. *Journal of Bacteriology*, *183*(23), 6746–6751. https://doi.org/10.1128/JB.183.23.6746-6751.2001

Sudmant, P. H., Alexis, M. S., & Burge, C. B. (2015). Meta-analysis of RNA-seq expression data across species, tissues and studies. *Genome Biology*, *16*(1), 287. https://doi.org/10.1186/s13059-015-0853-4

Suo, Y., Gao, S., Baranzoni, G. M., Xie, Y., & Liu, Y. (2018). Comparative transcriptome RNA-Seq analysis of Listeria monocytogenes with sodium lactate adaptation. *Food Control*, *91*, 193–201. https://doi.org/10.1016/j.foodcont.2018.03.044

Tilden, J., Young, W., McNamara, A. M., Custer, C., Boesel, B., Lambert-Fair, M. A., Majkowski, J., Vugia, D., Werner, S. B., Hollingsworth, J., & Morris, J. G. (1996). A new route of transmission for Escherichia coli: Infection from dry fermented salami. *American Journal of Public Health*, *86*(8 Pt 1), 1142–1145.

Tobe, T., Beatson, S. A., Taniguchi, H., Abe, H., Bailey, C. M., Fivian, A., Younis, R., Matthews, S., Marches, O., Frankel, G., Hayashi, T., & Pallen, M. J. (2006). An extensive repertoire of type III secretion effectors in Escherichia coli O157 and the role of lambdoid phages in their dissemination. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(40), 14941–14946. https://doi.org/10.1073/pnas.0604891103

Trautmann, S. M., & Sharkey, K. A. (2015). The Endocannabinoid System and Its Role in Regulating the Intrinsic Neural Circuitry of the Gastrointestinal Tract. *International Review of Neurobiology*, *125*, 85–126. https://doi.org/10.1016/bs.irn.2015.10.002

Tsuyuzaki, K., Ishii, M., & Nikaido, I. (2019). Uncovering hypergraphs of cell-cell interaction from single cell RNA-sequencing data. *BioRxiv*, 566182. https://doi.org/10.1101/566182

Tyler, J. S., Mills, M. J., & Friedman, D. I. (2004). The operator and early promoter region of the Shiga toxin type 2-encoding bacteriophage 933W and control of toxin expression. *Journal of Bacteriology*, *186*(22), 7670–7679. https://doi.org/10.1128/JB.186.22.7670-7679.2004

Https://www.usda.gov/sites/default/files/documents/2018-farm-bill-and-legislative-principles.pdf, (testimony of USDA).

Ueta, M., Yoshida, H., Wada, C., Baba, T., Mori, H., & Wada, A. (2005). Ribosome binding proteins YhbH and YfiA have opposite functions during 100S formation in the stationary phase of Escherichia coli. *Genes to Cells*, *10*(12), 1103–1112. https://doi.org/10.1111/j.1365-2443.2005.00903.x

UniProt: A worldwide hub of protein knowledge. (2019). *Nucleic Acids Research*, *47*(D1), D506–D515. https://doi.org/10.1093/nar/gky1049

Vallance, B. A., & Finlay, B. B. (2000). Exploitation of host cells by enteropathogenic Escherichia coli. *Proceedings of the National Academy of Sciences*, *97*(16), 8799–8806. https://doi.org/10.1073/pnas.97.16.8799

Ventola, C. (2015). The Antibiotic Resistance Crisis. *Pharmacy and Therapeutics*, *40*(4), 277–283.

Walters, M., & Sperandio, V. (2006). Autoinducer 3 and epinephrine signaling in the kinetics of locus of enterocyte effacement gene expression in enterohemorrhagic Escherichia coli. *Infection and Immunity*, *74*(10), 5445–5455. https://doi.org/10.1128/IAI.00099-06

Wang, G., Zhao, T., & Doyle, M. P. (1996). Fate of enterohemorrhagic Escherichia coli O157:H7 in bovine feces. *Applied and Environmental Microbiology*, *62*(7), 2567–2570.

Wang, H., Yu, Q., Ding, X., Hu, X., Hou, K., Liu, X., Nie, S., & Xie, M. (2019). RNA-seq based elucidation of mechanism underlying Ganoderma atrum polysaccharide induced immune activation of murine myeloid-derived dendritic cells. *Journal of Functional Foods*, *55*, 104–116. https://doi.org/10.1016/j.jff.2019.02.022

Weirauch, M. T. (2011). Gene Coexpression Networks for the Analysis of DNA Microarray Data. In *Applied Statistics for Network Biology* (pp. 215–250). John Wiley & Sons, Ltd. https://doi.org/10.1002/9783527638079.ch11

White, C. M. (2019). A Review of Human Studies Assessing Cannabidiol's (CBD) Therapeutic Actions and Potential. *The Journal of Clinical Pharmacology*, *59*(7), 923–934. https://doi.org/10.1002/jcph.1387

Wong, A. R. C., Pearson, J. S., Bright, M. D., Munera, D., Robinson, K. S., Lee, S. F., Frankel, G., & Hartland, E. L. (2011). Enteropathogenic and enterohaemorrhagic Escherichia coli: Even more subversive elements.

Molecular Microbiology, *80*(6), 1420–1438. https://doi.org/10.1111/j.1365-2958.2011.07661.x

Wong, C. S., Mooney, J. C., Brandt, J. R., Staples, A. O., Jelacic, S., Boster, D. R., Watkins, S. L., & Tarr, P. I. (2012). Risk Factors for the Hemolytic Uremic Syndrome in Children Infected With Escherichia coli O157:H7: A Multivariable Analysis. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, *55*(1), 33–41. https://doi.org/10.1093/cid/cis299

Zhang, X., McDaniel, A. D., Wolf, L. E., Keusch, G. T., Waldor, M. K., & Acheson, D. W. K. (2000). Quinolone Antibiotics Induce Shiga Toxin-Encoding Bacteriophages, Toxin Production, and Death in Mice. *The Journal of Infectious Diseases*, *181*(2), 664–670. https://doi.org/10.1086/315239

Zhao, Y., Qin, X., Yang, J., Liao, Y., Wu, X., & Zheng, H. (2019). Identification and expression analysis of ceftriaxone resistance-related genes in Neisseria gonorrhoeae integrating RNA-Seq data and qRT-PCR validation. *Journal of Global Antimicrobial Resistance*, *16*, 202–209. https://doi.org/10.1016/j.jgar.2018.10.008

Zhu, M., & Dai, X. (2020). Bacterial stress defense: The crucial role of ribosome speed. *Cellular and Molecular Life Sciences*, *77*(5), 853–858. https://doi.org/10.1007/s00018-019-03304-0

Zoja, C., Angioletti, S., Donadelli, R., Zanchi, C., Tomasoni, S., Binda, E., Imberti, B., te Loo, M., Monnens, L., Remuzzi, G., & Morigi, M. (2002). Shiga toxin-2 triggers endothelial leukocyte adhesion and transmigration via NF-kappaB dependent up-regulation of IL-8 and MCP-1. *Kidney International*, *62*(3), 846–856. https://doi.org/10.1046/j.1523-1755.2002.00503.x

Zoja, C., Buelli, S., & Morigi, M. (2010). Shiga toxin-associated hemolytic uremic syndrome: Pathophysiology of endothelial dysfunction. *Pediatric Nephrology*, *25*(11), 2231–2240. https://doi.org/10.1007/s00467-010-1522-1