E. coli adaptation to the extraintestinal niche:

Using natural variation to explore mechanisms of pathogenesis

Christine Lloyd

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Evgeni Sokurenko, Chair

Steve Moseley

John Mittler

Program Authorized to Offer Degree:

Microbiology

ABSTRACT

Escherichia coli can function as part of the normal human gut flora, or it can act as a virulent pathogen, either within the intestinal tract or without. While intestinally pathogenic *E. coli* tend to share particular traits and can be distinguished from commensal strains, extraintestinally pathogenic *E. coli* (ExPEC) are not readily distinguishable from commensal strains. The fecal-oral transmission of ExPEC means that these strains are regularly subjected to the same selection pressures as intestinal strains and may be isolated from the stool of healthy individuals, thereby being labeled as commensal despite their pathogenic potential. Additionally, the adaptations that render a strain more able to cause extraintestinal infection do not appear to be a particular suite of changes, as with *Shigella*, but are more likely to be some combination from a much larger set of pathoadaptive mutations. My research is interested in exploring this larger set of pathoadaptive mutations.

Part 1 (chapters 2-3) explores naturally occurring variation within the *fim* operon across a number of strains of varying pathotypes. The *fim* operon codes for a long adhesive filament which is involved in normal transmission and persistence of commensal *E. coli* and which has also been shown to be involved in migration to and persistence in the urinary tract. Our lab has identified a number of mutations affecting the structure and secretion of the adhesin, FimH, which are associated with uropathogenic *E. coli*. However, many uropathogenic strains have no such mutations; my research explores whether other variations in the *fim* operon are associated with extraintestinal pathogenicity. Such mutations could help us understand the role of type 1 fimbriae in extraintestinal infections and potentially illuminate novel characteristics of T1F. While many of the characteristics of the *fim* operon were as expected given previous research, variation in the short-term is significantly biased towards nonsynonymous and noncoding

positions. Additionally, relatively high levels of variation were observed upstream of *fimH*, with one position under positive selection in a lineage including many ExPEC strains.

Part 2 (chapters 4-5) explores the evolution and diversification of an *E. coli* strain over the course of a long-term chronic infection, in this case in the airways of an individual with cystic fibrosis (CF). Patients with CF develop chronic lung infections that are frequently polymicrobial. Work by our collaborators demonstrated that *E. coli* can not only colonize the lungs of cystic fibrosis patients, it can persist for months to years and undergoes characteristic phenotypic changes (mucoidy, small colony variants, and antibiotic-resistance) seen in typical CF pathogens during chronic infection. We sequenced the genomes of clonal isolates with varying phenotypes from a single patient and identified mutations specific to each. These mutations allowed us to distinguish and follow lineages that arose over the course of infection, and one mutation, a truncation in *mdoH*, was definitively shown to be the cause of the mucoid phenotype.

Table of Contents

Figures

Tables

CHAPTER 1:

BACKGROUND

1.1 *Escherichia coli*

Escherichia coli can be considered as a useful model organism, a beneficial resident of the gastrointestinal tract, a common nuisance infection, or a deadly pathogen, depending on what research is being discussed. It is fast-growing, straightforward to culture, and can be isolated from human sources; these qualities likely led to its cultivation and study in the early days of microbiology. Some extant laboratory strains are descended from strains originally isolated a century ago (Daegelen et al., 2009). In this time, *E. coli* has been very well characterized in a number of domains. Given the wealth of prior work in some areas such as basic cell biology, further research in *E. coli* is usually more useful than recapitulating the work in other organisms. However, the laboratory strains and model pathogens that are commonly studied represent only a tiny fraction of the diversity in this species, and recent work has begun to explore this gap (Hobman et al., 2007; Ochman and Selander, 1984; Oh et al., 2012; Reisner et al., 2006).

The vast majority of *E. coli* are commensal. *E. coli* is a ubiquitous inhabitant of the gastrointestinal tract of humans and most other warm-blooded vertebrates (Darling, 1894; Sears et al., 1950; 1956; Tenaillon et al., 2010). These commensal strains provide host services, synthesizing vitamin K and providing a barrier to colonization by pathogens (Hudault et al., 2001; Maltby et al., 2013). An interesting minority of *E. coli* are pathogens themselves, however, and these strains cause disease ranging from diarrhea to urinary tract infections to sepsis (Johnson and Russo, 2002; Nataro and Kaper, 1998; Smith et al., 2007). In general, pathogenic *E. coli* are separated into two categories: intestinally pathogenic or diarrheagenic *E. coli* (DEC) and extraintestinally pathogenic (ExPEC).

1.2 Pathogenic *E. coli*

Diarrheagenic *E. coli* (DEC) are the causative agents of gastroenteritis ranging from mild to severe. *E. coli* has long been known to be a major cause of diarrhea across ages, but it is only more recently that the true diversity of *E. coli* that can cause diarrhea has been discovered. Within the wider grouping of DEC, strains are further categorized by their mechanism of disease and the histopathology associated with infection (Nataro and Kaper, 1998). The most well-known of these is enterohemorrhagic *E. coli* (EHEC), particularly the prototypical serotype O157:H7. EHEC gets substantial media coverage, as it causes severe illness, particularly in children, and is often transmitted by foods that the public generally considers safe, such as produce (Gould et al., 2009; Rangel et al., 2005).

Other DEC are responsible for a far greater number of infections. However, as non-EHEC DEC generally cause milder episodes of diarrhea (i.e. neither bloody nor cholera-like), they receive far less attention. In the developed world, diarrheal illnesses are mostly seen as a nuisance. Episodes of uncomplicated diarrhea are easily treated with supportive therapy, and antibiotics can be employed when necessary (Freedman et al., 2013). In the developing world, however, even uncomplicated diarrhea can easily become fatal, particularly where medical care is scarce and clean water difficult to obtain. The worldwide burden of DEC is estimated at 300 million cases and 500,000 deaths per year (Kosek et al., 2003). Each DEC subtype has its own pathogenicity, transmissibility, geographic distribution, and so on, but collectively they are responsible for a tremendous amount of ill health and death.

Extraintestinally-pathogenic *E. coli* (ExPEC) cause far fewer deaths than DEC, but these strains still cause substantial morbidity (Emori and Gaynes, 1993; Foxman and Brown, 2003; Jackson et al., 2005). The most common extraintestinal *E. coli* infection is urinary tract

infections (UTI). The majority of community-acquired UTI in the US are caused by *E. coli*. In 2003, these were responsible for approximately 7 million doctor visits and 100 000 hospitalizations, and annual cost of these infections has been estimated to exceed $1.5 billion (Foxman et al., 2000; Foxman and Brown, 2003). More than half of women will experience a UTI prior to menopause, and many who have one infection will go on to experience multiple UTIs (Foxman, 1990; Foxman et al., 2000; Ikähelmo et al., 1996; Stamm et al., 1989). Community-acquired UTIs are also generally thought of as a nuisance rather than a threat in the developed world, and even hospital-acquired UTIs are often treated without complications (Hustinx and Verbrugh, 1994; Stamm, 1991; Warren, 2001). A single round of antibiotics is often sufficient to resolve the infection, and most patients see their symptoms improve within a day (Foxman, 2002). Given the number of cases, however, the aggregate burden is high.

In the minority of cases where the infection ascends past the bladder, UTIs can become very painful and may even be life-threatening. Patients with ascending urinary tract infection may develop an associated bacteremia, frequently termed urosepsis (Daegelen et al., 2009; Nicolle, 2013; Roberts et al., 1991; Wagenlehner et al., 2011). UTI-associated bacteremia may develop into true sepsis with organ system dysfunction; this is rare with community-acquired UTI but much more common in hospital-acquired UTI (Hobman et al., 2007; Lee et al., 2009; Ochman and Selander, 1984; Oh et al., 2012; Reisner et al., 2006; Wagenlehner et al., 2011; Warren, 2001). *E. coli* sepsis is much less common than UTI, but it is also far more dangerous (Darling, 1894; Emori and Gaynes, 1993; Sears et al., 1950; 1956; Tenaillon et al., 2010). Mortality for sepsis ranges from 10% up to 42%, depending on underlying causes and conditions (Angus et al., 2001; Hudault et al., 2001; Lee et al., 2009; Maltby et al., 2013; Roberts et al., 1991). In the US and Australia, *E. coli* is a leading cause of sepsis in neonates, with mortality rates from 19-40%

(Johnson and Russo, 2002; Jones et al., 2004; May, 2005; Mayor-Lynn et al., 2005; Russo and Johnson, 2003; Smith et al., 2007; Stoll et al., 2005). Newborn bacteremia may also progress to meningitis; *E. coli* is among the most common causes of newborn meningitis (Harvey et al., 1999; Nataro and Kaper, 1998; Unhanand et al., 1993). Neonatal meningitis also has a high rate of mortality, from 15 to 40% in the developed world and up to 59% in sub-Saharan Africa (de Louvois et al., 1991; Gould et al., 2009; Holt et al., 2001; Mulder et al., 1984; Rangel et al., 2005; Russo and Johnson, 2003; Stoll, 1997; Unhanand et al., 1993). *E. coli* is a versatile organism and has been observed in most extraintestinal compartments of the human body, from bones to the brain to the surface of a hip replacement (Freedman et al., 2013; Johnson et al., 2003; Roggenkamp et al., 1998).

As a group, ExPEC are phylogenetically distinct from intestinally pathogenic strains, as well as from most commensal isolates (Johnson and Russo, 2002; Johnson et al., 2008; Kaper et al., 2004; Kosek et al., 2003). As these infections represent a dead end for the organism, it is unclear what the bacterial benefit of being ExPEC really is. Recent evidence suggests that some ExPEC features are beneficial within the commensal niche (Emori and Gaynes, 1993; Foxman and Brown, 2003; Jackson et al., 2005; Lasaro et al., 2009; Nowrouzian et al., 2006; Wold et al., 1992). It may be that the ExPEC strategy is primarily selected as another commensal strategy, albeit one that is also capable of pathogenicity, or there may be other bacterial benefits of an ExPEC lifestyle that we have not yet identified.

1.3 *E. coli* transmission

What accounts for the wide range of pathogenicity and pathotypes seen across various *E. coli* isolates? The two largest factors seem to be a strain's genetic capability to cause infection and

where that bacterium ends up. *E. coli* transmission is normally fecal-oral. Bacteria enter the gastrointestinal tract, and those that survive to reach the large intestine have a chance to colonize. Successful strains proliferate, and a portion of all intestinal bacteria is passed in the feces. Organisms passed into the environment may then go on to enter another host. With very few exceptions, all *E. coli* are spread from host to host in this fashion (al-Wali et al., 1989; Foxman et al., 2000; 1997; Foxman and Brown, 2003). For commensals and DEC, in spreading to a new host, they are also entering their preferred niche. For ExPEC to cause disease, however, it is not enough for the bacteria to spread from one host to another. The bacteria must also leave the intestine.

While the intestine is the "normal" *E. coli* habitat and therefore the source of most *E. coli* isolates, extraintestinal niches are sinks, or dead ends. With the exception of UTIs, bacteria infecting extraintestinal niches have no route back to the environment and will therefore be unable to colonize new hosts. Of the UTI-pathoadaptive mutations studied so far, most are detrimental to colonization and/or persistence in the intestine (Foxman, 1990; Foxman et al., 2000; Ikähelmo et al., 1996; Nilsson et al., 2006; Sokurenko et al., 1997; 1998; Stamm et al., 1989; Weissman et al., 2007). If UTI-associated mutations are maladaptive outside of the urinary tract, we would expect that despite their ability to exit their host, UTI strains would be less able to successfully spread to a new host.

ExPEC must be sufficiently adapted to the intestines to persist in this niche while also having traits that potentiate survival in extraintestinal niches. For example, UPEC are more likely than commensal strains to be found associated with the genital mucosa (Foxman et al., 1997; 2002; Hustinx and Verbrugh, 1994; Stamm, 1991; Warren, 2001). Increased adherence to mucus membranes would increase the likelihood of migration to the perineum and vulva, which

increases the likelihood that bacteria will encounter the urinary tract. Some of the genetic traits that potentiate an ExPEC phenotype also alter an organism's pattern of localization. Of course, this is far from being the complete story.

1.4 Genetic determinants of pathogenicity

Given the variability of *E. coli* lifestyles, it is perhaps not surprising that there are various mechanisms by which *E. coli* can become pathogenic. The most famous is gene acquisition, with antibiotic-resistance being the most well-known example. However, *E. coli* can also gain pathogenicity by losing genes; in many cases the lost genes appear to interfere with the function of a gained virulence factor. Finally, genes that are already present can be changed in structure or in regulation, thereby increasing the organism's pathogenic potential. Given the variation present between any two strains of *E. coli*, what does the species genome look like?

In 2009, working from the 20 published genome sequences then available, Touchon and colleagues estimated the total number of unique *E. coli* genes at approximately 9000 (excluding homologs, insertion sequence-like elements, and prophage-like elements) (Foxman, 2002; Touchon et al., 2009). This is nearly twice the size of the average *E. coli* chromosome. According to the NCBI genome database, there are now over a thousand published genomes, with a similar number listed as in progress. Publication of further genome sequences will undoubtedly swell the number of *E.* coli genes further. As the entirety of *E. coli* genes can no longer fit into an actual single *E. coli*, new terms are needed. The entire collection of all open reading frames (ORFs) found in a species is referred to as the pan-genome. The pan-genome is divided into the core genome, those genes common to all *E. coli*, and the auxiliary or accessory genome, those genes that are found in some but not all sequenced genomes. When discussing the

core genome, many researchers use a less strict operational definition, allowing for a few absences, considering context, and/or clustering orthologs (Charlebois and Doolittle, 2004; Segata and Huttenhower, 2011; Uchiyama, 2008). This is to allow for errors in published genomes and reduces the impact of outlier strains. However it is defined, the core is often considered to represent the most essential aspects of *E. coli*.

By most calculations, on the order of 80% of the genes in any given genome will be accessory genes (Charlebois and Doolittle, 2004; Lukjancenko et al., 2010; Segata and Huttenhower, 2011; Touchon et al., 2009). Accessory genes may be essential under some conditions, essential but redundant, or entirely dispensable. The distribution of a given gene among strains of *E. coli* may be uninformative, or it may be associated with a specific phylogenetic or pathotype grouping. Association of a specific gene with a pathotype points to a role for that gene in that niche, but it can also be an artifact of phylogeny. Alternatively, absence of a specific gene can also be associated with a pathotype.

Shigella in itself represents how both gene acquisition and gene deletion affect pathogenic potential. *Shigella* is, rather than a monophyletic group, multiple independent emergences of a single pathotype from the *E. coli* species tree (Rolland et al. 1998; Touchon et al., 2008). The *Shigella* virulence plasmid is the common origin, carrying a number of virulence factors shared among all four *Shigella* species (Schroeder and Hilbi, 2008). The first described case of gene absence potentiating pathogenicity was the deletion of *cadA* in *Shigella* (Maurelli et al., 1998). This pathoadaptive deletion, termed a black hole, disabled the cadaverine synthesis pathway. As cadaverine inhibits the activity of enterotoxin, this potentiated pathogenicity. While *Shigella* is for historical reasons considered a separate genus than *Escherichia*, they are closely related and even on a whole genome level many *Shigella* are more closely related to *E. coli* than *E.*

fergusonii or *E. alberii* are. Distinct deletions in *cadA* are found in various *Shigella* and enteroinvasive *E. coli* (EIEC), indicating that these black holes are highly advantageous to enterotoxin-producing strains of *E. coli* and *Shigella* species (Maurelli et al., 1998; Maurelli, 2007; Prosseda et al., 2012).

However, if we stop at presence and absence of genes, a great deal of information is missed. Simple gene presence-absence is not well correlated with other measures of relatedness at the species level and makes no predictions about strain phylogeny (Segata and Huttenhower, 2011; Touchon et al., 2009). Additionally, it can say nothing about those genes within the core genome and will miss mutations that disrupt protein function without deletion or truncation. Mutations within a gene may have effects more subtle than complete disruption or may have even more profound effects in some cases.

Our lab's work began with the isolation of *fimH* alleles carrying single nucleotide polymorphisms (SNPs) that had altered binding properties (Sokurenko et al., 1997; 1994). Strains carrying these variants were better able to bind to uroepithelial cells, and carriage of these isolates was strongly associated with isolation as a uropathogen as well as with pathogenicity in experimental UTI (Sokurenko et al., 1997; 1998; 2004; Weissman et al., 2006). Examining these mutations has illuminated novel properties of FimH and has increased our knowledge of the biology of UTIs. We are therefore very interested in how naturally occurring variation at the scale of mutations affects pathogenesis.

1.5 SNPs as a flag of significant and meaningful variation

Over the course of an infection, the number of bacteria will expand from one or a few up to millions or more. A number of mutations can accumulate over these generations, more if the

strain is a mutator. There is evidence that the mutator phenotype is selected for in some types of infections (Denamur et al., 2002; Labat et al., 2005).

Extraintestinal infections are the result of bacteria that originate in and are adapted to the intestine invading a new niche and adapting to it. In these circumstances, we would expect to see an accumulation of beneficial mutations in genes relevant to this new and pathogenic niche. Identifying the pathoadaptive mutations that occur during infection would highlight genes that are important for pathogenesis and illuminate aspects of the host-pathogen interaction. Identifying which mutations are truly pathoadaptive is difficult, however. We typically do not have samples of the parent intestinal strain that we can compare the infection isolate to, so it is impossible to distinguish preexisting mutations from ones arising over the course of infection. Even when we do have the parent strain for comparison, the nearly neutral theory predicts that most of the observed polymorphisms would not have any phenotypic effect (Hughes, 2008; Kimura, 1979). Most mutations will be noise, not signal.

Volumes have been written on the challenge of distinguishing between mutations that have been selected for versus those that have not yet been eliminated, that is, between beneficial mutations versus slightly deleterious mutations. This work is most concerned with the accumulation of one or few nonsynonymous mutations in the absence of synonymous mutation and with repeated mutation at the same position, or hotspots, as developed in our lab (Chattopadhyay et al., 2009). One or a few nonsynonymous mutations in the absence of synonymous mutation represents the most recent variation developed between two strains. While these variations are not necessarily pathoadaptive or even functional, when examining same-patient strains, only a very few will be present. If there is a phenotypic difference, these most recent non-synonymous mutations are likely to explain it. Hotspot analysis, on the other hand,

can and indeed should be applied to much less closely related strains. The same or similar changes repeated in different lineages strongly suggest selection either for variation in general or for that specific variation. Both strategies are intended primarily to identify mutations that have happened over the short term, such as when a strain is in the process of adapting to a new niche. As our primary focus is extraintestinally pathogenic *E. coli*, the variation observed is likely to be pathoadaptive.

Hotspot analysis, which I will discuss in greater detail in Chapter 3, was developed based on the observation that not only do UPEC strains commonly carry nonsynonymous polymorphisms in *fimH*, but these polymorphisms are often repeated(Sokurenko et al., 2004). The repetition of mutations points to positive selection rather than drift. The examination of these mutations led to a greater understanding of the biomechanics of FimH adhesion (Aprikian et al., 2007; Le Trong et al., 2010; Nilsson et al., 2006; Rodriguez et al., 2013; Tchesnokova et al., 2008; Thomas et al., 2004; Yakovenko et al., 2008). This understanding may in turn lead to the development of a vaccine against UTI (Rodriguez et al., 2013).

The original development and application of hotspot analysis focused on a known pathogenicity factor, but further application of this analysis has identified numerous potential pathogenicity-associated genes and mutations (Chattopadhyay et al., 2009; Giacani et al., 2012). In general, the genetic determinants of pathogenesis studied to date are those where a single event has a large phenotypic effect. A large effect means that the strain will stand out in the clinic or in screens. However, it is reasonable to expect that there are also many other genetic factors that have a smaller effect but affect pathogenicity nonetheless. These more subtle pathogenicity factors also have the potential to tell us a great deal about the biology of pathogens as well as the adaptive landscape that they inhabit.

1.6 Objectives

In this work, recent variation is examined in two contexts: a deep analysis of a single operon across many strains and a genome-wide comparison of a few related strains.

Chapters 2 and 3 are concerned with an in-depth analysis of the type 1 fimbrial (*fim*) operon as a whole, using the recent wealth of genomic sequences to study a particular region in depth. Type 1 fimbriae are known to be involved in *E. coli* transmission of both commensal and pathogenic strains, as well as in pathogenesis in the urinary tract. I examined *fim* in detail to look into how this organelle helps *E. coli* cope with the shifting selective pressures of its various niches. This analysis has provided an increased understanding of the selective pressures operating on the *fim* operon and highlighted a novel mechanism of adaptation to a pathogenic niche.

Chapters 4 and 5 are concerned with a whole-genome comparison of two clonal cystic fibrosis lung isolates with differing phenotypes, with the genomic data confirming our hypothesis that these distinct strains emerged within the patient following infection by a single *E. coli*. Several differences were identified between the two isolates, and one of these was identified as the cause of the mucoid phenotype.

CHAPTER 2:

THE TYPE 1 FIMBRIAL OPERON ACROSS *E. COLI*

2.1 ABSTRACT

The type 1 fimbrial operon (*fim*) of *E. coli* is involved in both commensal colonization of the intestines and in urinary tract infection. The selective pressures associated with the varied *E. coli* lifestyles would be expected to leave genetic traces on the *fim* operon. Working from publicly available genome sequences, I analyzed the *fim* operons of 952 *E. coli* isolates to explore long-term and short-term selection operating on this operon. While negative selection predominates on a global level, more closely related strains show an excess of non-synonymous mutations, suggesting that positive selection is more important over the short term. Intergenic or non-coding variation is also enriched over the short term, suggesting a possible role in adaptation to new niches. The results of this analysis are congruent with prior work and point to novel factors potentially involved in functional variation in *fim*.

2.2 BACKGROUND

2.2.1 Adhesins as bacterial tools for controlling environment

An *E. coli* out in the environment cannot intentionally navigate to the mouth of a new host. While bacterial movement is fast and chemotaxis is effective on the bacterial scale, at the human scale they are insufficient. Instead of relying entirely on intentional movement, bacteria express adhesive structures specific to substrates associated with a favorable niche. This ensures that at least when a bacterium encounters its preferred environment, it has a higher chance of remaining there.

E. coli expresses a number of adhesins, fimbriae as well as afimbriate adhesins, with a wide range of specificities (Antão et al., 2009). Fimbriae are long adhesive filaments extending from the outer membrane of the bacterium (Brinton, 1959). Different adhesins with different specificities recognize targets present in various environments (Duguid et al., 1979; Korhonen et al., 1982; Olsén et al., 1989; Parkkinen et al., 1983). What adhesin or adhesins a bacterium expresses at a given point in time helps determine whether or not it is capable of colonizing a given niche. A single *E. coli* can carry multiple fimbrial operons, possibly including more than one distinct copy of a given operon (Korea et al., 2010). By expressing different fimbriae in response to different stimuli, a single organism can colonize a variety of niches. For ExPEC, an ability to switch niches is crucial.

The study of microbiology is strongly biased towards what is relevant to humans, be it biotechnology or bacterial pneumonia. Organisms with pathogenic potential tend to be considered largely or even solely in light of their pathogenic niche, even though many virulence factors are functional within the commensal or environmental niche (Adlerberth et al., 1995; Alsam, 2006; Cookson et al., 2002; Lasaro et al., 2009; Rendón et al., 2007; Schierack et al., 2008; Wold et al., 1992). This focus on disease biology can lead us to overlook basic biology. While Shiga toxin is of great human importance, its importance to *E. coli* is uncertain. Inducing diarrhea leads to greater transmission between hosts, but the O157:H7 clonal group appears to have acquired Shiga toxin well before its emergence as a cause of human diarrheal illness (Leopold et al., 2009; Rangel et al., 2005; Zhang, 2006; Zhou et al., 2010). However, some studies suggest that it is involved in survival in the environment, helping the bacteria survive predation by protozoa (Meltz Steinberg and Levin, 2007). This hypothesis merits further study,

as it may mean that our current understanding of EHEC toxin production and persistence is incomplete.

The *E. coli* common pilus, Ecp, was initially described as Mat, or meningitis-associated and temperature-regulated fimbria. A newborn meningitis strain had all known fimbrial operons deleted, but it still expressed functional adhesive fimbriae. In the initial paper, the novel type of fimbriae was shown to be induced in response to higher temperatures and to be essential for biofilm formation at 37 C (Pouttu et al., 2001). However, further work showed that Mat/Ecp was not limited to pathogenic bacteria but was in fact widespread, being present in 91% of strains examined. Examination of Mat/Ecp fimbriation in other strains showed repression at high temperatures, not induction (Rendón et al., 2007). Aberrant expression of Ecp is associated with invasive infections. However, given that Ecp is frequently induced by low temperatures, Ecp likely also functions in the environmental niche (Lehti et al., 2010; Rendón et al., 2007).

Other fimbriae are also most often studied in the context of pathogenesis despite their important role in commensal or environmental niches. Curli, often studied in the context of DEC, is also involved in survival in the environment (Patel et al., 2011; Prigent-Combaret et al., 2000; Ryu et al., 2004). Type 1 fimbriae (T1F) are critical for UTI, but they are also important for commensal colonization of the intestine (Bloch et al., 1992). While these structures can play a role in pathogenesis, they are not strictly virulence factors. These dual roles inform us that it is not carriage of the operon that matters, it is the specific sequence of the operon that determines pathogenic potential.

2.2.2 The *fim* operon

Our lab studies the *E. coli fim* operon, which codes for T1F. T1F bind to terminal mannose residues on glycoproteins, which are found on a variety of mammalian cells (Klemm, 1984). On

the order of 90% of strains studied to date carry the *fim* operon (Duguid et al., 1979; Yamamoto et al., 1995). The prevalence of *fim* carriage is explained by the role of T1F in the spread of *E. coli* from one host to another. In an experimental model of normal transmission, rats were colonized with *E. coli* with or without the *fim* operon. Colonized animals were then housed with uninfected animals. Strains lacking T1F colonized new hosts at a much reduced rate compared to the strain carrying T1F (Bloch et al., 1992). Carriage of the *fim* operon does not mean that a strain is capable of producing T1F, however. O157:H7 strains carry alleles with an 16-bp deletion in the regulatory region; this deletion completely blocks expression of T1F under lab conditions (Enami et al., 1999; Iida et al., 2001; Li et al., 1997; Roe et al., 2001). It may be that in these strains T1F is inducible under *in vivo* conditions even if it is not inducible in the lab, or these strains may use another type of fimbriae, such as the type 4 pilus, to adhere to the intestine (Xicohtencatl-Cortes et al., 2009).

In the pathogenic context, the role of T1F is mixed. T1F are a critical virulence factor in cystitis (Bahrani-Mougeot et al., 2002; Connell et al., 1996; Snyder et al., 2006). However, there is strong selection against T1F expression in extraintestinal infections outside of the bladder, as well as in many diarrheal strains (Holden et al., 2006; Iida et al., 2001; Lim et al., 1998; Xia et al., 2000). Regulatory proteins carried in other pathogenicity-associated fimbrial operons bind to the regulatory region of the *fim* operon and suppress transcription (Holden et al., 2001; 2006; Xia et al., 2000). In these contexts, it appears that mannose binding is inappropriate or that the costs of immune recognition of T1F outweigh the benefits on mannose binding.

The *fim* operon consists of 9 genes, as shown in Figure 2.1. In strains that can express T1F, some fraction of the cells will be piliated while the remainder will be unpiliated. The proportion of bacteria expressing fimbriae depends on a number of genetic and environmental factors.

Whether or not a bacterium can produce T1F mRNA is controlled by the fimbrial switch region, *fimS*. The switch region, which is flanked by inverted repeats, contains a promoter that drives transcription of the fimAICDFGH mRNA when in the ON orientation (marked with asterisks) (Abraham et al., 1985).

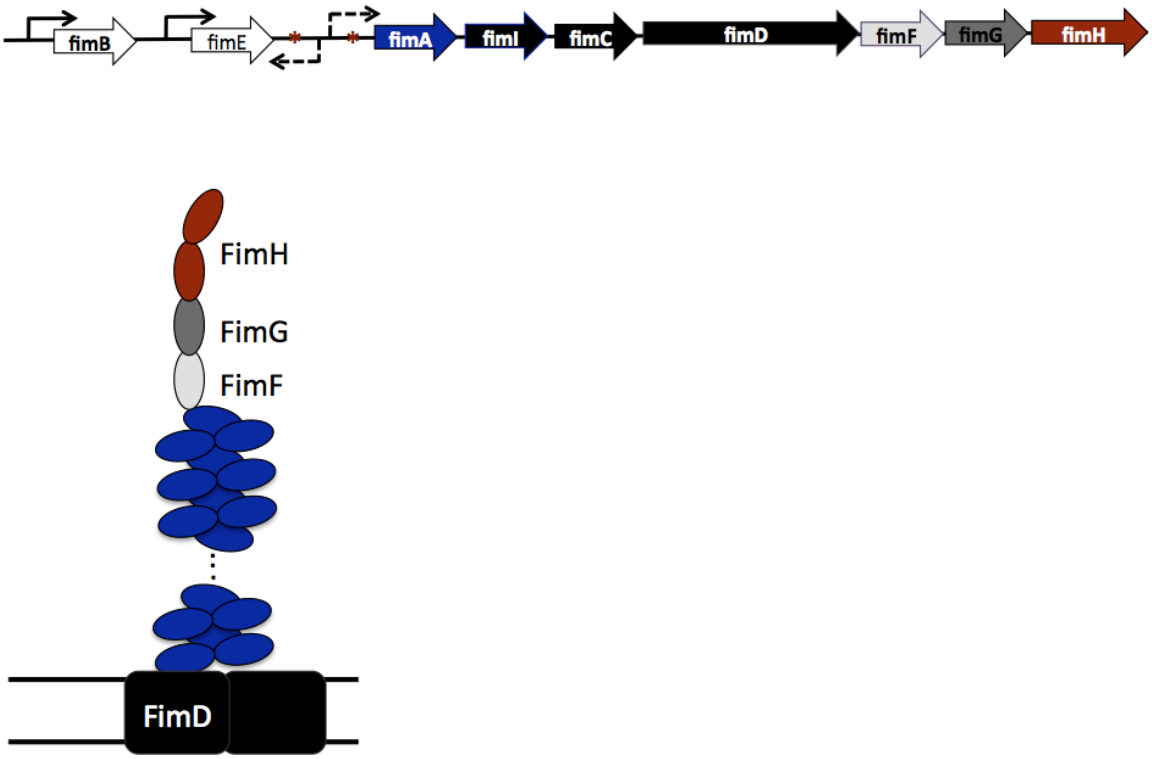

Figure 2.1: Operon and fimbria diagram.
(top) *fim* operon. Promoters indicated by corner arrows, with *fimS* promoter shown in both directions. Inverted repeats flanking promoter indicated by asterisks.
(left) Diagram of type 1 fimbrial structure

FimB and FimE are the primary regulatory proteins for *fim* (Klemm, 1986). Regulation of *fim* expression is by recombination. The FimB and FimE recombinases recognize the inverted repeats flanking *fimS* and flip it, switching the directionality of the promoter contained in the switch region. FimB is unbiased in its interactions with *fimS*, changing promoter direction randomly. FimE has a stronger affinity for *fimS*-ON than –OFF and so is biased towards turning off expression of T1F. In the absence of other factors, the ratio of these two proteins determines

the ratio of fimbriated to non-fimbriated cells (Gally et al., 1993; Holden et al., 2007; McClain et al., 1991; Schwan et al., 2007; Stentebjerg-Olesen et al., 2000). However, a number of genes outside of the operon also regulate type 1 fimbriation. Unlinked recombinases present in some strains also recognize and act on *fimS* (Bryan et al., 2006; Xie et al., 2006). Transcriptional regulators affect the expression of one or both recombinases, biasing the ratio of FimB to FimE (Blumer et al., 2005; Dorman and Higgins, 1987; Dove et al., 1997; Eisenstein et al., 1987; El-Labany et al., 2003; Olsen et al., 1998; Xia et al., 2000). A wide range of DNA-binding proteins, both universal and pathogenicity-associated, are able to bind to *fimS* and affect the rate and direction of inversion by altering the ability of the recombinases to interact with the inverted repeats (Gally et al., 1994; O'gara and Dorman, 2000; Roesch and Blomfield, 1998).

FimA is the major structural component of T1F and is present in hundreds of copies per fimbria (see Figure 2.1). FimF, G, and H are the minor subunits, each present in a single copy at the tip of the fimbria (Klemm and Christiansen, 1987; Krogfelt and Klemm, 1988; Krogfelt et al., 1990; Russell and Orndorff, 1992; Schembri et al., 2001). FimH is the adhesin, while FimF and FimG serve as adaptors. Structural data indicates that FimF and FimG serve as a flexible tether, increasing the range of contact angles over which FimH is able to bind to a surface (Aprikian et al., 2011; Gossert et al., 2008). FimI resembles a pilin subunit and has been shown to be required for synthesis of normal fimbria. FimI is not incorporated into the fimbria, however, and its function is unknown (Valenski et al., 2003).

Polymerization of the structural subunits is catalyzed by FimC and FimD, the chaperone and usher of the system, respectively (Jones et al., 1993; Klemm, 1992). Subunits are attached to each other through a process known as donor strand exchange. Each pilin domain resembles an immunoglobulin fold with a beta strand missing. Each pilin domain as well as the chaperone has

an N-terminal extension that can complete the fold (Choudhury, 1999). The energetics of protein folding drive fimbrial assembly. It is more energetically favorable for a pilin domain to fold with the chaperone than to fold alone. It is more energetically favorable for a pilin domain to interact with the appropriate pilin domain than to interact with the chaperone. The ΔG_0 of these interactions is such that fimbriae can be assembled without input of energy beyond synthesis and secretion of the subunits, and it results in extremely stable subunit-subunit interactions (Geibel et al., 2013; Gossert et al., 2008; Puorger et al., 2008; Remaut et al., 2006; Sauer et al., 2002; Zavialov et al., 2005). The fimbriae are assembled from the tip down. FimC-H heterodimers interact with FimD to initiate fimbrial synthesis, followed by addition of appropriate subunits (Phan et al., 2011).

2.2.3 The *fim* operon and pathogenicity

The *fim* operon is present in the vast majority of strains of all pathotypes, so presence or absence alone is not what makes some strains pathogenic (Duguid et al., 1979; Yamamoto et al., 1995). Rather, the sequence of the *fim* operon that a strain carries affects pathogenic potential. In particular, specific FimH polymorphisms have been associated with pathogenesis. Certain SNPs in FimH have been shown to greatly increase binding to mannose under static conditions, and these are strongly associated with UPEC strains (Sokurenko et al., 1994; 1997; 1995; 2004; Weissman et al., 2006). The wild-type FimH has catch-bond kinetics and only binds strongly to mannose under conditions of flow (Aprikian et al., 2007; Rodriguez et al., 2013; Tchesnokova et al., 2008; Thomas et al., 2004; 2002; Yakovenko et al., 2008). This is thought to be an adaptation to help bacteria bind to the gut mucosa during peristalsis without the binding pocket being blocked by the numerous soluble mannose moieties present. UTI-associated high-binding mutant

fimH alleles produce an adhesin that is locked into a high-binding conformation. Strains carrying these variants bind strongly to mannose under all conditions, which helps these strains colonize the bladder and persist (Sokurenko et al., 1997). While this is adaptive in the context of cystitis, high-binding FimH variants at a selective disadvantage in the gut, however. These adhesins bind to mannose even in the absence of flow and would therefore be blocked by soluble substrates (Weissman et al., 2007).

In order to examine the selective pressures acting on *fim*, I acquired *fim* sequences from a number of published genomes using BLAST and in some cases, manual reassembly of operons from BLAST results. The entire set of operons was analyzed for recombination as well as pairwise diversity (by gene and within a sliding window). As we were also interested in the most recent variation, which is most likely to occur during adaptation to a new niche, we identified groups of closely-related *fim* operons from closely-related strains and examined the variation occurring within these groups.

2.3 MATERIALS AND METHODS

2.3.1 Published genomes examined

A pilot study was performed using published genome sequences of 161 *Escherichia coli* strains as a starting point for our analysis, as these were the published genomes available at the time. Initial studies pointed to high variation upstream of *fimH* and suggested an excess of nonsynonymous variation amongst the most related *fim* operons. We subsequently performed further studies using a total of 1144 (983 additional) genome sequences. These sequences consisted of whole genome shotgun sequence data from GenBank, finished sequences from the PubMed database, and draft genomes from the Broad Institute, as well as genomes sequenced by

our lab. See Table 2.1 (genome strain data) for summary data on the strains used; a more detailed accounting of the strains and their various characteristics is available on request. These strains cover a wide variety of bacterial pathotypes, though the set is by no means a balanced sampling of *E. coli*. Strains selected for sequencing are typically of interest for humans, generally due to their pathogenic potential (isolates designated as "stool" are not necessarily from healthy samples or diarrheal samples). Our sample is therefore biased.

2.3.2 Retrieval of genes and operons

Nucleotide BLAST was used to retrieve sequences of interest from the genomic data, with thresholds of 90% identity to the query sequence and 95% of query sequence present (Altschul et al., 1990). For all genes, several alleles were used as query sequence, representing the most divergent alleles present in the PubMed Gene database for that gene. For multi-locus sequence typing (MLST), the query sequences consisted of the gene fragments used for typing (Wirth et al., 2006). Operon retrieval was sometimes complicated by the fact that draft and shotgun genome data consisted of many contigs instead of a fully assembled sequence, resulting in fragmented *fim* operons in many cases. When possible, the *fim* operon was extracted as a single sequence from 500 bp upstream of the *fimB* start through the stop codon of *fimH*. Otherwise, portion of contigs containing genes of the *fim* operon were identified and aligned to reference operons to reconstruct a full operon. Reconstruction of a complete operon was not always possible, and these reconstructions were discarded if more than 900 nt (10% of operon) was missing or if the operon was rearranged within in an assembly.

(a) ***fim* operons in small set by genome sequence source**

Source	n	Fim	%
Broad[1]	42	35	83
Published[2]	31	28	90
UW[3]	8	6	75
WGS[4]	78	46	59
total	**159**	**115**	**72**

(b) **Pathotypes and genome sequence sources of small set**

Isolation	n	from	n	Fim
DEC	41	PubMed	8	8
		WGS	33	16
Environment	1	PubMed	1	1
ExPEC	15	PubMed	8	8
		UW	6	6
		WGS	1	0
Stool	77	Broad	42	35
		PubMed	4	3
		UW	2	1
		WGS	29	18
Lab	10	PubMed	6	5
		WGS	4	3
Unspecified	17	PubMed	4	3
		UW	2	1
		WGS	11	9

(c) **Pathotypes and genome sequence sources of large set**

Isolation	n	from	n	Fim
DEC	281	PubMed	25	18
		WGS	256	228
Environment	5	PubMed	1	1
		WGS	4	3
ExPEC	38	PubMed	11	11
		UW	4	4
		WGS	23	20
Stool	345	Broad	87	82
		PubMed	3	2
		UW	2	1
		WGS	253	186
Lab	15	PubMed	8	5
		UW	1	1
		WGS	6	5
Unspecified	436	Broad	233	204
		PubMed	2	1
		UW	2	2
		WGS	199	155

[1]Broad: whole genome shotgun sequences from the Broad Institute Antibiotic Resistance Database.
[2]Published: fully assembled genome sequences deposited in the NCBI Genome Database.
[3]UW: whole genome shotgun sequences of 8 ST 95 strains completed in-house.
[4]WGS: whole genome shotgun sequences from the NCBI Genome Database.

Table 2.1: Genome strain data.
More detailed strain data is available as supplementary data file.

2.3.3 Programs used for alignment, visualization, and analysis

Alignments were created and manipulated using ClustalX2 and MEGA5.2 (Larkin et al., 2007; Tamura et al., 2011). Alignment figures were created using BioEdit 7.0.9.0 (Hall, 1999). Trees were constructed using the maximum likelihood method with DNAml from the PHYLIP package (Felsenstein, 2005). Trees were displayed and figures were created using TreeView 1.6.6(Page, 1996). Calculations of π and dN/dS were made with DnaSP 5 (Rozas et al., 2003). Calculations of nucleotide variance and polymorphism made using DnaSP 5 and ProSeq2.91 (Filatov, 2009; Rozas et al., 2003). PhiPack was used to detect recombination within a set of sequences using the Phi test as well as Max χ^2 and NSS (Bruen, 2006). RDP 3.44beta was used to identify likely breakpoints using X-over analysis (Martin et al., 2010).

2.4 RESULTS

2.4.1 Presence and characteristics of *fim* operon in sequenced strains

From the 161 genomes examined, I retrieved 108 complete and 9 partial *fim* operon sequences, of which 77 were unique. With 72% of the genome sequences carrying a *fim* operon, this rate of carriage is far below previously published data on *fim* prevalence (Duguid et al., 1979; Yamamoto et al., 1995). The shortfall is most likely due to the incomplete assembly of genome sequences. Operons were retrieved from 90% of published fully assembled genomes but were only retrieved from 68% of genomes assembled to the contig/supercontig level. The BLAST search included a requirement that 95% of the query sequence be present for a target sequence to be returned. Given that draft sequences consisted of anywhere from 2 to 2939 contigs, some operons were almost certainly missed due to gene fragmentation across contigs. Similar results were seen with the larger genome set.

Two highly divergent *fim* operons were observed in the initial sample set, in E1118 and H605. Though these strains had been identified as *E. coli* within the genome database, examination of housekeeping genes showed them clustering with other *Escherichia* species. Seven additional divergent *fim* operons were identified in the larger genome set. In all cases, the strains carrying the divergent operons proved to be non-*coli Escherichia* species, with the topology of the *fim* tree resembling the concatenated MLST tree. Within the set, there does not appear to be acquisition of *fim* from other species by *E. coli*. This is most likely due to the relative rarity of cross-species horizontal gene transfer as compared to within-species horizontal gene transfer. If an apparent cross-species transfer of *fim* is observed as more genomes are published, it will be interesting to look into the traits of the transferred operon and more specific traits of the strain with the divergent operon. If no apparent cross-species *fim* transfer is observed even after the publication of many more genomes, it may be fruitful to look into whether this is unusual compared to other similarly mobile genetic loci and whether there may be other factors limiting the range of *fim*.

2.4.2 Global variation within the *fim* operon

With the exception of *fimA*, which is known to be recombinant, genes within the operon showed similar amounts of pairwise variation or π (see Figure 2.2). When examined on a finer scale using a 100-bp sliding window, π was more variable but there were no significant minima or maxima of variation aside from the already observed increased variation in *fimA* (not shown). Similarly, outside of *fimA*, dN/dS of *fim* was significantly below 1, indicating purifying selection. On a species level, *fimA* is under positive selection for variation while the remainder of the operon is under strong negative selection. The large amount of recombination seen within

fimA contributes to both dS as well as dN, resulting in relatively high local variation without a dN/dS significantly above 1.

2.4.3 Recombination within the *fim* operon

Operons were analyzed using RDP3.44beta to identify the most likely breakpoints within operons flagged as being recombinant. The analysis indicated clusters of breakpoints upstream of *fimB* and downstream of *fimH*. This agrees with previous observations that some *fim* operons are promiscuous; identical or nearly identical *fim* operons are present across various groups of closely-related *E. coli*, indicating horizontal gene transfer of entire operons. Within the operon, breakpoints were concentrated in and around *fimA*, which is again congruent with previous research showing that *fimA* is highly recombinant. FimA is the major pilin of T1F. Since it is both surface exposed and present in such high copy numbers relative to the other pilin subunits, it is the dominant target of the immune system and as such is subject to strong positive selection for variability.

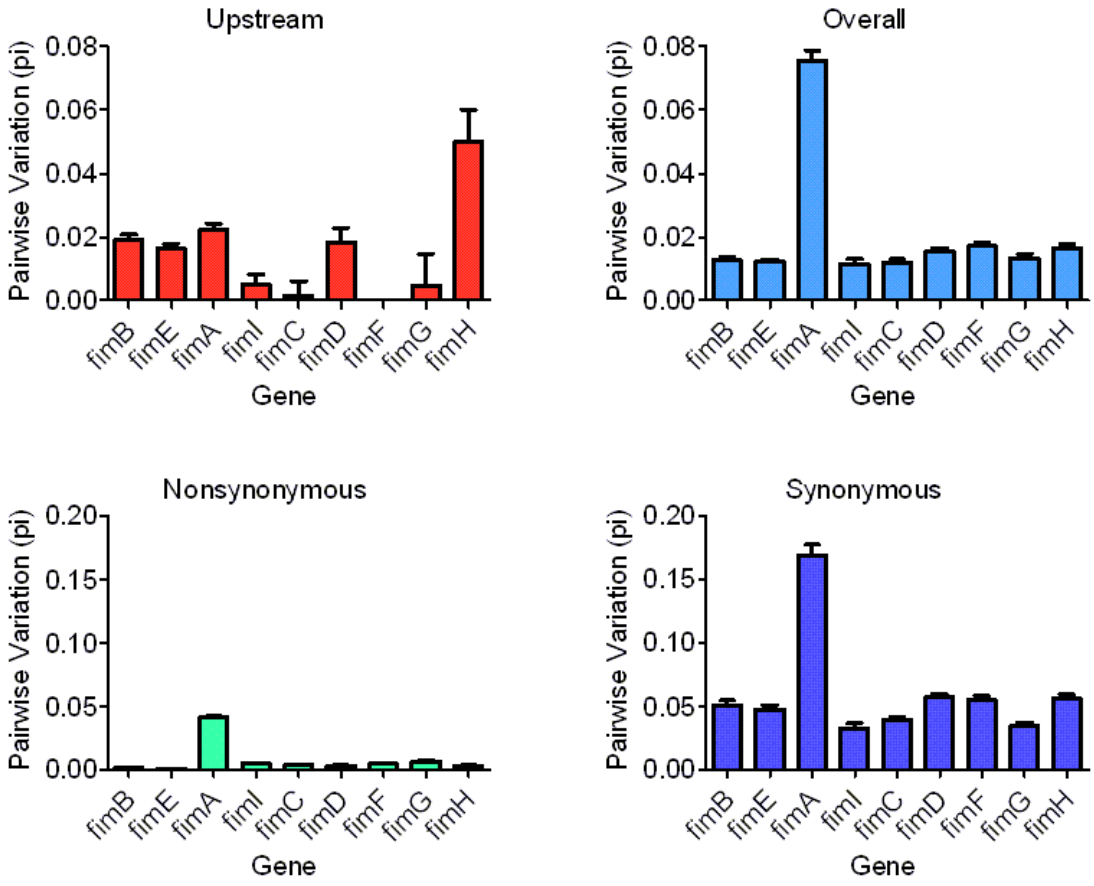

Figure 2.2: Pairwise diversity of *fim* genes.
Pairwise diversity among analyzed fimbrial operons by region, graphed by type (overall coding, non-synonymous, and synonymous) with standard error. y-axis is the same for all graphs.

Figure 2.3: recombination within the operon.
Schematic of *fim* operon with predicted breakpoints marked with vertical lines. For clarity, predicted breakpoints within a single 10-nt stretch are denoted with a single line. Number of hits at a specific breakpoint is not denoted.

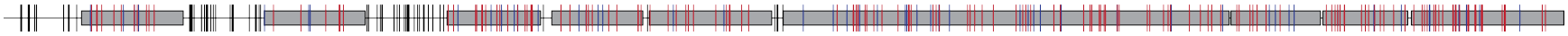

Figure 2.5: polymorphisms within closely-related *fim* operons.
Schematic of *fim* operon with polymorphisms observed between closely-related *fim* operons among strains from the same ST denoted with vertical line. Black lines indicate non-coding positions, blue lines indicate synonymous positions, and red lines indicate non-synonymous positions.

2.4.4 Bias in most recent mutations

We are particularly interested in recent mutations, i.e. among closely related strains. One of the most common methods for assessing strain relatedness is multi-locus sequence typing, or MLST. Strains are typed by sequencing regions of approximately 500 nt from seven housekeeping genes spaced along the *E. coli* chromosome. Strains with identical sequences at all seven positions have the same sequence type, or ST, and are assumed to be highly related. Mutations between strains from the same ST should be recent. I examined polymorphisms between highly similar *fim* operons within the same ST (sub-ST clades) to look for patterns in the most recent mutations. Clustering of mutations within a certain region could indicate either positive selection or drift in that region. An overabundance of synonymous, non-synonymous, or non-coding mutations could indicate the dominant type of selection operating on the operon in recent evolutionary history.

In the pilot study, six STs had representatives with two or more distinct but closely related *fim* operons. ST11 was the only cluster with more than three unique *fim* sequences. ST11 contains O157:H7 strains and is therefore well represented in the available genome sequences. Of the 7 within-ST mutations observed in ST11, 5 were non-synonymous and 2 were non-coding. The non-coding mutations were both within *fimS*. The number of mutations was too small for this bias to be statistically significant (with 77% of coding sites in *fim* being non-synonymous, $p = 0.26$). Nonetheless, the pattern was striking. Unexpectedly, one of the non-synonymous polymorphisms was an inactivating mutation in the binding pocket of FimH. O157:H7 strains carry a *fim* operon with an 16-bp deletion in *fimS* that blocks expression of T1F, at least under lab conditions (Shaikh et al., 2007). An inactivating mutation in FimH subsequent

to the *fimS* deletion implies that the block to *fim* expression is not absolute and that there is very strong selective pressure against T1F expression in O157:H7 *E. coli* (Shaikh et al., 2007).

When the larger genome set was considered, 540 strains with *fim* operons could be grouped into ST clusters, representing 364 unique operons. Not all operons from strains with the same ST were closely related, however, due to recombination. However, within a single ST, there were typically distinct clusters of operons (see Figure 2.4), allowing for straightforward division into clusters of same-ST strains with closely related *fim* operons. These clusters consisted of 464 strains with 302 unique operons, and are henceforth referred to as sub-ST clades. In all but one case, each sub-ST clade was separated by many changes (10 or more) from the next nearest *fim* operon found within that ST.

There was a strong bias towards nonsynonymous mutations in sub-ST clades. In the case of the less clearly demarcated cluster, there were 8 non-synonymous and 4 synonymous polymorphisms in the larger set. Eliminating the less closely related operons from the cluster resulted in 6 non-synonymous and 1 synonymous polymorphisms. This pattern held true at the larger scale, as well. Considering all *fim* changes within an ST, regardless of relatedness of *fim* operons, 612 non-synonymous and 1400 synonymous polymorphisms were observed, compared to 167 non-synonymous and 64 synonymous changes within sub-ST clades, showing a highly significant skew towards non-synonymous changes in the most recent mutations ($p < 0.001$ by χ^2).

	Within-ST*	Sub-ST clade	p-value (χ^2)
Non-synonymous	612 / 30.4%	167 / 72.3%	< 0.001
Synonymous	1400 / 69.6%	64 / 27.7%	

* Numbers shown are total polymorphisms observed / non-synonymous or synonymous as percent of total polymorphisms within that data set

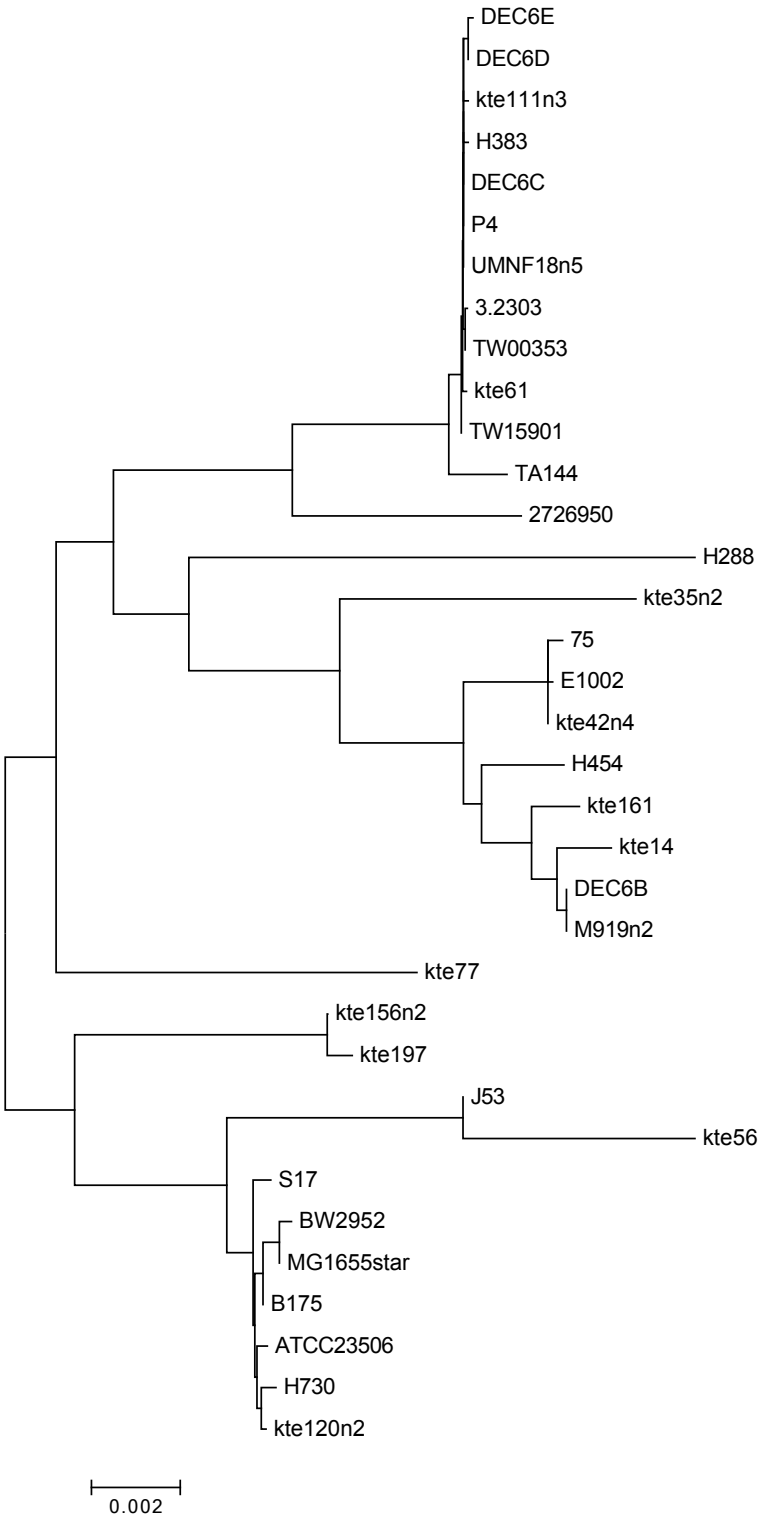

Figure 2.4: Sub-ST clades of *fim* operons within a single sequence type.

Within the sub-ST clades, there was also a smaller but nonetheless significant increase in the proportion of non-coding polymorphisms relative to coding. The within-ST set had 349 non-coding polymorphisms (compared to 2002 coding); the sub-ST clades had 82 non-coding polymorphisms (compared to 231 coding). This skew towards non-coding mutations is significant compared with both the observed proportion of non-coding mutations in the larger set and the proportion of the *fim* operon which is non-coding, i.e. what would be expected from random chance ($p = 0.021$ versus the larger set and $p < 0.001$ versus random chance by χ^2).

	Within-ST*	Sub-ST clade	Actual sites
Coding	2002 / 85.2%	231 / 73.8%	82.5%
Non-coding	349 / 14.8%	82 / 26.2%	17.5%
p-value†	close v. within: 0.021	close v. random: <0.001	

* Numbers shown are total polymorphisms observed / coding or non-coding polymorphisms as percent of total within that data set

† χ^2 of percent coding and non-coding in the closely-related set versus percent coding in either the within-ST set or the percent expected by random chance

2.5 CONCLUSIONS

2.5.1 Recent variation shows a skew towards non-synonymous and non-coding mutations

The marked trend towards a greater proportion of non-synonymous mutations in the most recent variation strongly suggests that positive selection dominates in the short term. Increased non-coding variation suggests either selection for non-coding variation (i.e. positive selection for altered transcription/translation) or relaxed selection operating at non-coding positions. However, the increased non-coding variation among *fim* operons from closely related strains and the trend towards a larger proportion of non-coding mutations as operons become more closely related would seem to support the positive selection hypothesis. Selection for non-coding variation will be discussed further in chapter 3.

2.5.2 Drawbacks of using genome sequences to study variation

While mining published genome sequences for variation in genes of interest is a promising approach, there are also drawbacks. Many of the genomes used were still in draft form, existing as multiple contigs, and most have remained in this format for years after their initial release. Full assemblies are unlikely to be constructed until one of the long read-length next-next-gen sequencing techniques is commercialized at a low price. Current methods of completing assemblies are now more time-intensive and expensive than the shotgun sequencing. The numerous breaks in the shotgun genome sequences compromise the ability of BLAST to identify genes of interest. This results in the relatively low hit rate of *fim* operons retrieved by this analysis compared to the rate expected from previous studies. While the low hit rate was disappointing for *fim*, the frequency of *fim* carriage meant that it was still possible to do a comparative analysis. Unfortunately, the reduction in operon retrieval severely restricted analysis of less-common operons such as Pap and Sfa, as there were too few haplotypes for informative analyses.

Additionally, the vast majority of the *E. coli* genomes in the NCBI database had little or no metadata associated with individual strains. Several large studies included the general makeup of their strain collection, e.g. *E. coli* isolates for comparison of fecal isolates from individuals with varying degrees of diarrheal illness. Individual strains within studies were infrequently or never marked with isolation source or pathogenicity, however, so we cannot make any associations of genetic variation with pathogenicity or pathotype. In this case, when the researchers who have sequenced these strains publish their findings, it is likely that the metadata will be released or available through their publications.

2.5.3 Further work on published genomes

Despite these drawbacks, published genome sequences represent an impressive data collection, and one that is well worth mining. Upcoming sequencing technologies show gains in read length, accuracy, and cost-effectiveness, suggesting that future published genomes will be of higher quality and greater number. A drop in cost will also contribute to more sequencing of a wider variety of strains, not just those from the scariest or most-studied isolates. However, in order to take advantage of these data, more of the analysis pipeline would need to be automated. The time needed to manually curate the *fim* BLAST results from a 100-genome set was not particularly great, but curating the results from the 1000-genome set became a substantial task; doing the same for 10 000-genome set would be unreasonable.

Resources such as the Biopython library render it fairly straightforward, if not exactly trivial, to automate tasks such as ST assignment, operon BLASTing, and simple operon reconstruction. Integrating existing programs such as those in the PHYLIP suite may make it possible to construct phylogenies, identify sub-ST clades, and report on variation found in the increased set. Additionally, as the number of published genomes increases, it will become possible to do things such as look at co-variation among the various pathogenicity-associated operons and identify potential nucleotides of interest for cross-talk between operons. With a solid framework for analysis and some programing know-how, the possibilities are amazing.

CHAPTER 3:

INCREASED NON-CODING VARIATION UPSTREAM OF THE *fimH* ADHESIN

3.1 ABSTRACT

The enrichment in non-coding variation in the *E. coli fim* operon over the short term discussed in Chapter 2 pointed to a role for non-coding variation in phenotypic variation of type 1 fimbriae. Previous work in the lab demonstrated that variation in FimH concentration could lead to pathoadaptive variation in fimbrial structure. I analyzed the non-coding regions of *fim* operons from sequenced genomes to look for evidence of positive selection for non-coding variation. A repeated change upstream of the *fimH* start site was found which had hallmarks of positive selection and was associated with extraintestinally pathogenic strains. Modeling suggests that this repeated change may alter mRNA stability and therefore translational efficiency of FimH.

3.2 BACKGROUND

3.2.1 Untangling positive and relaxed selection in the *fim* operon

Knowing what kind of selection is operating on a gene can give us important clues as to its function and importance, whether it's essential and tightly conserved or drifting as a pseudogene. The standard mechanism for detecting positive selection is dN/dS, the ratio of the number of amino acid changes out of all possible over the number of silent changes out of all possible, corrected for multiple substitutions at single sites. While a high dN/dS in the absence of recombination is a strong sign of diversifying selection, it is a very blunt instrument, operating as it does at the level of entire genes. In the case of FimH, there is strong selection for altered binding kinetics under certain circumstances, but there are still tight functional constraints: it

must still be able to interact appropriately with other components of the fimbria, and it must still bind mannose. As only certain portions of the *fimH* sequence are free to vary under these conditions and variation is potentially beneficial in only a few regions, the local positive selection is swamped by the purifying selection acting at the vast majority of positions. Examining selective pressures in a more fine-grained way can provide clues to protein structure and function. For example, *fimA*, which is under very strong positive selection and experiences a high level of recombination, still contains stretches of amino acids which vary little; these correspond to portions of the protein under structural constraints, such as interactions with other fimbrial proteins (Boyd and Hartl, 1998; Klemm, 1984; Peek et al., 2001).

Our lab developed hotspot analysis as a method of detecting positive selection at the amino acid scale, particularly against a background of negative or neutral selection. A hotspot is a position that has mutated more than once, independently, over the evolutionary history of a sequence, indicating that at that position there is selection either for variation in general or towards a particular variant. I am using the same approach to examine changes at the nucleotide level, examining supposedly silent variation for evidence of positive selection.

3.2.2 Pathogenicity-associated changes in available concentration of adhesin subunit

Our lab had previously observed a *fimH* mutation in the signal peptide associated with uropathogenesis. While mature protein was unchanged, the fimbrial structure was altered. This nonsynonymous SNP in the signal peptide greatly reduced the efficiency of FimH secretion into the periplasm, which in turn reduced the number of fimbriae that could be initiated. As FimA translation and secretion remained unchanged, the fimbriae that were initiated tended to be longer than normal due to the change in stoichiometry between subunits. In assays of binding

kinetics under flow, strains with the *fimH* mutation had a greatly reduced off-rate after switching from high flow to static conditions compared with wild-type. This is hypothesized to be due to bacteria becoming cross-braced by their T1F; fimbriae on each side extend in opposite directions and provide tension to each other, maintaining the catch bond of FimH in its high-binding state (Ronald et al., 2008). As with previously-observed the high-binding FimH variants, there appears to be strong selection for this in the context of cystitis.

3.2.3 Hypothesis

If an effect on secretion could have such a dramatic effect on phenotype, it suggests that noncoding or synonymous mutations that change protein levels could also affect the fimbrial phenotype and pathogenic potential. This observation led us to examine the *fim* operon in more detail, focusing particularly on variation in intergenic regions. Changes in protein levels can have a significant phenotype, just as changes in protein sequence can, and I am interested in positively selected noncoding and synonymous variation near the translation start sites of fimbrial genes, as this may represent an understudied mechanism of adaptation.

3.3 MATERIALS AND METHODS

3.3.1 Hotspot nucleotide analysis

Hotspots refer to positions that have independently undergone mutation at least twice during the evolutionary history of a sequence. The position may have mutated to different residues in different branches of the phylogeny, but independent mutation to the same residue is frequently seen and is a strong indication of positive selection for that mutation. Previous work in this lab has focused on hotspot amino acids, or nonsynonymous hotspots as markers of positive

selection. For this study, I used the same technique to look for hotspot nucleotides that could affect the expression of a gene or genes of interest. Briefly, sequences are analyzed for hallmarks of recombination by Phi (Bruen, 2005) and by RDP v. 3.44 using RDP, GENECONV, MaxChi, BootScan, and SiScan (Martin et al., 2010; 2005). Phi signals the likelihood of recombination being present in a given set of sequences, and RDP indicates the location of likely breakpoints in these sequences. As hotspots are by definition homoplasies, it is necessary to rule out recombination as much as possible in order to minimize false positives. After eliminating likely recombinant sequences and duplicate sequences, the remainder are aligned using ClustalW (Larkin et al., 2007). PAUP* is used to construct a maximum likelihood tree of the sequences (Swofford, 1991), and ZPS (Zonal Phylogeny Software) places mutations on the tree and identifies hotspots (Chattopadhyay et al., 2007). This analysis was done for genes, intergenic regions, and "start-proximal" regions, which were defined as -300 to +300 relative to the start site of the gene indicated.

3.3.2 Strains sequenced for fimG-fimH intergenic region.

See Table 3.1 for summary data of additional strains sequenced.

Isolation	n
Asymptomatic bacteriuria	1
Fecal	3
ExPEC	9
Lab	1
Vaginal	23
Unspecified	14

Table 3.1 Isolation source of strains Sanger sequenced for *fimH* upstream non-coding region and adjacent.
See supplemental data for full list of strain names and rates of *fim* carriage by source.

3.3.3 Sequencing nucleotides proximal to fimH start site

fimH and upstream were amplified using primers CL17 (CGGCCTGGCATGATGTTG) and CL18B (TAATTTGCCGTTAATCCCAGAC) specific for 330 nucleotides upstream of and the

stop site of *fimH*. Samples were cleaned up with ExoSAP (Affymetrix) according to manufacturer instructions before being sent out for sequencing. Sanger sequencing was performed by Genewiz using the CL17 primer.

3.4 RESULTS

3.4.1 Increased variation seen immediately upstream of the fimH start site

Each gene and noncoding region was analyzed for various markers of variation and selection. With the exception of *fimA*, dN/dS was consistent and significantly below 1, indicating purifying selection. Excluding *fimA*, both synonymous and nonsynonymous variation was consistent across the operon, as measured by π (Figure 2.2). The level of intergenic variation was similar to the overall variation in coding regions. Intergenic variation was also consistent across the operon, with the exception of the intergenic region between *fimG* and *fimH* (Figure 3.1). The pairwise variation of this short stretch was significantly higher than the overall non-coding variation ($p < 0.01$). Of the five observed polymorphisms in this 19-nt stretch, none disrupted the predicted ribosome-binding site of *fimH*, but this does not necessarily mean that the changes are functionally silent. One polymorphic position was immediately upstream of the predicted ribosome binding site while another was at -1 relative to the translation start site; though these do not directly disrupt translation, they may perturb it in some fashion.

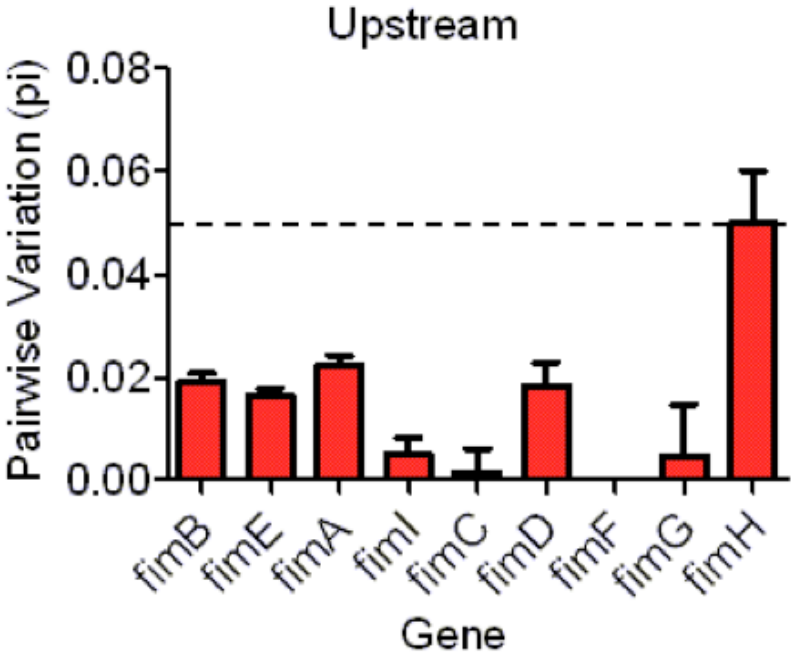

Figure 3.1 Non-coding variation in the *fim* operon
Bars represent noncoding π for intergenic region upstream of each gene with standard error. Dotted line indicates average synonymous π for fimbrial genes excluding *fimA*.

3.4.2 Hotspot nucleotides upstream of *fimH* start site

Of the five polymorphisms observed in the *fimG-fimH* intergenic regions, none directly interfered with the ribosome-binding site, so there is no obvious potential effect. In order to ascertain whether the increased variation was likely to be the result of relaxed selective pressure or positive selection, I applied hotspot analysis to a region spanning -300 to +300 of the *fimH* start site, excluding those sequences with recombination predicted within that region (Figure 2.3). ZPS analysis of this region identified two of the five polymorphic positions in the intergenic region as hotspot nucleotides. Intriguingly, one of these hotspot polymorphisms (-15a) was significantly associated with extraintestinally pathogenic *E. coli* (χ^2, $p < 0.001$).

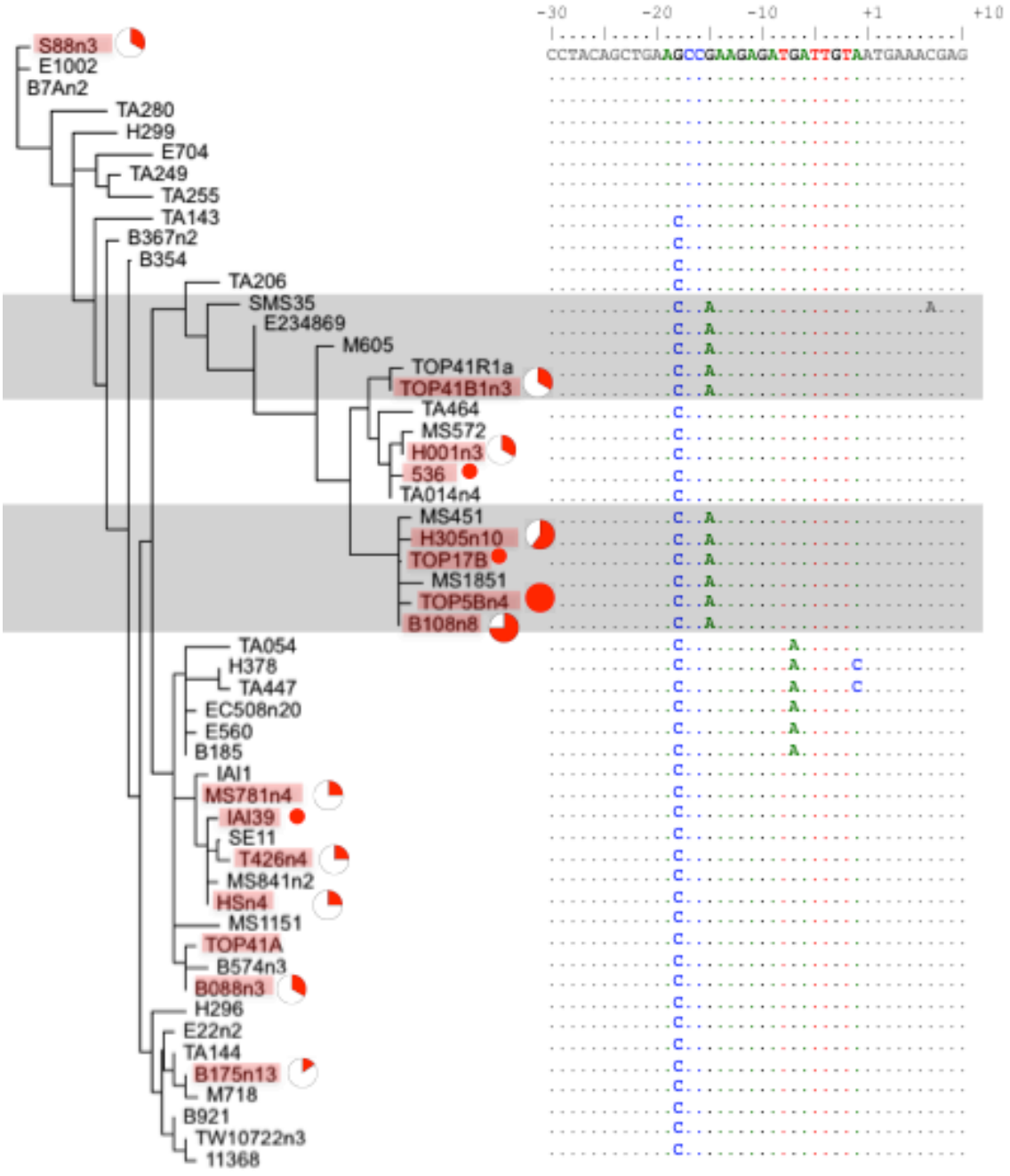

Figure 3.2: Association of -15a *fimH* polymorphism with ExPEC strains.
Maximum-likelihood tree of -192 to +312 relative to *fimH* start with sequences containing -15A shaded. Where more than one strain had identical sequences, these were collapsed to a single tip. Number of isolates with that sequence signified by “n” followed by the number of sequences. Any tips which include ExPEC strains are indicated by red shading. For tips with at least one ExPEC strain, proportion of strains that are ExPEC is indicated with red section of circles.

The fact that hotspot analysis is based on detecting homoplasies in nucleotide sequence-based trees means that undetected recombination is always a potential issue. It is possible that in very closely related *fim* operons, recombination events would not be detectable with the techniques used, resulting in a genuinely silent noncoding change being inappropriately flagged as a hotspot. In order to determine the likelihood of this, I examined the distribution of the -15 *fimH* alleles on maximum likelihood phylogenies of the entire *fim* operon, as well as of *fimG* and *fimH*. If the polymorphism were found only within a single subtree of the *fim* operon phylogeny, this would indicate a common background for the polymorphism. While this would not rule it out as a genuine hotspot, it would indicate the need for caution as well as additional tests. If the sequences carrying the -15a polymorphism were monophyletic in either the *fimG* or *fimH* tree, this would draw attention to the possibility that this polymorphism may not be selected for per se but instead carried along by recombination selecting for a polymorphism in the gene in question. When this analysis was performed, the *fim* operons that carried the -15a polymorphism were not monophyletic and in two cases were substantially divergent from the closest neighbor carrying the -15a polymorphism, indicating that there is not a common background for this polymorphism. When placed on the *fimG* and *fimH* trees, the -15a polymorphism was again not monophyletic, strengthening the argument that this is a noncoding hotspot selected for on its own merits.

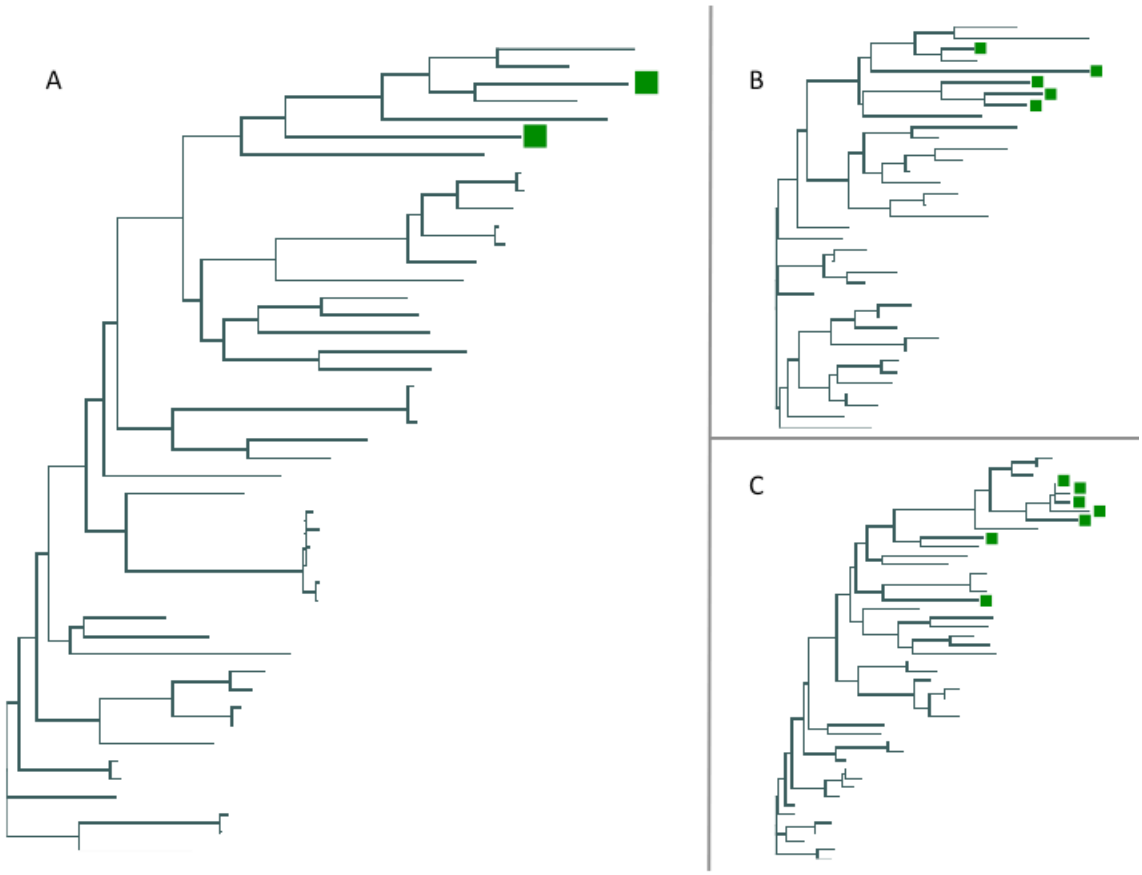

Figure 3.3: Non-monophyly of -15a polymorphism.
Background of the -15A polymorphism at other sequence levels; maximum-likelihood trees of the *fim* operon (A), *fimG* (B), and *fimH* (C), with boxes indicating presence of the -15A polymorphism. Strain name labels have been omitted for clarity.

3.4.3 Examining polymorphism proximal to *fimH* start in further strains

I was interested in the distribution of this polymorphism in a larger sample of strains, as well as whether expanding the sample set for the analysis would yield other hotspot nucleotides of interest. To that end, I sequenced this region in a further 110 strains from our collection (see Table 3.1). The sequence from -279 to +389 relative to *fimH* start was consistently of high quality, and this region was used for analysis. No further hotspots were identified, but position -15 continued to be flagged as a hotspot nucleotide. -15a was also significantly associated with ExPEC in those strains where a source of isolation was specified.

3.4.4 Possible effect of -15a *fimH* polymorphism on translation

While synonymous variation is often referred to as "silent," research by Kudla and colleagues demonstrated that this is not necessarily the case (Kudla et al., 2009). A library of 154 synonymous variants of GFP was created; when expression vectors carrying these sequences were transformed into *E. coli*, fluorescence levels varied widely (up to 250-fold). Expression level was not correlated with the codon adaptation index but instead correlated with the predicted folding energy of the mRNA; the more negative the free energy of the folded mRNA immediately around the translational start site, the less fluorescence was seen in cells expressing that construct. This raises the possibility that something similar could be happening in the case of the -15a *fimH* polymorphism. Analysis of the *fimA-fimH* mRNA by mFold (Zuker, 2003) indicated a likely hairpin structure including the *fimH* start site. When free energy was calculated for the nucleotides immediately around the start site, presence of the -15a polymorphism was associated with lower values, suggesting a possible reduction in translation. The calculated ΔG_0 of the RNA hairpin structures predicts and 2-fold decrease in translation with the UTI89 sequence compared to MG1655 (de Smit and van Duin, 1990; Seo et al., 2009). This is less than is seen with the secretion signal mutant, but it is still potentially able to affect the structure of type 1 fimbriae.

3.5 CONCLUSIONS

3.5.1 Non-coding variation upstream of *fimH*

The -15a *fimH* polymorphism appears as a hotspot in all the data sets considered (small set of genomes, small set plus strains sequenced for region surrounding *fimH* noncoding region, and large set of genomes plus sequenced strains). This polymorphism is also associated with an ExPEC strain origin, implying that selection at this site may be related to an extraintestinal lifestyle and/or the alternating selection pressures experienced by ExPEC strains.

3.5.2 Further examination of the effects of -15a *fimH* hotspot nucleotide

The non-coding position at -15 relative to the *fimH* start site appears to be subject to repeated variation associated with strain isolation from extraintestinal infections. Models of mRNA folding indicate that this polymorphism may be associated with reduced translation, which could have effects similar to the FimH secretion signal mutations observed previously. The secretion signal mutant strains displayed longer fimbriae than strains with a normal signal sequence, and the data suggest that individual bacteria could be cross braced by their fimbriae, inducing a conformation shift in FimH as is seen with shear force in the absence of any flow.

The -15a polymorphism could also affect *fimH* transcription or translation by a variety of other means. The -15 nucleotide is in or immediately adjacent to the *fimH* ribosome binding site and so could directly affect translational efficiency. Alternatively, it could potentially affect mRNA stability, endonucleolytic cleavage involved in translational regulation, or binding by regulatory factors. The first step in characterizing the -15a polymorphism would be to construct variant *fim* operons carrying various combinations of the observed noncoding polymorphisms upstream of *fimH* and determine their effect on the number and length of pili produced by

bacteria carrying the operon. If an effect were observed, the challenge would then be to determine the mechanism by which the polymorphism affects the change in FimH expression.

3.5.3 Utility of work for future studies

I am intrigued by the possibility that non-coding changes like the one discussed here could be part of the suite of pathoadaptive changes that alter the pathogenic potential of ExPEC. Certain STs are strongly associated with extraintestinal pathogenicity, implying that these strains are pre-adapted to an ExPEC niche, needing only an opportunity and one or a few genetic changes in order to thrive in the bladder, blood, or brain. By biasing the operon in the direction of fewer and longer fimbriae, would this provide a better background for future changes? Could smaller changes like this provide an advantage in the urinary tract without also causing a fitness defect in the intestines? We study the pathoadaptive mutations that cause significant phenotype effects because that is what is visible to us.

However, some mutations causing small phenotypic variations would be expected to be beneficial, just as with mutations causing large variations. Indeed, given the necessity for ExPEC to be able to switch between niches, small changes that are beneficial within the extraintestinal niche, even if only to a small degree, could be quite beneficial in the long run if they do not interfere with fitness in the intestinal niche. As we gain more sequence data, I believe that we will become able to detect more and more of these subtle variations, which will in turn illuminate more about the pathways involved in pathogenesis. While these data are pointing back to a factor that we already knew was important, it supports our hypothesis that FimH expression levels as well as amino acid sequence can impact on the pathogenic potential of a given strain. I believe that future studies can and will illuminate novel participants in pathogenesis.

CHAPTER 4:

E. COLI AS A PERSISTENT RESIDENT OF THE CF LUNG AND WITHIN-PATIENT VARIATION OF LONG-TERM INFECTING *E. COLI*

4.1 ABSTRACT

Individuals with cystic fibrosis (CF) frequently suffer from lung infections, causing substantial morbidity and mortality, and by adulthood they have often developed chronic polymicrobial lung infections. Recent work by our collaborators has demonstrated that *E. coli* is capable of causing long-term infections of the CF lung and undergoing phenotypic changes characteristic of long-term CF lung colonization. We sequenced the genomes of two clonal strains from a single patient with differing phenotypes (mucoidy and ampicillin-resistance) to track down the genetic changes associated with these phenotypic changes. Genome sequencing strongly supported the hypothesis that these strains were clonal and had arisen from a single strain over the course of the patient's lung infection.

4.2 BACKGROUND

4.2.1 Lung infections in cystic fibrosis

In the USA, cystic fibrosis (CF) is the most common life-shortening genetic disease, after sickle cell anemia, occurring in 1 in 3500 live births in the US (Cystic Fibrosis Foundation, 2012; O'Sullivan and Freedman, 2009). Patients with CF have a defect in chloride transport, and the primary effect of this defect is altered mucus composition. Decreased net ion export leads to decreased net water movement, but normal amounts of other components are produced, resulting in mucus that is thick, dry, and sticky (Boucher, 2007; Matsui et al., 1998). This mucus can block the pancreatic duct, the small intestine, and small airways of the lung (Cohen and Prince,

2012). It is this last that causes the greatest number of problems for patients (Cystic Fibrosis Foundation, 2012).

In the lungs of patients with CF, mucus cannot be cleared properly. Instead of trapping microbes and particulates and carrying them away, the mucus instead retains inhaled particles in the lungs. The mucus itself can also serve as a food source for a number of bacteria. Many organisms can then proliferate and cause lung infections, even species that would never be pathogenic to healthy individuals. Initially, patients suffer acute infections caused by a single species. However, CF patients have a reduced ability to clear lung infections, due both to depletion of airway surface liquid and immunological abnormalities apparently resulting from the defect in the cystic fibrosis transmembrane conductance regulator (CFTR) (Cohen and Prince, 2012; Matsui et al., 1998).

The most common and most well studied causes of infection in the CF lung are *Pseudomonas aeruginosa* and *Staphylococcus aureus*. Nearly 75% of CF patients have a *P. aeruginosa* infection by the time they reach adulthood (Cystic Fibrosis Foundation, 2012). Chronic *Pseudomonas* infection is generally considered to be a condition to be managed rather than a curable condition. Patients with chronic infections appear to be more susceptible to infection by further species. The polymicrobial infections that develop are complex, with the microbial community varying over time as well as from region to region within the lungs (Fodor et al., 2012; Sibley and Surette, 2011). The bacteria undergo a number of adaptations to the CF lung. Several adaptations are found not only across isolates from a single species but also across species.

CF patients undergo many rounds of antibiotics to control lung infections, but symptomatic improvement is not necessarily accompanied by reduced bacterial counts. Unsurprisingly,

antibiotic resistance has been observed in CF isolates of a number of species. CF isolates of both *S. aureus* and *P. aeruginosa* form small colony variants (SCVs) as well (Besier et al., 2006; Folkesson et al., 2012; Kahl et al., 1998; 2003). *P. aeruginosa* develops a number of other phenotypes in the CF lung, including a mucoid or slimy phenotype, defects in quorum sensing, or loss of motility (Folkesson et al., 2012).

Our organism of interest, *E. coli*, is less commonly isolated from CF sputum, and it is less studied. *E. coli* is generally considered to be a transient, not a long-term colonizer (Høiby, 1974a) Nevertheless, our dataset demonstrates that *E. coli* can cause long-term infections. Antibiotic resistant and mucoid *E. coli* have previously been observed in CF sputum isolates and are frequently observed in our strain set. We have also observed phenotypically unstable SCV *E. coli*, with pure isolates that are streaked out on agar growing as a mixture of normal and SCV colonies; these have been termed springer variants.

4.2.2 Adaptive benefit of abnormal CF-associated phenotypes

The adaptive benefit to the bacterium of these CF lung-associated phenotypes cannot yet be tested directly. Mice with CFTR defects do not display a lung phenotype. The agar bead model of CF lung infection imitates a chronic lung infection that cannot be cleared but does not address the innate immune system abnormalities associated with CF. Ferrets and pigs carrying the most common CF-causing mutation have been created recently, and these are the first genetic models to replicate the human lung phenotype. While these animal models are promising, they are still under characterization and not yet suitable for infection studies (Fisher et al., 2011; Keiser and Engelhardt, 2011; Li and Engelhardt, 2003; Rogers et al., 2008; Sun et al., 2010). Nevertheless, a great deal can be inferred from *in vitro* work and longitudinal studies of patients.

The repeated rounds of antibiotics used in CF patient care are sufficient to explain the adaptive benefit of bacterial resistance. Antibiotic treatment may also explain some or possibly all of the SCV *E. coli* seen. Aminoglycosides have been shown to induce SCV in both *E. coli* and *Staphylococcus aureus*, and SCV of various species have been shown to be resistant to aminoglycosides (Clowes and Rowley, 1955; Eiff et al., 1997; Proctor et al., 1998). The antibiotic resistance of SCVs appears to be due to their slower growth, with slow growth being associated with antibiotic resistance in many species. Springer strains may represent a more flexible approach. Slower-growing individuals, which would produce an SCV phenotype when plated, survive antibiotic challenge better. Faster-growing phenotypic revertants would be more able to proliferate and compete for resources than bacteria from a strictly SCV strain.

Mucoidy obviously provides some benefit to *E. coli* in the CF lung, though it is not entirely clear how mucoidy is beneficial. Mucoid *P. aeruginosa* are resistant to certain chemical insults as well as to phagocytosis by macrophages, both stresses encountered in the CF lung (Bayer et al., 1991; Benincasa et al., 2009; Boucher et al., 1997; Ciofu et al., 2001; Govan and Deretic, 1996; Leid et al., 2005; Slack and Nichols, 1981). However, mucoid *E. coli* are neither resistant to phagocytosis nor resistant to antimicrobials relative to a non-mucoid parent strain. So far, mucoid *E. coli* have been shown to be resistant to desiccation and serum-mediated killing, as well as being slightly resistant to both reactive oxygen species and simulated intestinal fluid (Chen et al., 2004; Miajlovic et al., 2013; Ophir and Gutnick, 1994; Phan et al., 2013). Colanic acid overproduction is frequently the cause of the mucoid phenotype; knocking out colanic acid decreases serum resistance in an ExPEC strain (Miajlovic et al., 2013; Phan et al., 2013). Taken together, these traits may be sufficient to explain selection for mucoidy in the CF lung, but there may also be other factors that we have not yet discovered.

4.3 MATERIALS AND METHODS

4.3.1 Strain collection and culture methods

Strains were collected at University Hospital, Muenster by culturing expectorated sputum, nasal secretions, and throat swabs from patients with cystic fibrosis. Strains were isolated, identified, and characterized using standard microbiological techniques. *E. coli* isolates were provided to us as pure cultures. Strains were grown in LB for DNA isolation. Strains were grown on tryptic-soy agar (TSA) to determine colony morphology, as we have observed that the mucoid phenotype is more pronounced under these conditions. See Table 4.2 for more information about the strains.

4.3.2 Strain typing

Strains were typed by our collaborators using Enterobacterial Repetitive Intergenic Consensus (ERIC) PCR. We also typed all strains using multi-locus sequence typing (MLST) or a variant of MLST, either extended MLST or CH typing, as described below. MLST uses 500-bp sequences from 7 housekeeping genes to estimate relatedness between strains. These genes are core, rarely recombinant, and under strong purifying selection. Strains that are identical in all seven sequences are considered to be of the same sequence type, or ST, and are assumed to be relatively closely related, sharing a common ancestor within the last few thousand years (Wirth et al., 2006). For more fine-grained determination of clonality, extended MLST adds four virulence-associated genes, *fimH, fyuA, malX*, and *usp* (Jakobsen et al., 2009; Nakano et al., 2001; Ostblom et al., 2011; Weissman et al., 2012). These genes are known to be under positive selection and to change more rapidly than the genes used for MLST. As these are virulence-associated genes, they are particularly useful for typing pathogens. We also used an abbreviated

form of sequence typing, termed C-H typing (for *fumC-fimH* typing). Our lab has previously demonstrated that typing with one MLST gene, *fumC*, and one widely carried virulence-associated gene, *fimH*, can type strains to superior level of discrimination than standard MLST with less sequencing. See the UCC MLST website for standard MLST primers (Wirth et al., 2006) and Table 4.1 for other primers used.

All strains were typed by ERIC-PCR and CH typing. Strains of interest were further analyzed with standard or extended MLST.

All PCR was carried out with Taq JumpStart ReadyMix (Sigma), using the following program: 94 C for 5 m; 30 cycles of 94 C for 30 s, 58 C for 30 s, and 72 C for an appropriate extension time; 72 C for 10 m; and a 4 C hold. Extension time was 1 minute for every 1 kb of product plus an additional 30 seconds.

Table 4.1: Primers used in this study.

Name	Gene	Primer sequence	for	Source
fimF	fimH	CACTCAGGGAACCATTCAGGCA	PCR, seq	Weissman et al (2012)
fimR	fimH	CTTATTGATAAACAAAAGTCAC	PCR, seq	Weissman et al (2012)
fyuF	fyuA	GTTGATGACGGCGACATGATT	PCR, seq	Jakobsen et al (2008)*
fyuR	fyuA	CGCAGTAGGCACGATGTTG	PCR, seq	Jakobsen et al (2008)*
malF	malX	CCACGCAATACGCCAAAGC	PCR, seq	Jakobsen et al (2008)*
malR	malX	CGCGTGACGATTCCTTTTGG	PCR, seq	Jakobsen et al (2008)*
uspF	usp	GCTACTGTTTCCGGGTAGTG	PCR, seq	Nakano et al (2001)*
uspR	usp	CATGTAGTCGGGGCGTAAC	PCR, seq	Nakano et al (2001)*
mdo1	mdoH	CAAACGCGTCGTTCAACG	PCR, seq	This study
mdo2R	mdoH	TTATTGCGAAGCCGCATCC	PCR, seq	This study
mdo3	mdoH	GTAAAGGCGCGTCTGGAAC	seq	This study
mdo4	mdoH	CTTATTGGTCGCGATAAATACAG	seq	This study
mdo5	mdoH	GATGGAAGCCAACCCGAAC	seq	This study
mdo6	mdoH	CGGGCGTGATGTCTTATCT	seq	This study
mdo7	mdoH	CGATTGTCTTCTCGTTGATCC	seq	This study
mrdF	mrdB	CTGGGATAAAGTCCATCTCG	PCR, seq	This study
mrdR	mrdB	CACGCTTTTCGACAACATTTTC	PCR, seq	This study
mrd1	mrdB	CCTCTCTGGCCTTAGCTG	seq	This study
mrd2	mrdB	CAGCTAAGGCCAGAGAGG	seq	This study

* Primers modified slightly (length or small location shift) for annealing temperature and terminal bases

4.3.3 Genome sequencing

Highly purified genomic DNA from O-8 and O9 was provided to our collaborators, who performed genome sequencing with 454. Sequencing with Illumina technology was used for finishing. Contigs were assembled using Velvet. Genome sequencing with the Illumina miSeq was performed in-house using a resequencing protocol with UTI89 as the template genome.

4.3.4 Genome sequence analysis

Genome sequences were analyzed using scripts developed in our lab; many of these are included in the TimeZone analysis pipeline (Chattopadhyay et al., 2012). UTI89 was used as the reference genome, with each annotated ORF in UTI89 being used as an nBLAST query sequence to identify potential homologues in the O-8 and O-9 genomes. BLAST results were aligned using ClustalX. TimeZone modules were used to identify polymorphisms within aligned ORFs.

Polymorphic ORFs of interest from the first sequencing run (454/Illumina) were checked against miSeq results (both aligned sequences and fastq reads). The genes of greatest interest were sequenced using Sanger chemistry to confirm polymorphisms. See 4.3.2 for PCR conditions and Table 4.1 for primers used.

4.4 RESULTS

4.4.1 Characteristics of *E. coli* isolated from CF patient sputum

In order to characterize the prevalence, persistence, and phenotypic traits of *E. coli* isolated from the CF lung, routine samples from patients at 2 CF centers in Muenster, Germany were cultured and analyzed. 176 patients were included in the study. Patients were involved in the study for 6-90 months (median 6.8 y) and 4-69 visits (median 26 visits). *E. coli* was isolated from airway samples in more than a quarter of patients. Of those patients positive for *E. coli*, approximately 2/3 were positive at two or more visits (29 patients or 16.5% of total). Strains were typed using Enterobacterial Repetitive Intergenic Consensus (ERIC) PCR (Versalovic et al., 1991) to determine if these patients were colonized transiently or long-term. 19 patients showed long-term infection, as defined by isolation of strains with the same ERIC-PCR profile on 3 or more visits over at least 6 months (median 21 months). In almost all cases, profiles differed from patient to patient. This is not a case of a special and epidemic strain of *E. coli*, but rather a demonstration that at least some strains of *E. coli* are capable of infecting the CF lung long-term.

Table 4.2: Patient strains.

Patient	Isolates	Duration*	CF-associated phenotypes observed
A	15	34 mo	Mucoid, SCV
B	2	1 mo	
C	3	5 mo	
D	40	50 mo	Mucoid, antibiotic resistance**
E	1	-	
F	4	11 mo	
G	3	10 mo	
H	46	50 mo	SCV, antibiotic resistance
J	4	31 mo	
K	5	7 mo	Antibiotic resistance
L	3	9 mo	Antibiotic resistance
M	6	16 mo	
N	31	50 mo	SCV, springer, antibiotic resistance
O	33	46 mo	Mucoid, SCV, springer, antibiotic resistance
P	8	23 mo	Mucoid
Q	1	-	
R	6	15 mo	Mucoid, antibiotic resistance
S	10	29 mo	
T	2	6 mo	
U	5	19 mo	
V	6	24 mo	

*Duration indicates span of time from first to last *E. coli* isolate in the sample set.
**Antibiotic resistance defined as significantly increased resistance to any of a panel of common clinical antibiotics as compared to clonal strains from the same patient, implying within-patient evolution of resistance to at least one relevant antibiotic.

For this work, we were provided with 230 samples from the 18 patients with long-term *E. coli* colonization as well as 4 samples from 3 patients with transient colonization (see Table 4.2). Each patient was represented by 2-46 samples (median 6, mean 12.8) isolated over a period of 5-50 months (median 16.5, mean 23.2). A number of strains were antibiotic resistant. Variant colony morphologies were also observed, from SCV to springer to mucoid.

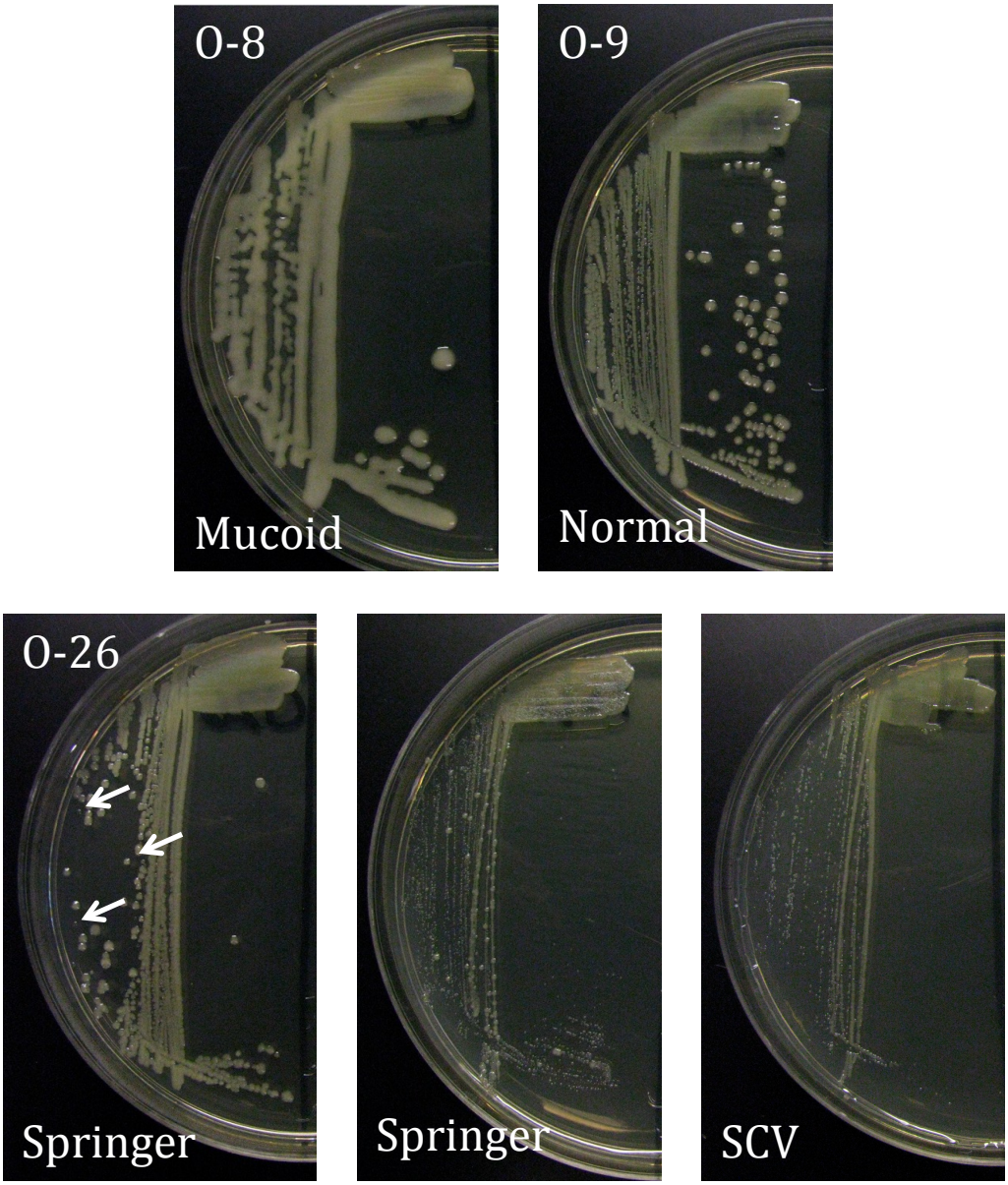

Figure 4.1: Colony morphology on TSA.
Colony morphology of 5 isolates from patient O. Note that the balance of SCV:normal colonies varies between springer isolates; arrows on O-26 (bottom left) point to scattered SVC colonies.

Colony morphology was media-dependent for a number of patient isolates, particularly intermediately mucoid isolates. Previous work demonstrated that mucoidy due to colanic acid overproduction is typically induced by lower incubation temperatures (20-25 C) and growth on minimal media (Sledjeski and Gottesman, 1996; Yeh and Chen, 2004). One of the few demonstrated selective advantages of colanic acid overproduction in *E. coli* is resistance to desiccation (Ophir and Gutnick, 1994). This suggests that the primary role of colanic acid is survival in the environment, at least for strains where colanic acid expression patterns resemble those of K-12.

No strains from the CF strain set showed induction of mucoidy in response to low temperature and minimal media, however. All strains with a mucoid phenotype on minimal glucose agar at 25 C possessed a mucoid phenotype on TSA at 37 C. Similar results were seen with a set of urinary tract isolates, and with very few exceptions, clinical isolates that were mucoid under one condition were mucoid under both conditions.

4.4.2 Phenotypic variation in *E. coli* strains from patient O

The isolates from patient O were of the greatest interest, as strains were present with ampicillin resistance, springer, and mucoid phenotypes (see Figure 4.3, Table 4.2). ERIC-PCR indicated that the patient isolates were clonal. Seven strains from this patient were typed by extended MLST and found to have the same type. CH typing of the remaining isolates from patient O showed 32 of the 33 isolates to be clonal. The strains from this patient constitute a closely related population that likely originated from a single strain and displays multiple phenotypes of interest.

Of the 32 strains from patient O, 10 are mucoid, 3 are springer/small colony variant, and 19 have normal morphology. Two strains were ampicillin resistant, one with normal morphology and one with an intermediately mucoid phenotype. O-8 and O-9, the first mucoid and ampicillin-resistant strains, respectively, were both isolated from a single patient sample, more than a year after the first *E. coli* was isolated from this patient. Both mucoid and non-mucoid strains of *E. coli* continued to be isolated from patient samples for the duration of observation. This clone of *E. coli*, in both its mucoid and non-mucoid variants, is capable colonizing the CF lung and persisting in the long term.

4.4.3 Genomic variation in strains from patient O

Two strains were selected for genome sequencing in order to identify mutations leading to the mucoid and ampicillin-resistant phenotypes. These were strains O-8 and O-9, respectively (Figure 4.1). Strains O-8 and O-9 were isolated from a single patient sample. These strains were sequenced by Illumina with resequencing on 454, as well as a resequencing pass with Illumina miSeq. ORFs in the resulting contigs were identified by nBLAST. UTI89 was used as the reference strain, as it is closely related to the patient isolates and is well annotated. The resulting sequences were aligned with ClustalX. Aligned sequences were analyzed for polymorphism between strains using a module of TimeZone, a genome analysis software package developed in our lab (Chattopadhyay et al., 2012). Based on the reference genome annotation, strains O-8 and O-9 shared 5,230 genes, with O-8 having an extra 8 ORFs (Tabel 4.3). In the shared genes, only 5 single nucleotide polymorphisms (SNPs) and a single 9-bp deletion were identified as the differences between the strains.

Next, the level of SNP and gene content differences between O-8 and O-9 was compared to that of strain O-8 and three sequenced uropathogenic strains that have diverse MLST profiles: the reference strain UTI89 from ST95; strain CFT073 from ST73; and strain NA114 from ST131 (Avasthi et al., 2011). A total of 4292 genes were found to be present in all four strains. As expected, the number of SNPs was the smallest between O-8 and UTI89, which belongs to the same clonal group, ST95 (Table 4.3). However, this level of polymorphism was over 250 fold higher than that found between the O-8 and O-9 strains. The difference in ORF content was also drastically lower between the O-8 and O-9 than between O-8 and any other strain, including the clonally-related UTI89.

Assembled contigs from O-8 were compared with three model ExPEC strains: UTI89 (uropathogen), CFT073 (uropathogen), and NA114 (ST131 antibiotic-resistant). Numbers above the diagonal represent the number of ORFs present in both strains. Numbers below the diagonal represent polymorphic nucleotides in aligned core genes between the two strains.

Table 4.3: Genomic differences among O-8, O-9, and 3 ExPEC strains.

nt changes in core genome	number shared ORFs				
	O-8	**O-9**	**UTI89**	**CFT073**	**NA114**
O-8		5326	5225	4899	4479
O-9	5		5217	4899	4479
UTI89	1233	1232		4904	4459
CFT073	27421	27420	28057		4540
NA114	39863	29862	40070	40026	

The strains from patient O shared a number of differences from UTI89, including a number of gene truncations (Table 4.4.) These truncations in the ancestral patient strain may be due to relaxed selective pressure, or the truncations may have contributed to the strain's ability to colonize the CF lung. If the truncations are due to relaxed selective pressure, we would expect the pattern of truncations to be random. DAVID is an algorithm for clustering genes by the

function of their encoded proteins to look for categories that are enriched in the test set. DAVID analysis of patient-specific truncations indicated two functional clusters, nucleotide binding and membrane-associated proteins. These were not significant ($p = 0.19$ and $p = 0.25$, respectively), however. It is of interest to note that the ORFs with only non-synonymous changes within patient O are also categorized as membrane-associated and nucleotide-binding. By DAVID, these ORFs were enriched for membrane-associated genes at $p = 0.042$.

All together, these results strongly suggest that O-8 and O-9 not only belong to the same MLST group but are very recent clonal derivatives that could have emerged within the patient O in the course of chronic persistence.

All differences identified by Illumina sequencing between the O-8 and O-9 genomes were subsequently confirmed by Sanger sequencing (Table 4.5, Figure 5). Detailed analysis has revealed that, out of the 5 SNPs, 4 were amino acid replacement mutations in 4 different ORFs (imp, gltX, dhaK and mrdB) and 1 was a premature stop-codon mutation in mdoH. The 9-bp in-frame deletion was within ycfT, and the presence/absence difference in 8 ORFs was the result of a single large deletion of approximately of 6.5 kb. The O-8/O-9 differences were distributed across the reference genome (Supplemental Figure 1), indicating that they had been acquired independently rather than in a single recombination event.

Based on the reference genome, non-synonymous changes in imp and mrdB and the large deletion had occurred in O-9, with the deletion affecting a cluster of 8 co-directionally transcribed chromosomal genes of hypothetical nature. All other changes - non-synonymous changes in gltX and dhaK, 9bp deletion in ycfT and premature stop-codon in mdoH - were predicted to be mutations in the O-8 strain.

Table 4.4: Truncations specific to patient O isolates.

GI	Gene name	DNA	Protein	Annotation
91071104	*YlaB*	2289del11		hypothetical with EAL domain
91071452	*ybjI*	571C>T	Q191X	haloacid dehalogenase-like hydrolase
91071764	UTI89_C1163	1101C>A	C367X	hypothetical
91071993	*YcgS*	113insG		provisional kinase subunit DhaL
91072814	UTI89_C2227	44insA		hypothetical
91073098	*gyrA*	580G>T	E194X	DNA gyrase
91073190	*usg*	559C>T	Q187X	putative, provisional dehydrogenase
91073686	*surE*	470T>A	W161X	provisional acid phosphatase
91074038	*ygiC*	399G>A	W133X	synthase/amidase
91074098	*yqjA*	524G>A	W175X	hypothetical membrane protein
91074500	*yrhA*	360delG		hypothetical
91074577	UTI89_C4029	532C>T	Q178X	hypothetical
91075281	*yjeA*	555insG		putative lysyl-tRNA synthetase
91075407	UTI89_C4885	703insC		hypothetical
91075459	UTI89_C4937	73insT		hypothetical putitive ATPase

The 29 other clonal isolates from patient O in this study were sequenced for the 7 coding differences seen between O-8 and O-9 (Table 4.6). Thirteen strains possessed neither the polymorphisms specific to O-8 nor those specific to O-9 (imp, mrdB, and the large deletion). Ten strains carried all four O-8-specific polymorphisms. None of the isolates had all O-9-specific mutations, but 5 isolates carried two of them (the imp SNP and 8-ORF deletion). Interestingly, 2 strains carried the O-9-like mrdB mutation and also a mutation in imp but of different nature than that found in O-9. Strains from 3 largest lineages ('non-mutated', O-8-like, and that carrying imp SNP and deletion) were present over a span of multiple years.

Thus, all differences seen between the O-8 and O-9 strains affected the structure or presence of coded proteins, with both strains acquiring mutational changes. Also, based on the O-8/O-9

polymorphisms at least 5 distinct clonal lineages could be identified in the patient O isolates, with the O-8-like bacteria being isolated throughout the period of observation.

The mutations in *mdoH*, and *mrdB* were of particular interest based on current knowledge. MrdB is a penicillin-binding protein and involved in mecillinam sensitivity (Clarke et al., 1999; Steenbergen et al., 1992; Tamaki et al., 1980). MdoH is not directly involved in capsule biosynthesis, but transposon disruption of *mdoH* has been associated with increased colanic acid synthesis in a lab strain (Ebel et al., 1997).

Table 4.5: Genes with strain-specific non-synonymous variation.

Strain	Gene	Accession	Nucleotide	Amino acid	Annotation*
O-8	*mdoH*	3993939	1011 C>A	C367X	periplasmic glycans synthesis
O-8	*ycfT*	3992701	374_382del	ΔAYA125_127	predicted membrane
O-8	*gltX*	3991693	1048 C>T	R350C	tRNA synthetase
O-8	*dhaK*	3992242	1063 G>A	G355R	glycerolipid metabolism
O-9	*imp*	3993622	1762 T>C	S588P	organic solvent tolerance
O-9	*mrdB*	3989726	1015 G>A	V339I	rod shape-determining protein
O-9	UTI89_C4936	3990785	deletion	deletion	hypothetical
	UTI89_C4937	3990786			histidine kinase-like
	UTI89_C4938	3990787			DNA methylase-like
	UTI89_C4939	3990827			hypothetical
	UTI89_C4940	3990828			TraG-like
	UTI89_C4941	3990829			TraG-like
	UTI89_C4942	3990830			ImpA domain
	UTI89_C4943	3990831			hypothetical

*Annotation based on gene annotations in DAVID database

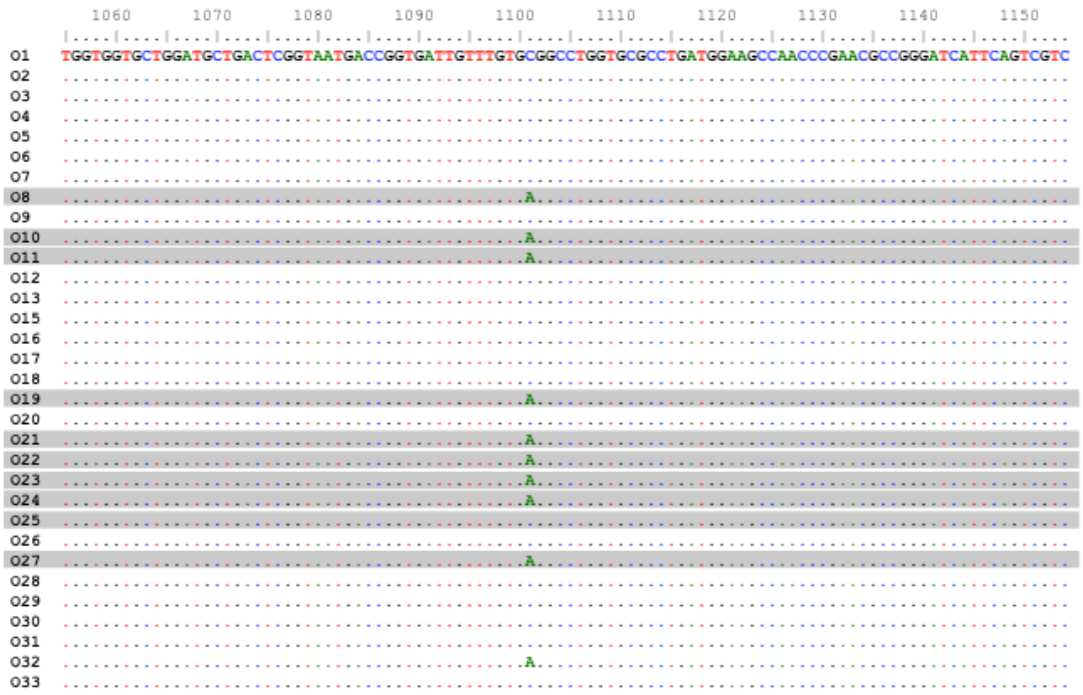

Figure 4.2: Alignment of *mdoH* sequence in the vicinity of nonsense mutation. Shading indicates strains with a mucoid phenotype on TSA plates.

4.4.4 Association of mdoH truncation and mucoidy in strains from patient O

Previous work has shown that transposon disruption of *mdoH* results in a mucoid phenotype. A nonsense mutation resulting in the truncation of MdoH would be expected to have similar effects as a transpon. I sequenced *mdoH* in full for the 32 clonal strains from patient O. The nonsense mutation was strongly associated with the mucoid phenotype (see Figures 4.2 and 4.3). No other polymorphisms were found in *mdoH* in any of the strains from this patient.

Both strains which broke the mucoidy-C367STOP association were isolated late in the infection. O25, which has an intermediately mucoid phenotype with an intact *mdoH*, was isolated 34 months after the first *E. coli* sample and may be a new emergence of mucoidy via a novel mechanism. O32, which has a normal phenotype with a truncated *mdoH*, arose 28 months after

the isolation of O-8. This revertant strain is almost certainly carrying a suppressing mutation in another gene. In mucoid *P. aeruginosa* with *mucA* mutations, pseudorevertants that lose the mucoid phenotype but retain the *mucA* mutation are commonly observed, both in patient isolates and in the lab (DeVries and Ohman, 1994).

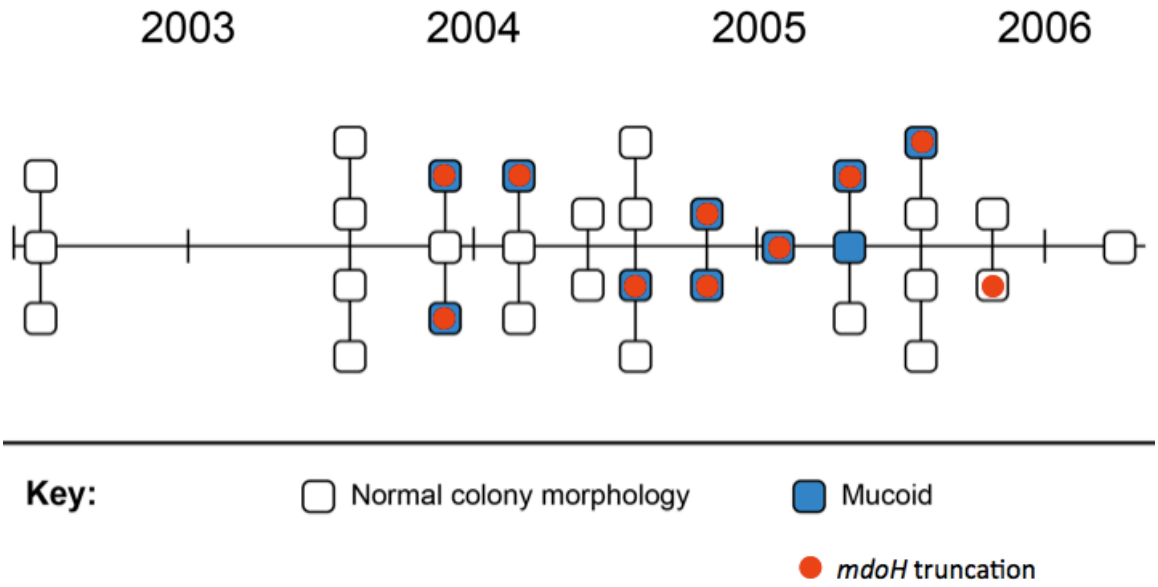

Figure 4.3: Association of nonsense mutation and mucoidy.
Clonal isolates from patient O ordered chronologically. Color of square indicates colony morphology (white: wild type, blue: mucoid). Red dot indicates presence of C367X polymorphism in *mdoH*.

4.5 CONCLUSIONS

E. coli is a common inhabitant of the cystic fibrosis lung, and it can cause both transient and long-term infections. Like other bacteria that chronically infect the CF lung, variant phenotypes are frequently observed over the course of infection. As seen with the isolates from patient O, these variants are also capable of chronic colonization. The overall prevalence of *E. coli* in this

dataset is substantially higher than what has been published previously. This may be due to differences in patient genetics, patient treatment, or other factors specific to the study sites in Muenster. The discrepancy may also be influenced by the common belief that *E. coli* in CF patient sputum isolates may be a contaminant or is likely to be a transient infection, resulting in underreporting of *E. coli*.

In isolates from patient O, mucoidy is strongly associated with truncation of *mdoH*. *mdoH* truncation has been demonstrated to cause mucoidy in the lab, but this mechanism has not previously been associated with mucoidy in wild isolates.

CHAPTER 5:

mdoH TRUNCATION IS A CAUSE OF MUCOIDY IN *E. COLI* CF ISOLATES

5.1 ABSTRACT

The development of mucoidy in *E. coli* is typically associated with mutations in *lon*, a negative regulator of the Rcs phosphorelay. I have observed two discrete instances of a different mutation associated with mucoidy in the CF lung, truncation of *mdoH*. Mucoidy can be complemented by transformation with a plasmid carrying the wild-type *mdoH*. Though previous work showed *mdoH* acting through the Rcs phosphorelay to induce colanic acid overproduction, in the strains from patient O, Rcs activation is regulating a different capsule.

5.2 BACKGROUND

5.2.1 *mdoH* and the Rcs phosphorelay

The *mdoGH* locus consists of two overlapping ORFs and is regulated by osmolarity (Bohin and Kennedy, 1984; Lacroix et al., 1991; 1989). The mdo genes are involved in the synthesis of membrane-derived oligosaccharides (MDOs), which are modified glucose polymers exported to the periplasm to protect the cell from osmotic stress. MdoH is constitutively produced at low levels and activated by osmotic stress, allowing the cell to respond immediately (Debarbieux et al., 1997). Through an unknown mechanism, lack of functional MdoH activates the Rcs phosphorelay (Ebel et al., 1997).

The Rcs phosphorelay, named for regulation of capsule synthesis, is a signaling cascade originally identified in a mutagenesis screen for altered expression of colanic acid biosynthesis genes (Gottesman et al., 1985). Excessive capsule production, stimulated by Rcs dysregulation, is responsible for the slimy phenotype of mucoid strains of *E. coli.* Further studies of the Rcs

phosphorelay have shown that it affects some 5% of the *E.* coli genome, including flagellar genes, the cell-division gene *ftsZ*, type 1 fimbriae, and in some strains, group I capsule production (Ferrières and Clarke, 2003; Hagiwara et al., 2003; Jayaratne et al., 1993; Laubacher and Ades, 2008; Lehti et al., 2012; Majdalani and Gottesman, 2005; Navasa et al., 2013; Russo and Singh, 1993; Schwan et al., 2007; Sledjeski and Gottesman, 1996; Stout and Gottesman, 1990). Mutations in *lon*, which codes for a negative regulator of the Rcs phosphorelay, are the most common cause of mucoidy in studies to date (Markovitz, 1964; Skorupski et al., 1988).

Constitutive Rcs activation therefore has a host of effects, some of which are deleterious to the cell, dysregulation of cell division in particular (Majdalani and Gottesman, 2005). Early studies of mucoid *E. coli* remarked on the difficulty of studying mucoid strains, as they tend to revert to a normal colony phenotype after a few passages on plates under laboratory conditions (Macone et al., 1981). In the face the steep costs of constitutive Rcs activation, it seems that mucoidy or some other effect must confer significant benefits during chronic infection of the CF lung for the phenotype to be maintained.

5.3 MATERIALS & METHODS

5.3.1 Sequencing

Sequences were amplified using JumpStart Taq ReadyMix (Roche) and cleaned up using ExoSAP (Affymetrix), as per manufacturer instructions. The resulting DNA was Sanger sequenced by GeneWiz. See Table 4.1 for primers used in amplification.

5.3.2 Plasmids

Complementation vectors (wt and truncated) were constructed by TOPO-TA cloning into pCR2.1 (Invitrogen. *mdoH* from O-26 (wt) and O-8 (truncated) were amplified using CL68 and CL69 with 2.5 units JumpStart Taq ReadyMix (Sigma-Aldrich) supplemented with 2 units Pfu Turbo (Agilent) for increased fidelity. PCR was run with a standard program (94 C for 3 m; 30 cycles of 94 C for 30 s, 58 C for 30 s, and 72 C for 3.5 m; 72 C for 10 m) After electrophoresis on a 1% agarose gel to confirm appropriate size, samples were incubated for 10 m at 72 C with an additional 2.5 u JumpStart in order to produce TA overhangs. Samples were immediately used in the TOPO-TA reaction as per manufacturer instructions. The vector control was constructed by excising *mdoH* using the NdeI and SacI sites of the amplification primers (NEB). The gel-purified vector backbone was made blunt-ended by treatment with Antarctic phosphatase (NEB), and after heat inactivation of the phosphatase was re-ligated using T4 DNA ligase (NEB). Plasmids were transformed into chemically competent AAEC191a cells. Plasmids were confirmed by restriction digest.

5.3.3 Transformation of clinical strains

Cells were grown in LB to mid-log phase, and then washed twice with ice-cold 10% glycerol. Pellet was resuspended in retained glycerol solution. Cells were incubated 10-15 m on ice with plasmid; they were then transferred to a chilled 1 mm cuvette and electroporated using a BioRad GenePulserII at 1.8 kV, 25μF, and 200 ohms. Ice-cold SOC was added and cells were allowed to recover at least 1 h at 37 C. 10% and 1% of the recovered cells were plated onto TSA agar containing appropriate antibiotics and grown overnight at 37 C. Transformed strains were

grown on TSA containing 0.1 mg/mL ampicillin and 0.025 mg/mL kanamycin or on blood agar to score phenotype. Plates were spread with 40 μL 0.1 M IPTG for induction.

5.3.4 Preparation of extracellular polysaccharide for sugar assays

O-8 and O-9 were compared with a number of strains that had previously been characterized as colanic acid-producing and non-mucoid (CFT073, JJP90, 83972, and RS218) as well as with the closely-related UTI89. Cultures were grown overnight in LB before being split. Cultures were plated at 10^{-7}, 10^{-8}, and 10^{-9} for cfu/mL calculations. 1 mL of each culture was spun down, resuspended in 500 μL dH2O, and frozen for protein quantification using the BCA assay (Pierce). 100 mL of culture was centrifuged for 5 m at 8000 x and the supernatant used for measurement of soluble EPS (the soluble sample). 100 mL of uncentrifuged culture (the total sample) and the soluble sample were boiled for 15 m. Once cool, samples were reconstituted to original volume with dH_2O. Three volumes of 70% ethanol was added to precipitate EPS. After overnight incubation at 4 C, samples were centrifuged for 30 m at 10 000 x g and 4 C. The resulting pellet was dried at room temperature for 30-60 m. The precipitated EPS was dissolved in 1 mL dH_2O and stored at 4 C.

5.3.5 Fucose assay

6 mL of 6:1 concentrated sulfuric acid:dH_2O was added to borosilicate test tubes containing 10-100 μL of precipitated EPS, standards, or blank. After mixing, these were incubated 20 m in a boiling water bath. Absorbance was measured at 396 nm and 427 nm for caramelization color correction. 15-20 m after addition of 100 μL 1M L-cysteine and mixing, absorbances were

measured again. Corrected OD396 – corrected OD427 was used to determine fucose concentration (Obadia et al., 2007).

5.4 RESULTS

5.4.1 *mdoH* of mucoid isolates from other patients

Is disruption of *mdoH* unique to mucoid isolates from patient O, or is it found elsewhere in our sample set? 6 other patients had at least one mucoid isolate, and I sequenced *mdoH* for clonal pairs of mucoid and non-mucoid isolates from the same patient. Paired strains were from the same visit where possible and from the nearest visit with a non-mucoid isolate when necessary. For five of the patients, there was no intrapatient variation in *mdoH*. The mucoid isolate from patient A, however, had a nonsense mutation. This SNP was distinct from the nonsense mutation observed in strains from patient O (see Figure 5.1). The nonsense mutations in mucoid strains from patients A and O were the only SNPs observed among clonal strains from any single patient. This absence of within-patient variation other than the two nonsense mutations implies that *mdoH* is not mutation-prone.

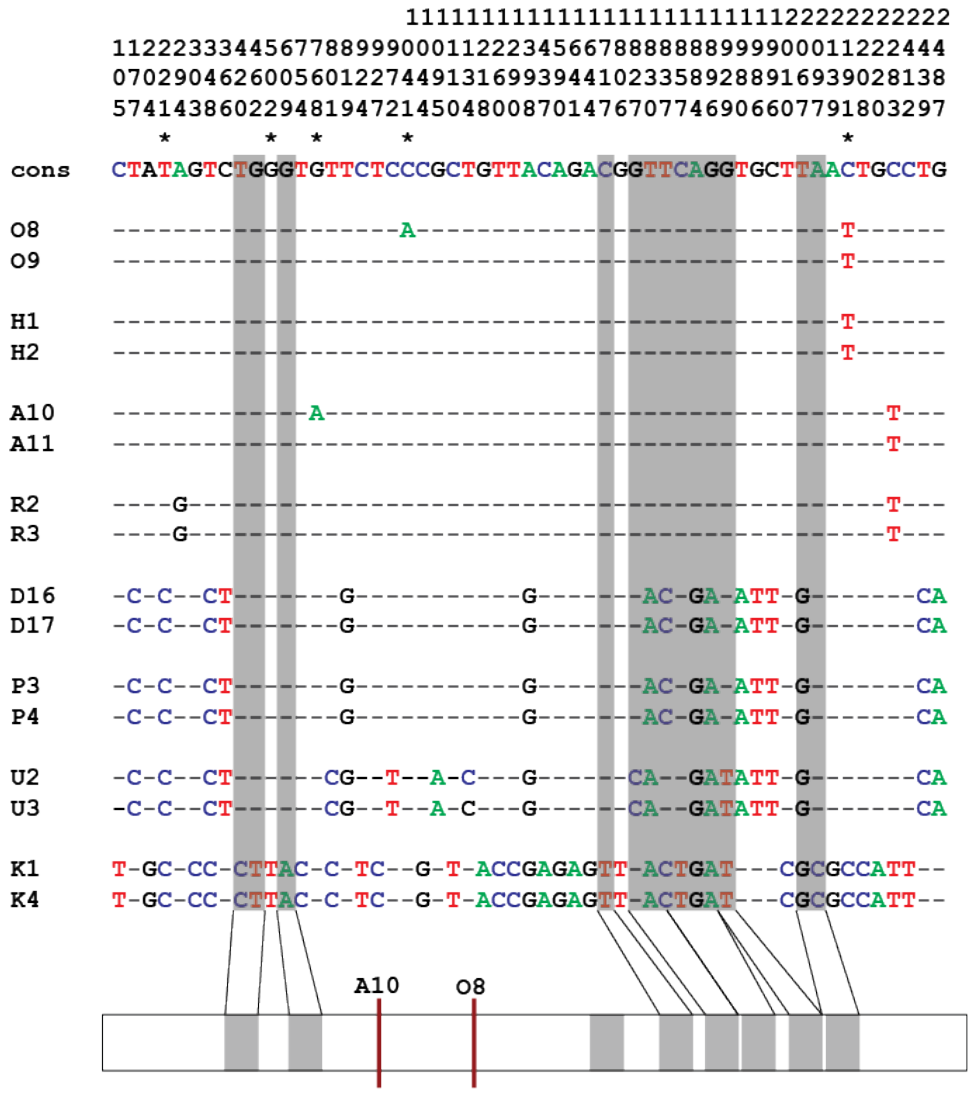

Figure 5.1: *mdoH* truncation in patient strains. Polymorphic positions in *mdoH* from seven patient pairs of mucoid and non-mucoid strains. Nonsynonymous polymorphisms denoted with asterisk. Transmembrane domains indicated by shading. Note that within-patient differences are only seen in patients A and O, with both differences being premature codon mutations.

5.4.2 *mdoH* truncation is the cause of the mucoid phenotype

The mechanism of Rcs activation due to *mdoH* truncation is unknown. Therefore, complementation experiments were designed to test for both recessive and dominant negative effects of the *mdoH* truncation. Initial efforts centered around construction of medium copy-number plasmids containing *mdoH* under its native promoter. This was unsuccessful, however, as the insert-containing plasmids that were recovered proved to have non-synonymous mutations within *mdoH*. As MdoH is a large protein (857 amino acids) and predicted to be transmembrane, there may be problems with toxicity or aggregation when it is overexpressed. Later cloning was designed for expression under an inducible promoter to minimize any such effects.

Plasmids were constructed to examine the effect of *mdoH* complementation *in trans* using TOPO-TA cloning with pCR2.1. Variants were constructed containing both the normal *mdoH*, amplified from O-26 (pCR2.1-mdoH.wt) and the truncated *mdoH*, amplified from O-8 (pCR2.1-mdoH.X). Both O-8 and O-26 were transformed with pCR2.1-mdoH.wt, pCR2.1-mdoH.X, and pCR2.1 as the empty vector control.

O-8 carrying pCR2.1-mdoH.wt displayed normal colony morphology after overnight growth, whereas O-8 carrying the vector control or truncated *mdoH in trans* displayed the same mucoid phenotype as the parental strain. No significant differences in colony morphology were observed between O-26 and O-26 carrying any of the plasmid constructs. This rules out a dominant negative effect of *mdoH* truncation. On LB and TSA, complementation in O-8 was most pronounced in fresh transformants. After passage, the phenotype became much more subtle. Complementation was consistent and dramatic on blood plates, however. The reason for this difference in complementation on different media is unknown. One explanation may be nutrition,

as nutrient balance and availability are known to affect capsule production (Yeh and Chen, 2004).

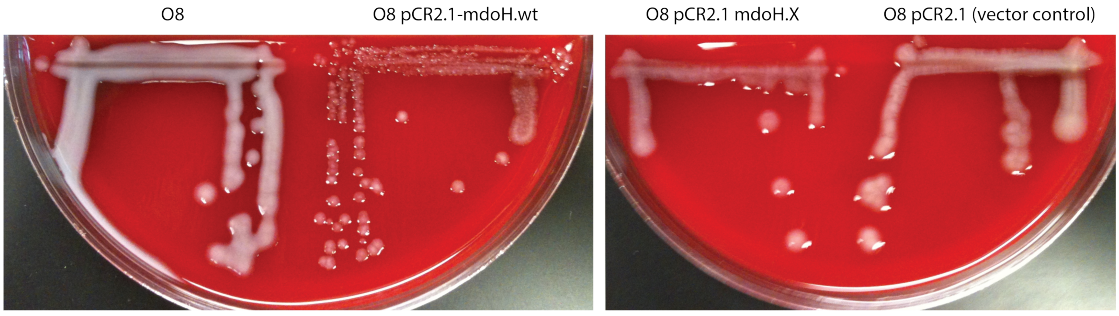

Figure 5.2: Complementation of *mdoH* in trans. O-8 and complemented O-8 on blood agar after 24 h incubation at 37 C.

5.4.3 Mucoidy is not due to colanic acid overproduction

Previous work with *mdoH* disruption suggested this increased colanic acid synthesis, as transposon insertion in cells carrying a *cpsB::lacZ* reporter fusion resulted in increased β-galactosidase activity (Ebel et al., 1997). In order to confirm that this was the case in our strains of interest, fucose in extracellular polysaccharide was quantitated, as has been described previously (Obadia et al., 2007). Colanic acid is a capsular polysaccharide present in a wide range of *E. coli*. The colanic acid repeat unit consists of 2 galactose, 2 glucose, 2 fucose, and 1 glucuronic acid (Goebel, 1963). Quantification of fucose and of uronic acid are both accepted proxies for colanic acid. Total and cell-free sugars were quantified for the purposes of increasing the specificity of colanic acid identification. Colanic acid is variably attached to the cell surface, with a significant minority of polysaccharide chains being unattached and only loosely

associated with the cell. Fucose is only found in a few capsules, and expected amounts of fucose with a significant fraction being cell-free strongly suggests colanic acid.

Control non-mucoid strains had the expected amounts of fucose per sample, with approximately a third of that being cell-free (Figure 5.3). This fraction corresponds with previous measurements, with total amounts also being congruent with previous work. Levels in the strains from patient O and the UTI89 reference strain were significantly lower, however. Additionally, the mucoid strain O-8 did not display elevated levels of fucose relative to other strains. These results rule out colanic acid overproduction as the mechanism of mucoidy in O-8. The low levels of fucose seen in UTI89 indicate that lack of colanic acid synthesis is not a patient-specific trait.

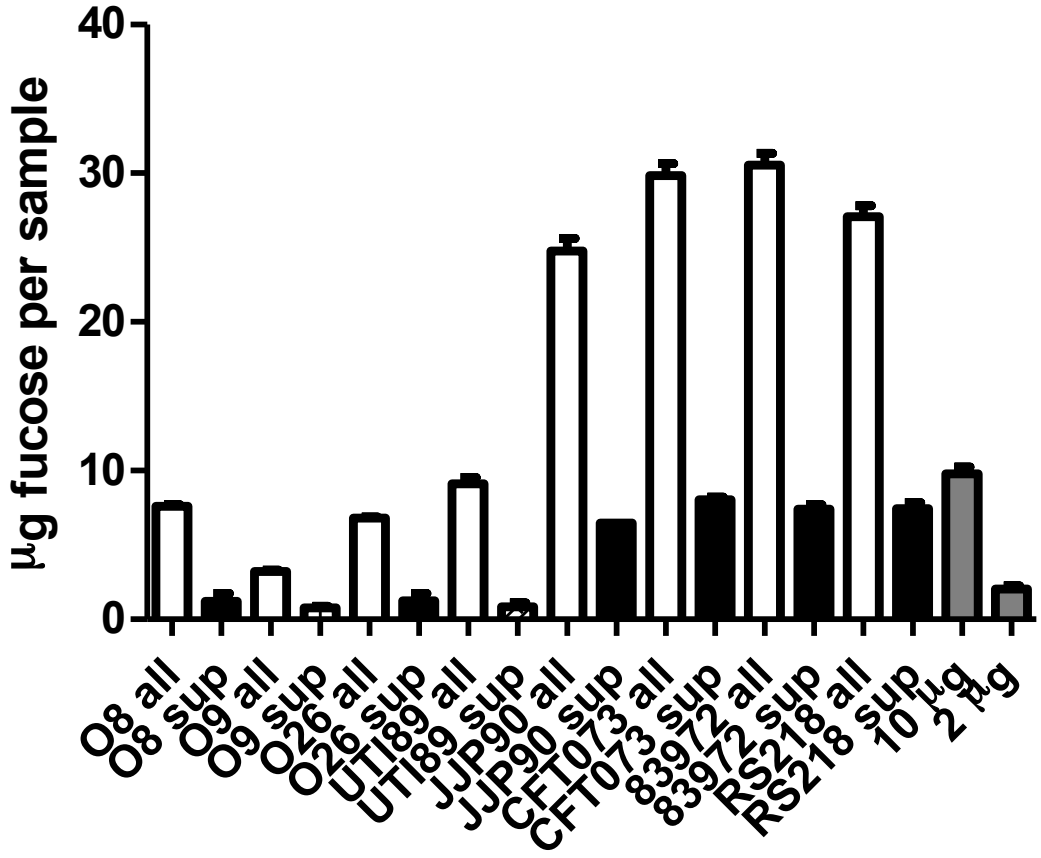

Figure 5.3: Sugar content of strains.
Micrograms fucose per sample, experiment run in triplicate; all samples normalized to O-8 by cfu of input culture. Total samples are in white, supernatant samples are in black, standards are in grey.

However, this defect is also not shared by all strains from ST 95; RS218 showed expected levels of colanic acid-associated sugars for a non-mucoid strain and is a member of ST95. Examination of the *cps* locus in UTI89 and in the sequenced patient strains showed no obvious defect in the promoter nor any truncation or deletion of proteins in the *cps* gene cluster. Serotype

also fails to explain this difference in colanic acid production. UTI89 has K1 capsule; this is a group 2 capsule and can be coexpressed with colanic acid (Keenleyside et al., 1993; Whitfield, 2006). However, while Rcs is canonically a regulator of colanic acid production, the Rcs phosphorelay has been shown to affect synthesis of capsule polysaccharides other than colanic acid (Jayaratne et al., 1993; Navasa et al., 2013). This may suggest a role for another capsular polysaccharide in the O-8 mucoid phenotype.

5.5 CONCLUSIONS

5.5.1 *mdoH* truncation results in a mucoid phenotype in isolates outside the laboratory

lon is a negative regulator of the Rcs phosphorelay, and mutations of *lon* are the dominant mechanism of mucoidy in laboratory strains of *E. coli*. Previous work characterizing naturally-occurring and chemically induced mutations in *E. coli* causing mucoidy have identified *lon* mutations. The screen that identified *mdoH* truncation as a cause of mucoidy was designed specifically to rule out mutations that could be complemented by *lon* in trans (Ebel et al., 1997). This study represents the first time that *mdoH* truncation resulting in mucoidy has been observed outside the lab.

5.5.2 Possible benefit of mucoidy to strain O-8

Mucoidy is not without its costs; synthesizing additional exopolysaccharides is a not insignificant metabolic drain. During propagation in the lab, mucoid strains of *E. coli* typically revert to a normal phenotype within a few passages. However, in the context of certain infections, this phenotype is apparently selected for, and in the context of the CF lung it may persist for months to years. I initially assumed that the increased exopolysaccharide itself was

somehow protective in the context of the CF lung, acting to shield the bacteria in some fashion. In *P. aeruginosa*, increased biofilm formation is associated with increased antibiotic resistance (Ciofu et al., 2001). However, when mature colonies of the clinical isolates were challenged with antibiotics or secreted factors of *P. aeruginosa*, the mucoid strain O-8 was no more resistant than other strains from the same patient. Under the conditions tested, mucoidy did not confer protection from soluble chemicals. However, in other strains and species, mucoidy has also been shown to provide protection from phagocytosis, dessication, and serum. The question remains: does the mucoidy itself provide a direct benefit to strain O-8?

Other work conducted by members of the lab demonstrated that the mucoid strain O-8 produces a more structured and thicker biofilm than the non-mucoid strain O-9. Formation of a biofilm in and of itself is of benefit to bacteria during chronic infection. Bacteria in a biofilm are better able to stick together and stick to a substrate, enhancing persistence. The extracellular matrix and cohesive nature of the biofilm also protect microbes from immune cells. It may be that simply building a better biofilm is enough to confer a selective advantage for mucoidy in the CF lung, even with the biochemical costs of polysaccharide overproduction. However, *mdoH* truncation leading to Rcs activation has a plethora of effects beyond induction of capsular polysaccharide biosynthesis.

Lack of normal MdoH triggers Rcs activation, directly or indirectly, which causes a host of downstream effects (Ferrières and Clarke, 2003; Francez-Charlot et al., 2004; Gervais et al., 1992; Jayaratne et al., 1993; Laubacher and Ades, 2008; Lehti et al., 2012; Majdalani and Gottesman, 2005; Navasa et al., 2013; Schwan et al., 2007). This pathway is essential for normal biofilm formation, and it regulates cell division as well as repressing motility (Ferrières and Clarke, 2003; Francez-Charlot et al., 2004; Gervais et al., 1992). Flagella are strongly

immunogenic, and loss of motility is common in CF isolates of *P. aeruginosa*. It may be that the selective pressure is not for mucoidy but instead for loss of flagella, for example, with mucoidy simply being a side effect of this particular mechanism of flagellar loss of function. As O32, one of the later-isolated strains from patient O, carries a truncated *mdoH* but does not express a mucoid colony phenotype, it would be interesting to examine the expression of genes downstream of the Rcs phosphorelay in this strain versus mucoid and normal strains from the same patient. If Rcs activity were increased in the O32 revertant strain and mucoid strains relative to other non-mucoid strains, this would imply that the main benefit of *mdoH* inactivation is Rcs activation rather than mucoidy per se.

Mucoid strains of *P. aeruginosa* are commonly found during chronic infection of the CF lung, and this phenotype is often due to disruption of the *mucA* gene (Boucher et al., 1997; DeVries and Ohman, 1994; Doggett et al., 1964; Høiby, 1974b; Martin et al., 1993). As *mucA* represses an alternate sigma factor, disruption of this gene has pleiotropic effects (Wu et al., 2004). Non-mucoid revertants are commonly found later in infection, but many of these phenotypic revertants still carry a disrupted *mucA* (DeVries and Ohman, 1994). In some cases, it may be that *mucA* truncation confers a selective advantage other than mucoidy, leading to selection against exopolysaccharide production along with selection to maintain the *mucA* mutation. Further characterization of persistent non-mucoid revertants from patient O will lead to a better understanding of the selective pressures operating on *E. coli* in the CF lung.

CHAPTER 6:

DISCUSSION/ FUTURE DIRECTIONS

6.1 Using natural variation to explore pathoadaptation

I examined natural variation within *E. coli* in order to get a better understanding of the selective pressures operating on this organism in its various niches. Two approaches were used: a directed examination of a single operon across many genomes and a broad examination of all variation present across the genomes of two isolates. The directed examination of *fim* pointed to the potential importance of non-coding variation in adaptation to pathogenic niches, as well as highlighting the qualitatively different nature of short-term variation as opposed to long-term variation. The broad examination of the genomes of two clonal isolates from one CF patient again highlighted the strongly non-synonymous nature of recent variation. This genomic analysis also identified the *mdoH* mutation responsible for mucoidy in isolates from patient O, which is the first time this mutation has been shown to cause mucoidy outside of the laboratory.

6.2 Further examination of the *fim* operon

On a species level, the findings from the small set of genomes were very consistent with the findings from the larger set, even though the larger set included many more genome sequences than the pilot set and distribution of pathotypes differed greatly. It may be interesting to see if any further non-coding/synonymous hotspot nucleotides near translation start sites crop up with further expansion of the input set of operons. This is not likely to be a high-reward activity, however, given the relative paucity of new hotspots when the dataset was expanded. Nonetheless, automation of the pipeline could contribute to further understanding of the *fim* operon and the role of Fim in *E. coli*.

In terms of analyzing closely-related within-ST polymorphisms, expanding the dataset was quite fruitful. The 7-fold increase in input size corresponded to a more than 10-fold increase in the number of unique *fim* operons with comparable operons in clonal strains, and this increased the number of polymorphisms observed to the point that meaningful statistical analysis could be performed. I believe that adding further sequences to this analysis will be highly fruitful. Similarly, the current dataset will become more useful as papers are published on the currently available genomes, as strains are likely to go from being annotated only as members of a specific study to having pathotype data available. For example, the Broad Institute Defensins project consists of 228 strains from UTIs along with rectal isolates from patients with a UTI and patients who have never had a UTI. It is possible that by subsetting the closely-related within-ST operon clusters by their association with pathotype, further patterns will emerge.

6.3 Examination of variation within patient O

The identity of the capsule being produced by the mucoid strain O-8 is a curiosity, but it is not necessarily germane to the interests of this lab or to the health of CF patients. As mentioned previously, mucoidy itself is not necessarily adaptive, or at least may not provide sufficient advantage to overcome the energy drain represented by overproduction of exopolysaccharides. It is entirely possible that selection for mucoidy is simply a side effect of selection for a separate trait such as loss of flagella. On the patient side, there is not yet any evidence that persistent *E. coli* infection or colonization with mucoid *E. coli* has any effect on patient prognosis. Adult CF patients typically have a complex and stable lung microbiome, with few aspects of the microbiome accounting for significant fractions of measures of patient health. *E. coli* isolates, or at least most *E. coli* isolates, may turn out to be merely squatters and opportunists, neither

helpful nor particularly harmful. Should future work indicate a role for persistent *E. coli* infection or mucoidy of *E. coli* in patient prognosis, the frozen isolates will be available for the issue to be revisited.

Our next steps focus on further exploration of genomic variation within isolates from patient O. Our lab is preparing to sequence 4 additional strains from this patient with varying phenotypes, including strain O32, the non-mucoid isolate carrying the C367X truncation in *mdoH*. Identifying the compensatory mutation associated with loss of mucoidy will shed further light on mucoidy in *E. coli* CF isolates. These sequences will also help us get a clearer idea of the rates and types of mutations that are occurring within infection and may identify mutations responsible for antibiotic resistance and/or springer/SCV colony morphology in these patient strains.

6.4 Natural variation as natural experiment

Examining natural variation allows us to answer questions that we had not thought to ask, rather than limiting ourselves to the hypotheses that we are able to come up with. This will illuminate novel pathways involved in both pathogenic and commensal lifestyles, as well as highlighting new mechanisms for achieving the same outcome, as with decreased FimH secretion/translation. Due to the ever-decreasing cost of genome sequencing, we have access to an immense amount of data that is constantly increasing.

This wealth of data can easily become a glut, however, if we do not have tools to efficiently analyze these sequences and identify the most interesting/most likely positive results. As seen in Figure 2.4, even considering a very restricted subset of variation can result in thoroughly unwieldy sets of possible experimental targets. On the data handling side, clever and flexible

programming can be used to take most of the load. Growing interest in programming and in biology-specific libraries like Biopython and BioPerl means that an increasing number of people have both the knowledge and the skills to create scripts to take care of relatively routine tasks like downloading new genome sequences, BLASTing for genes of interest, and running comparisons.

Development of robust pipelines and better sharing of metadata on published genomes will open up a wealth of new possibilities for analysis. The biggest problem we're likely to encounter is deciding which of the many promising results to pursue first. Nature has already provided the experiments in abundance; now it's up to us to analyze the data and make sense of it all.

REFERENCES:

Abraham, J.M., Freitag, C.S., Clements, J.R., and Eisenstein, B.I. (1985). An invertible element of DNA controls phase variation of type 1 fimbriae of Escherichia coli. Proc Natl Acad Sci USA *82*, 5724–5727.

Adlerberth, I., Hanson, L.A., Svanborg, C., Svennerholm, A.M., Nordgren, S., and Wold, A.E. (1995). Adhesins of Escherichia coli associated with extra-intestinal pathogenicity confer binding to colonic epithelial cells. Microb Pathog *18*, 373–385.

al-Wali, W., Hamilton-Miller, J.M., Joshi, S., and Brumfitt, W. (1989). A case of recurrently sexually transmitted urinary tract infection. Genitourin Med *65*, 397–398.

Alsam, S. (2006). Escherichia coli interactions with Acanthamoeba: a symbiosis with environmental and clinical implications. J Med Microbiol *55*, 689–694.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403–410.

Angus, D.C., Linde-Zwirble, W.T., Lidicker, J., Clermont, G., Carcillo, J., and Pinsky, M.R. (2001). Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. Crit Care Med *29*, 1303–1310.

Antão, E.-M., Wieler, L.H., and Ewers, C. (2009). Adhesive threads of extraintestinal pathogenic Escherichia coli. Gut Pathog *1*, 22.

Aprikian, P., Interlandi, G., Kidd, B.A., Le Trong, I., Tchesnokova, V., Yakovenko, O., Whitfield, M.J., Bullitt, E., Stenkamp, R.E., Thomas, W.E., et al. (2011). The Bacterial Fimbrial Tip Acts as a Mechanical Force Sensor. Plos Biol *9*, e1000617.

Aprikian, P., Tchesnokova, V., Kidd, B., Yakovenko, O., Yarov-Yarovoy, V., Trinchina, E., Vogel, V., Thomas, W., and Sokurenko, E. (2007). Interdomain interaction in the FimH adhesin of Escherichia coli regulates the affinity to mannose. J Biol Chem *282*, 23437–23446.

Avasthi, T.S., Kumar, N., Baddam, R., Hussain, A., Nandanwar, S., Jadhav, S., and Ahmed, N. (2011). Genome of multidrug-resistant uropathogenic Escherichia coli strain NA114 from India. J Bacteriol *193*, 4272-4273.

Bahrani-Mougeot, F.K., Buckles, E.L., Lockatell, C.V., Hebel, J.R., Johnson, D.E., Tang, C.M., and Donnenberg, M.S. (2002). Type 1 fimbriae and extracellular polysaccharides are preeminent uropathogenic Escherichia coli virulence determinants in the murine urinary tract. Mol Microbiol *45*, 1079–1093.

Bayer, A.S., Speert, D.P., Park, S., Tu, J., Witt, M., Nast, C.C., and Norman, D.C. (1991). Functional role of mucoid exopolysaccharide (alginate) in antibiotic-induced and polymorphonuclear leukocyte-mediated killing of Pseudomonas aeruginosa. Infect Immun *59*, 302–308.

Benincasa, M., Mattiuzzo, M., Herasimenka, Y., Cescutti, P., Rizzo, R., and Gennaro, R. (2009). Activity of antimicrobial peptides in the presence of polysaccharides produced by pulmonary pathogens. J Pept Sci *15*, 595–600.

Besier, S., Smaczny, C., Mallinckrodt, von, C., Krahl, A., Ackermann, H., Brade, V., and Wichelhaus, T.A. (2006). Prevalence and Clinical Significance of Staphylococcus aureus Small-Colony Variants in Cystic Fibrosis Lung Disease. Journal of Clinical Microbiology *45*, 168–172.

Bloch, C.A., Stocker, B.A., and Orndorff, P.E. (1992). A key role for type 1 pili in enterobacterial communicability. Mol Microbiol *6*, 697–701.

Blumenkrantz, N., and Asboe-Hansen, G. (1973). New method for quantitative determination of uronic acids. Anal. Biochem. *54*, 484–489.

Blumer, C., Kleefeld, A., Lehnen, D., Heintz, M., Dobrindt, U., Nagy, G., Michaelis, K., Emödy, L., Polen, T., Rachel, R., et al. (2005). Regulation of type 1 fimbriae synthesis and biofilm formation by the transcriptional regulator LrhA of Escherichia coli. Microbiology (Reading, Engl) *151*, 3287–3298.

Bohin, J.P., and Kennedy, E.P. (1984). Mapping of a locus (mdoA) that affects the biosynthesis of membrane-derived oligosaccharides in Escherichia coli. J Bacteriol *157*, 956–957.

Boucher, J.C., Yu, H., Mudd, M.H., and Deretic, V. (1997). Mucoid Pseudomonas aeruginosa in cystic fibrosis: characterization of muc mutations in clinical isolates and analysis of clearance in a mouse model of respiratory infection. Infect Immun *65*, 3838–3846.

Boucher, R.C. (2007). Airway Surface Dehydration in Cystic Fibrosis: Pathogenesis and Therapy. Annu Rev Med *58*, 157–170.

Boyd, E.F., and Hartl, D.L. (1998). Diversifying selection governs sequence polymorphism in the major adhesin proteins fimA, papA, and sfaA of Escherichia coli. J Mol Evol *47*, 258–267.

BRINTON, C.C. (1959). Non-flagellar appendages of bacteria. Nature *183*, 782–786.

Bruen, T.C. (2005). A Simple and Robust Statistical Test for Detecting the Presence of Recombination. Genetics *172*, 2665–2681.

Bruen, T.C. (2006). PhiPack: PHI test and other tests of recombination.

Bryan, A., Roesch, P., Davis, L., Moritz, R., Pellett, S., and Welch, R.A. (2006). Regulation of type 1 fimbriae by unlinked FimB- and FimE-like recombinases in uropathogenic Escherichia coli strain CFT073. Infect Immun *74*, 1072–1083.

Charlebois, R.L., and Doolittle, W.F. (2004). Computing prokaryotic gene ubiquity: rescuing the core from extinction. Genome Research *14*, 2469–2477.

Chattopadhyay, S., Dykhuizen, D.E., and Sokurenko, E.V. (2007). ZPS: visualization of recent adaptive evolution of proteins. BMC Bioinformatics *8*, 187-192.

Chattopadhyay, S., Paul, S., Dykhuizen, D.E., and Sokurenko, E.V. (2012). Tracking recent adaptive evolution in microbial species using TimeZone. Nat Protoc *8*, 652–665.

Chattopadhyay, S., Weissman, S.J., Minin, V.N., Russo, T.A., Dykhuizen, D.E., and Sokurenko, E.V. (2009). High frequency of hotspot mutations in core genes of Escherichia coli due to short-term positive selection. Proc Natl Acad Sci USA *106*, 12412–12417.

Chen, J., Lee, S.M., and Mao, Y. (2004). Protective effect of exopolysaccharide colanic acid of Escherichia coli O157:H7 to osmotic and oxidative stress. International Journal of Food Microbiology *93*, 281–286.

Choudhury, D. (1999). X-ray Structure of the FimC-FimH Chaperone-Adhesin Complex from Uropathogenic Escherichia coli. Science *285*, 1061–1066.

Ciofu, O., Fussing, V., Bagge, N., Koch, C., and Hoiby, N. (2001). Characterization of paired mucoid/non-mucoid Pseudomonas aeruginosa isolates from Danish cystic fibrosis patients: antibiotic resistance, beta-lactamase activity and RiboPrinting. J Antimicrob Chemother *48*, 391–396.

Clarke, B.R., Pearce, R., and Roberts, I.S. (1999). Genetic organization of the Escherichia coli K10 capsule gene cluster: identification and characterization of two conserved regions in group III capsule gene clusters encoding polysaccharide transport functions. J Bacteriol *181*, 2279–2285.

Clowes, R.C., and Rowley, D. (1955). Genetic studies on small-colony variants of Escherichia coli K-12. J Gen Microbiol *13*, 461–473.

Cohen, T.S., and Prince, A. (2012). Cystic fibrosis: a mucosal immunodeficiency syndrome. Nat Med *18*, nm.2715–11.

Connell, I., Agace, W., Klemm, P., Schembri, M., Mărild, S., and Svanborg, C. (1996). Type 1 fimbrial expression enhances Escherichia coli virulence for the urinary tract. Proc Natl Acad Sci USA *93*, 9827–9832.

Cookson, A.L., Cooley, W.A., and Woodward, M.J. (2002). The role of type 1 and curli fimbriae of Shiga toxin-producing Escherichia coli in adherence to abiotic surfaces. International Journal of Medical Microbiology *292*, 195–205.

Cystic Fibrosis Foundation (2012). Cystic Fibrosis Foundation Patient Registry - 2012 Annual Data Report.

Daegelen, P., Studier, F.W., Lenski, R.E., Cure, S., and Kim, J.F. (2009). Tracing ancestors and relatives of Escherichia coli B, and the derivation of B strains REL606 and BL21(DE3). J Mol Biol *394*, 634–643.

Darling, E.A. (1894). The Bacillus Coli Communis. Boston Medical and Surgical Journal *131*, 479–482.

de Louvois, J., Blackbourn, J., Hurley, R., and Harvey, D. (1991). Infantile meningitis in England and Wales: a two year study. Archives of Disease in Childhood *66*, 603–607.

de Smit, M.H., and van Duin, J. (1990). Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. Proc Natl Acad Sci USA *87*, 7668–7672.

Debarbieux, L., Bohin, A., and Bohin, J.P. (1997). Topological analysis of the membrane-bound glucosyltransferase, MdoH, required for osmoregulated periplasmic glucan synthesis in Escherichia coli. J Bacteriol *179*, 6692–6698.

Denamur, E., Bonacorsi, S., Giraud, A., Duriez, P., Hilali, F., Amorin, C., Bingen, E., Andremont, A., Picard, B., Taddei, F., et al. (2002). High Frequency of Mutator Strains among Human Uropathogenic Escherichia coli Isolates. J Bacteriol *184*, 605–609.

DeVries, C.A., and Ohman, D.E. (1994). Mucoid-to-nonmucoid conversion in alginate-producing Pseudomonas aeruginosa often results from spontaneous mutations in algT, encoding a putative alternate sigma factor, and shows evidence for autoregulation. J Bacteriol *176*, 6677–6687.

Doggett, R.G., Harrison, G.M., And Wallis, E.S. (1964). Comparison of some properties of Pseudomonas aeruginosa isolated from infections in persons with and without cystic fibrosis. J Bacteriol *87*, 427–431.

Dorman, C.J., and Higgins, C.F. (1987). Fimbrial phase variation in Escherichia coli: dependence on integration host factor and homologies with other site-specific recombinases. J Bacteriol *169*, 3840–3843.

Dove, S.L., Smith, S.G., and Dorman, C.J. (1997). Control of Escherichia coli type 1 fimbrial gene expression in stationary phase: a negative role for RpoS. Mol Gen Genet *254*, 13–20.

Duguid, J.P., Clegg, S., and Wilson, M.I. (1979). The fimbrial and non-fimbrial haemagglutinins of Escherichia coli. J Med Microbiol *12*, 213-227.

Ebel, W., Vaughn, G.J., Peters, H.K., and Trempy, J.E. (1997). Inactivation of mdoH leads to increased expression of colanic acid capsular polysaccharide in Escherichia coli. J Bacteriol *179*, 6858–6861.

Eiff, von, C., Heilmann, C., Proctor, R.A., Woltz, C., Peters, G., and Götz, F. (1997). A site-directed Staphylococcus aureus hemB mutant is a small-colony variant which persists intracellularly. J Bacteriol *179*, 4706–4712.

Eisenstein, B.I., Sweet, D.S., Vaughn, V., and Friedman, D.I. (1987). Integration host factor is required for the DNA inversion that controls phase variation in Escherichia coli. Proc Natl Acad Sci USA *84*, 6506–6510.

El-Labany, S., Sohanpal, B.K., Lahooti, M., Akerman, R., and Blomfield, I.C. (2003). Distant cis-active sequences and sialic acid control the expression of fimB in Escherichia coli K-12. Mol

Microbiol *49*, 1109–1118.

Emori, T.G., and Gaynes, R.P. (1993). An overview of nosocomial infections, including the role of the microbiology laboratory. Clin Microbiol Rev *6*, 428–442.

Enami, M., Nakasone, N., Honma, Y., Kakinohana, S., Kudaka, J., and Iwanaga, M. (1999). Expression of type I pili is abolished in verotoxin-producing Escherichia coli O157. FEMS Microbiology Letters *179*, 467–472.

Felsenstein, J. (2005). PHYLIP (Phylogeny Inference Package) version 3.6. *Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.*

Ferrières, L., and Clarke, D.J. (2003). The RcsC sensor kinase is required for normal biofilm formation in Escherichia coli K-12 and controls the expression of a regulon in response to growth on a solid surface. Mol Microbiol *50*, 1665–1682.

Filatov, D.A. (2009). Processing and population genetic analysis of multigenic datasets with ProSeq3 software. Bioinformatics *25*, 3189–3190.

Fisher, J.T., Zhang, Y., and Engelhardt, J.F. (2011). Comparative Biology of Cystic Fibrosis Animal Models. Methods in Molecular Biology *742*, 311–334.

Fodor, A.A., Klem, E.R., Gilpin, D.F., Elborn, J.S., Boucher, R.C., Tunney, M.M., and Wolfgang, M.C. (2012). The Adult Cystic Fibrosis Airway Microbiota Is Stable over Time and Infection Type, and Highly Resilient to Antibiotic Treatment of Exacerbations. PLoS ONE *7*, e45001.

Folkesson, A., Jelsbak, L., Yang, L., Johansen, H.K., Ciofu, O., Høiby, N., and Molin, S. (2012). Adaptation of Pseudomonas aeruginosa to the cystic fibrosis airway: an evolutionary perspective. Nat Rev Micro *10*, 841–851.

Foxman, B. (1990). Recurring urinary tract infection: incidence and risk factors. Am J Public Health *80*, 331–333.

Foxman, B., Barlow, R., D'Arcy, H., Gillespie, B., and Sobel, J.D. (2000). Urinary tract infection: self-reported incidence and associated costs. Ann Epidemiol *10*, 509–515.

Foxman, B., Zhang, L., Tallman, P., Andree, B.C., Geiger, A.M., Koopman, J.S., Gillespie, B.W., Palin, K.A., Sobel, J.D., Rode, C.K., et al. (1997). Transmission of uropathogens between sex partners. J Infect Dis *175*, 989–992.

Foxman, B. (2002). Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. Am J Med *113 Suppl 1A*, 5S–13S.

Foxman, B., and Brown, P. (2003). Epidemiology of urinary tract infections: transmission and risk factors, incidence, and costs. Infect Dis Clin North Am *17*, 227–241.

Foxman, B., Manning, S.D., Tallman, P., Bauer, R., Zhang, L., Koopman, J.S., Gillespie, B.,

Sobel, J.D., and Marrs, C.F. (2002). Uropathogenic Escherichia coli are more likely than commensal E. coli to be shared between heterosexual sex partners. Am J Epidemiol *156*, 1133–1140.

Francez-Charlot, A., Laugel, B., Van Gemert, A., Dubarry, N., Wiorowski, F., Castanié-Cornet, M.-P., Gutierrez, C., and Cam, K. (2004). RcsCDB His-Asp phosphorelay system negatively regulates the flhDC operon in Escherichia coli. Mol Microbiol *49*, 823–832.

Freedman, S.B., Ali, S., Oleszczuk, M., Gouin, S., and Hartling, L. (2013). Treatment of acute gastroenteritis in children: an overview of systematic reviews of interventions commonly used in developed countries. Evid.-Based Child Health *8*, 1123–1137.

Gally, D.L., Bogan, J.A., Eisenstein, B.I., and Blomfield, I.C. (1993). Environmental regulation of the fim switch controlling type 1 fimbrial phase variation in Escherichia coli K-12: effects of temperature and media. J Bacteriol *175*, 6186–6193.

Gally, D.L., Rucker, T.J., and Blomfield, I.C. (1994). The leucine-responsive regulatory protein binds to the fim switch to control phase variation of type 1 fimbrial expression in Escherichia coli K-12. J Bacteriol *176*, 5665–5672.

Geibel, S., Procko, E., Hultgren, S.J., Baker, D., and Waksman, G. (2013). Structural and energetic basis of folded-protein transport by the FimD usher. Nature *496*, 243–246.

Gervais, F.G., Phoenix, P., and Drapeau, G.R. (1992). The rcsB gene, a positive regulator of colanic acid biosynthesis in Escherichia coli, is also an activator of ftsZ expression. J Bacteriol *174*, 3964–3971.

Giacani, L., Chattopadhyay, S., Centurion-Lara, A., Jeffrey, B.M., Le, H.T., Molini, B.J., Lukehart, S.A., Sokurenko, E.V., and Rockey, D.D. (2012). Footprint of Positive Selection in Treponema pallidum subsp. pallidum Genome Sequences Suggests Adaptive Microevolution of the Syphilis Pathogen. PLoS Negl Trop Dis *6*, e1698.

Goebel, W.F. (1963). Colanic acid. Proc Natl Acad Sci USA *49*, 464–471.

Gossert, A.D., Bettendorff, P., Puorger, C., Vetsch, M., Herrmann, T., Glockshuber, R., and Wüthrich, K. (2008). NMR structure of the Escherichia coli type 1 pilus subunit FimF and its interactions with other pilus subunits. J Mol Biol *375*, 752–763.

Gottesman, S., Trisler, P., and Torres-Cabassa, A. (1985). Regulation of capsular polysaccharide synthesis in Escherichia coli K-12: characterization of three regulatory genes. J Bacteriol *162*, 1111–1119.

Gould, L.H., Demma, L., Jones, T.F., Hurd, S., Vugia, D.J., Smith, K., Shiferaw, B., Segler, S., Palmer, A., Zansky, S., et al. (2009). Hemolytic Uremic Syndrome and Death in Persons with Escherichia coliO157:H7 Infection, Foodborne Diseases Active Surveillance Network Sites, 2000–2006. Clin Infect Dis *49*, 1480–1485.

Govan, J.R., and Deretic, V. (1996). Microbial pathogenesis in cystic fibrosis: mucoid

Pseudomonas aeruginosa and Burkholderia cepacia. Microbiol Rev *60*, 539–574.

Hagiwara, D., Sugiura, M., Oshima, T., Mori, H., Aiba, H., Yamashino, T., and Mizuno, T. (2003). Genome-Wide Analyses Revealing a Signaling Network of the RcsC-YojN-RcsB Phosphorelay System in Escherichia coli. J Bacteriol *185*, 5735–5746.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium *41*, 95–98.

Harvey, D., Holt, D.E., and Bedford, H. (1999). Bacterial meningitis in the newborn: a prospective study of mortality and morbidity. Semin. Perinatol. *23*, 218–225.

Hobman, J.L., Penn, C.W., and Pallen, M.J. (2007). Laboratory strains of Escherichia coli: model citizens or deceitful delinquents growing old disgracefully? Mol Microbiol *64*, 881–885.

Holden, N.J., Uhlin, B.E., and Gally, D.L. (2001). PapB paralogues and their effect on the phase variation of type 1 fimbriae in Escherichia coli. Mol Microbiol *42*, 319–330.

Holden, N.J., Totsika, M., Mahler, E., Roe, A.J., Catherwood, K., Lindner, K., Dobrindt, U., and Gally, D.L. (2006). Demonstration of regulatory cross-talk between P fimbriae and type 1 fimbriae in uropathogenic Escherichia coli. Microbiology (Reading, Engl) *152*, 1143–1153.

Holden, N., Blomfield, I.C., Uhlin, B.-E., Totsika, M., Kulasekara, D.H., and Gally, D.L. (2007). Comparative analysis of FimB and FimE recombinase activity. Microbiology (Reading, Engl) *153*, 4138–4149.

Holt, D.E., Halket, S., de Louvois, J., and Harvey, D. (2001). Neonatal meningitis in England and Wales: 10 years on. Archives of Disease in Childhood - Fetal and Neonatal Edition *84*, F85–F89.

Hudault, S., Guignot, J., and Servin, A.L. (2001). Escherichia coli strains colonising the gastrointestinal tract protect germfree mice against Salmonella typhimurium infection. Gut *49*, 47–55.

Hughes, A.L. (2008). Near neutrality: leading edge of the neutral theory of molecular evolution. Annals of the New York Academy of Sciences *1133*, 162–179.

Hustinx, W.N., and Verbrugh, H.A. (1994). Catherer-associated urinary tract infections: epidemiological, preventive and therapeutic considerations. International Journal of Antimicrobial Agents *4*, 117–123.

Høiby, N. (1974a). Epidemiological investigations of the respiratory tract bacteriology in patients with cystic fibrosis. Acta Pathologica Microbiologica Scandinavica Section B Microbiology and Immunology *82*, 541–550.

Høiby, N. (1974b). Pseudomonas aeruginosa infection in cystic fibrosis. Acta Pathologica Microbiologica Scandinavica Section B Microbiology and Immunology *82*, 551–558.

Iida, K., Mizunoe, Y., Wai, S.N., and Yoshida, S. (2001). Type 1 fimbriation and its phase switching in diarrheagenic Escherichia coli strains. Clin Diagn Lab Immunol *8*, 489–495.

Ikähelmo, R., Siitonen, A., Heiskanen, T., Kärkkäinen, U., Kuosmanen, P., Lipponen, P., and Mäkelä, P.H. (1996). Recurrence of Urinary Tract Infection in a Primary Care Setting: Analysis of a I-Year Follow-up of 179 Women. Clinical Infectious Diseases *22*, 91–99.

Jackson, L.A., Benson, P., Neuzil, K.M., Grandjean, M., and Marino, J.L. (2005). Burden of community-onset Escherichia coli bacteremia in seniors. J Infect Dis *191*, 1523–1529.

Jakobsen, L., Kurbasic, A., Skjøt-Rasmussen, L., Ejrnæs, K., Porsbo, L.J., Pedersen, K., Jensen, L.B., Emborg, H.-D., Agersø, Y., Olsen, K.E.P., et al. (2009). Escherichia coli Isolates from Broiler Chicken Meat, Broiler Chickens, Pork, and Pigs Share Phylogroups and Antimicrobial Resistance with Community-Dwelling Humans and Patients with Urinary Tract Infection. Foodborne Pathogens and Disease.

Jayaratne, P., Keenleyside, W.J., MacLachlan, P.R., Dodgson, C., and Whitfield, C. (1993). Characterization of rcsB and rcsC from Escherichia coli O9:K30:H12 and examination of the role of the rcs regulatory system in expression of group I capsular polysaccharides. J Bacteriol *175*, 5384–5394.

Johnson, J.R., Gajewski, A., Lesse, A.J., and Russo, T.A. (2003). Extraintestinal Pathogenic Escherichia coli as a Cause of Invasive Nonurinary Infections. Journal of Clinical Microbiology *41*, 5798–5802.

Johnson, J.R., and Russo, T.A. (2002). Extraintestinal pathogenic Escherichia coli: “the other bad E coli.” J Lab Clin Med *139*, 155–162.

Johnson, T.J., Wannemuehler, Y., Johnson, S.J., Stell, A.L., Doetkott, C., Johnson, J.R., Kim, K.S., Spanjaard, L., and Nolan, L.K. (2008). Comparison of extraintestinal pathogenic Escherichia coli strains from human and avian sources reveals a mixed subset representing potential zoonotic pathogens. Appl Env Microbiol *74*, 7043-7050.

Jones, B., Peake, K., Morris, A.J., McCowan, L.M., and Battin, M.R. (2004). Escherichia coli: a growing problem in early onset neonatal sepsis. Aust N Z J Obstet Gynaecol *44*, 558–561.

Jones, C.H., Pinkner, J.S., Nicholes, A.V., Slonim, L.N., Abraham, S.N., and Hultgren, S.J. (1993). FimC is a periplasmic PapD-like chaperone that directs assembly of type 1 pili in bacteria. Proc Natl Acad Sci USA *90*, 8397–8401.

Kahl, B., Herrmann, M., Everding, A.S., Koch, H.G., Becker, K., Harms, E., Proctor, R.A., and Peters, G. (1998). Persistent infection with small colony variant strains of Staphylococcus aureus in patients with cystic fibrosis. J Infect Dis *177*, 1023–1029.

Kahl, B.C., Duebbers, A., Lubritz, G., Haeberle, J., Koch, H.G., Ritzerfeld, B., Reilly, M., Harms, E., Proctor, R.A., Herrmann, M., et al. (2003). Population dynamics of persistent Staphylococcus aureus isolated from the airways of cystic fibrosis patients during a 6-year prospective study. Journal of Clinical Microbiology *41*, 4424–4427.

Kaper, J.B., Nataro, J.P., and Mobley, H.L. (2004). Pathogenic Escherichia coli. Nat Rev Micro *2*, 123-140.

Keenleyside, W.J., Bronner, D., Jann, K., Jann, B., and Whitfield, C. (1993). Coexpression of colanic acid and serotype-specific capsular polysaccharides in Escherichia coli strains with group II K antigens. J Bacteriol *175*, 6725–6730.

Keiser, N.W., and Engelhardt, J.F. (2011). New animal models of cystic fibrosis. Current Opinion in Pulmonary Medicine 1.

Kimura, M. (1979). Model of effectively neutral mutations in which selective constraint is incorporated. Proc Natl Acad Sci USA *76*, 3440–3444.

Klemm, P. (1984). The fimA gene encoding the type-1 fimbrial subunit of Escherichia coli. Nucleotide sequence and primary structure of the protein. Eur J Biochem *143*, 395–399.

Klemm, P. (1986). Two regulatory fim genes, fimB and fimE, control the phase variation of type 1 fimbriae in Escherichia coli. Embo J *5*, 1389–1393.

Klemm, P. (1992). FimC, a chaperone-like periplasmic protein of Escherichia coli involved in biogenesis of type 1 fimbriae. Res Microbiol *143*, 831–838.

Klemm, P., and Christiansen, G. (1987). Three fim genes required for the regulation of length and mediation of adhesion of Escherichia coli type 1 fimbriae. Mol Gen Genet *208*, 439–445.

Korea, C.-G., Badouraly, R., Prevost, M.-C., Ghigo, J.-M., and Beloin, C. (2010). Escherichia coli K-12 possesses multiple cryptic but functional chaperone-usher fimbriae with distinct surface specificities. Environmental Microbiology *12*, 1957–1977.

Korhonen, T.K., Väisänen, V., Saxén, H., Hultberg, H., and Svenson, S.B. (1982). P-antigen-recognizing fimbriae from human uropathogenic Escherichia coli strains. Infect Immun *37*, 286–291.

Kosek, M., Bern, C., and Guerrant, R.L. (2003). The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. Bull World Health Organ *81*, 197–204.

Krogfelt, K.A., and Klemm, P. (1988). Investigation of minor components of Escherichia coli type 1 fimbriae: protein chemical and immunological aspects. Microb Pathog *4*, 231–238.

Krogfelt, K.A., Bergmans, H., and Klemm, P. (1990). Direct evidence that the FimH protein is the mannose-specific adhesin of Escherichia coli type 1 fimbriae. Infect Immun *58*, 1995–1998.

Kudla, G., Murray, A.W., Tollervey, D., and Plotkin, J.B. (2009). Coding-sequence determinants of gene expression in Escherichia coli. Science *324*, 255–258.

Labat, F., Pradillon, O., Garry, L., Peuchmaur, M., Fantin, B., and Denamur, E. (2005). Mutator phenotype confers advantage in Escherichia coli chronic urinary tract infection pathogenesis.

FEMS Immunol Med Microbiol *44*, 317–321.

Lacroix, J.-M., Loubens, I., Tempête, M., Menichi, B., and Bohin, J.P. (1991). The mdoA locus of Escherichia coli consists of an operon under osmotic control. Mol Microbiol *5*, 1745–1753.

Lacroix, J.-M., Tempête, M., Menichi, B., and Bohin, J.P. (1989). Molecular cloning and expression of a locus (mdoA) implicated in the biosynthesis of membrane-derived oligosaccharides in Escherichia coli. Mol Microbiol *3*, 1173–1182.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947–2948.

Lasaro, M.A., Salinger, N., Zhang, J., Wang, Y., Zhong, Z., Goulian, M., and Zhu, J. (2009). F1C fimbriae play an important role in biofilm formation and intestinal colonization by the Escherichia coli commensal strain Nissle 1917. Appl Environ Microbiol *75*, 246–251.

Laubacher, M.E., and Ades, S.E. (2008). The Rcs Phosphorelay Is a Cell Envelope Stress Response Activated by Peptidoglycan Stress and Contributes to Intrinsic Antibiotic Resistance. J Bacteriol *190*, 2065–2074.

Le Trong, I., Aprikian, P., Kidd, B.A., Forero-Shelton, M., Tchesnokova, V., Rajagopal, P., Rodriguez, V., Interlandi, G., Klevit, R., Vogel, V., et al. (2010). Structural basis for mechanical force regulation of the adhesin FimH via finger trap-like beta sheet twisting. Cell *141*, 645–655.

Lee, D.-G., Jeon, S.H., Lee, C.-H., Lee, S.-J., Kim, J.I., and Chang, S.-G. (2009). Acute pyelonephritis: clinical characteristics and the role of the surgical treatment. J Korean Med Sci *24*, 296–301.

Lehti, T.A., Bauchart, P., Heikkinen, J., Hacker, J., Korhonen, T.K., Dobrindt, U., and Westerlund-Wikstrom, B. (2010). Mat fimbriae promote biofilm formation by meningitis-associated Escherichia coli. Microbiology *156*, 2408–2417.

Lehti, T.A., Heikkinen, J., Korhonen, T.K., and Westerlund-Wikstrom, B. (2012). The Response Regulator RcsB Activates Expression of Mat Fimbriae in Meningitic Escherichia coli. J Bacteriol *194*, 3475–3485.

Leid, J.G., Willson, C.J., Shirtliff, M.E., Hassett, D.J., Parsek, M.R., and Jeffers, A.K. (2005). The exopolysaccharide alginate protects Pseudomonas aeruginosa biofilm bacteria from IFN-gamma-mediated macrophage killing. J Immunol *175*, 7512–7518.

Leopold, S.R., Magrini, V., Holt, N.J., Shaikh, N., Mardis, E.R., Cagno, J., Ogura, Y., Iguchi, A., Hayashi, T., Mellmann, A., et al. (2009). A precise reconstruction of the emergence and constrained radiations of Escherichia coli O157 portrayed by backbone concatenomic analysis. Proc Natl Acad Sci USA *106*, 8713–8718.

Li, B., Koch, W.H., and Cebula, T.A. (1997). Detection and characterization of the fimA gene of Escherichia coli O157:H7. Mol Cell Probes *11*, 397–406.

Li, Z., and Engelhardt, J.F. (2003). Progress toward generating a ferret model of cystic fibrosis by somatic cell nuclear transfer. Reprod Biol Endocrinol *1*, 83.

Lim, J.K., Gunther, N.W., Zhao, H., Johnson, D.E., Keay, S.K., and Mobley, H.L. (1998). In vivo phase variation of Escherichia coli type 1 fimbrial genes in women with urinary tract infection. Infect Immun *66*, 3303-3310.

Lukjancenko, O., Wassenaar, T.M., and Ussery, D.W. (2010). Comparison of 61 Sequenced Escherichia coli Genomes. Microb Ecol *60*, 708–720.

Macone, A.B., Pier, G.B., Pennington, J.E., Matthews, W.J., and Goldmann, D.A. (1981). Mucoid Escherichia coli in cystic fibrosis. N Engl J Med *304*, 1445–1449.

Majdalani, N., and Gottesman, S. (2005). The Rcs phosphorelay: a complex signal transduction system. Annu Rev Microbiol *59*, 379–405.

Maltby, R., Leatham-Jensen, M.P., Gibson, T., Cohen, P.S., and Conway, T. (2013). Nutritional Basis for Colonization Resistance by Human Commensal Escherichia coli Strains HS and Nissle 1917 against E. coli O157:H7 in the Mouse Intestine. PLoS ONE *8*, e53957.

Markovitz, A. (1964). REGULATORY MECHANISMS FOR SYNTHESIS OF CAPSULAR POLYSACCHARIDE IN MUCOID MUTANTS OF ESCHERICHIA COLI K12. Proc Natl Acad Sci USA *51*, 239–246.

Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., and Lefeuvre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics *26*, 2462–2463.

Martin, D.P., Williamson, C., and Posada, D. (2005). RDP2: recombination detection and analysis from sequence alignments. Bioinformatics *21*, 260–262.

Martin, D.W., Schurr, M.J., Mudd, M.H., Govan, J.R., Holloway, B.W., and Deretic, V. (1993). Mechanism of conversion to mucoidy in Pseudomonas aeruginosa infecting cystic fibrosis patients. Proc Natl Acad Sci USA *90*, 8377.

Matsui, H., Grubb, B.R., Tarran, R., Randell, S.H., Gatzy, J.T., Davis, C.W., and Boucher, R.C. (1998). Evidence for periciliary liquid layer depletion, not abnormal ion composition, in the pathogenesis of cystic fibrosis airways disease. Cell *95*, 1005–1015.

Maurelli, A.T., Fernández, R.E., Bloch, C.A., Rode, C.K., and Fasano, A. (1998). "Black holes" and bacterial pathogenicity: a large genomic deletion that enhances the virulence of Shigella spp. and enteroinvasive Escherichia coli. Proc Natl Acad Sci USA *95*, 3943–3948.

Maurelli, A.T. (2007). Black holes, antivirulence genes, and gene inactivation in the evolution of bacterial pathogens. FEMS Microbiology Letters *267*, 1–8.

May, M. (2005). Early onset neonatal meningitis in Australia and New Zealand, 1992-2002. Archives of Disease in Childhood - Fetal and Neonatal Edition *90*, F324–f327.

Mayor-Lynn, K., González-Quintero, V.H., O'Sullivan, M.J., Hartstein, A.I., Roger, S., and Tamayo, M. (2005). Comparison of early-onset neonatal sepsis caused by Escherichia coli and group B Streptococcus. American Journal of Obstetrics and Gynecology *192*, 1437–1439.

McClain, M.S., Blomfield, I.C., and Eisenstein, B.I. (1991). Roles of fimB and fimE in site-specific DNA inversion associated with phase variation of type 1 fimbriae in Escherichia coli. J Bacteriol *173*, 5308–5314.

Meltz Steinberg, K., and Levin, B.R. (2007). Grazing protozoa and the evolution of the Escherichia coli O157:H7 Shiga toxin-encoding prophage. Proceedings of the Royal Society B: Biological Sciences *274*, 1921–1929.

Miajlovic, H., Cooke, N.M., Moran, G.P., Rogers, T.R., and Smith, S.G. (2013). The response of Extra-intestinal Pathogenic Escherichia coli to human serum reveals a protective role for Rcs-regulated exopolysaccharide colanic acid. Infect Immun.

Mulder, C.J., van Alphen, L., and Zanen, H.C. (1984). Neonatal meningitis caused by Escherichia coli in The Netherlands. J Infect Dis *150*, 935–940.

Nakano, M., Yamamoto, S., Terai, A., Ogawa, O., Makino, S.I., Hayashi, H., Nair, G.B., and Kurazono, H. (2001). Structural and sequence diversity of the pathogenicity island of uropathogenic Escherichia coli which encodes the USP protein. FEMS Microbiology Letters *205*, 71–76.

Nataro, J.P., and Kaper, J.B. (1998). Diarrheagenic Escherichia coli. Clin Microbiol Rev *11*, 142–201.

Navasa, N., Rodríguez-Aparicio, L., Ferrero, M.Á., Monteagudo Mera, A., and Martínez-Blanco, H. (2013). Polysialic and colanic acids metabolism in Escherichia coli K92 is regulated by RcsA and RcsB. Biosci. Rep. *33*, 405–415.

Nicolle, L.E. (2013). Urinary Tract Infection. Critical Care Clinics *29*, 699–715.

Nilsson, L.M., Thomas, W.E., Sokurenko, E.V., and Vogel, V. (2006). Elevated Shear Stress Protects Escherichia coli Cells Adhering to Surfaces via Catch Bonds from Detachment by Soluble Inhibitors. Appl Environ Microbiol *72*, 3005–3010.

Nowrouzian, F.L., Adlerberth, I., and Wold, A.E. (2006). Enhanced persistence in the colonic microbiota of Escherichia coli strains belonging to phylogenetic group B2: role of virulence factors and adherence to colonic cells. Microbes and Infection *8*, 834–840.

O'gara, J.P., and Dorman, C.J. (2000). Effects of local transcription and H-NS on inversion of the fim switch of Escherichia coli. Mol Microbiol *36*, 457–466.

Obadia, B., Lacour, S., Doublet, P., Baubichon-Cortay, H., Cozzone, A.J., and Grangeasse, C. (2007). Influence of Tyrosine-Kinase Wzc Activity on Colanic Acid Production in Escherichia coli K12 Cells. J Mol Biol *367*, 42–53.

Ochman, H., and Selander, R.K. (1984). Standard reference strains of Escherichia coli from natural populations. J Bacteriol *157*, 690–693.

Oh, S., Buddenborg, S., Yoder-Himes, D.R., Tiedje, J.M., and Konstantinidis, K.T. (2012). Genomic Diversity of Escherichia Isolates from Diverse Habitats. PLoS ONE *7*, e47005.

Olsen, P.B., Schembri, M.A., Gally, D.L., and Klemm, P. (1998). Differential temperature modulation by H-NS of the fimB and fimE recombinase genes which control the orientation of the type 1 fimbrial phase switch. FEMS Microbiology Letters *162*, 17–23.

Olsén, A., Jonsson, A., and Normark, S. (1989). Fibronectin binding mediated by a novel class of surface organelles on Escherichia coli. Nature *338*, 652–655.

Ophir, T., and Gutnick, D.L. (1994). A role for exopolysaccharides in the protection of microorganisms from desiccation. Appl Environ Microbiol *60*, 740–745.

Ostblom, A., Adlerberth, I., Wold, A.E., and Nowrouzian, F.L. (2011). Pathogenicity island markers, virulence determinants malX and usp, and the capacity of Escherichia coli to persist in infants' commensal microbiotas. Appl Environ Microbiol *77*, 2303–2308.

O'Sullivan, B.P., and Freedman, S.D. (2009). Cystic fibrosis. The Lancet *373*, 1891–1904.

Page, R.D. (1996). TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci *12*, 357–358.

Parkkinen, J., Finne, J., Achtman, M., Väisänen, V., and Korhonen, T.K. (1983). Escherichia coli strains binding neuraminyl alpha 2-3 galactosides. Biochemical and Biophysical Research Communications *111*, 456–461.

Patel, J., Sharma, M., and Ravishakar, S. (2011). Effect of curli expression and hydrophobicity of Escherichia coli O157:H7 on attachment to fresh produce surfaces. Journal of Applied Microbiology *110*, 737–745.

Peek, A.S., Souza, V., Eguiarte, L.E., and Gaut, B.S. (2001). The interaction of protein structure, selection, and recombination on the evolution of the type-1 fimbrial major subunit (fimA) from Escherichia coli. J Mol Evol *52*, 193–204.

Phan, G., Remaut, H., Wang, T., Allen, W.J., Pirker, K.F., Lebedev, A., Henderson, N.S., Geibel, S., Volkan, E., Yan, J., et al. (2011). Crystal structure of the FimD usher bound to its cognate FimC-FimH substrate. Nature *474*, 49–53.

Phan, M.-D., Peters, K.M., Sarkar, S., Lukowski, S.W., Allsopp, L.P., Moriel, D.G., Achard, M.E.S., Totsika, M., Marshall, V.M., Upton, M., et al. (2013). The Serum Resistome of a Globally Disseminated Multidrug Resistant Uropathogenic Escherichia coli Clone. PLoS Genet *9*, e1003834.

Pouttu, R., Westerlund-Wikstrom, B., Lang, H., Alsti, K., Virkola, R., Saarela, U., Siitonen, A., Kalkkinen, N., and Korhonen, T.K. (2001). matB, a Common Fimbrillin Gene of Escherichia

coli, Expressed in a Genetically Conserved, Virulent Clonal Group. J Bacteriol *183*, 4727–4736.

Prigent-Combaret, C., Prensier, G., Le Thi, T.T., Vidal, O., Lejeune, P., and Dorel, C. (2000). Developmental pathway for biofilm formation in curli-producing Escherichia coli strains: role of flagella, curli and colanic acid. Environmental Microbiology *2*, 450–464.

Proctor, R.A., Kahl, B., Eiff, von, C., Vaudaux, P.E., Lew, D.P., and Peters, G. (1998). Staphylococcal small colony variants have novel mechanisms for antibiotic resistance. Clin Infect Dis *27 Suppl 1*, S68–S74.

Prosseda, G., Di Martino, M.L., Campilongo, R., Fioravanti, R., Micheli, G., Casalino, M., and Colonna, B. (2012). Shedding of genes that interfere with the pathogenic lifestyle: the Shigella model. Res Microbiol 1–8.

Puorger, C., Eidam, O., Capitani, G., Erilov, D., Grütter, M.G., and Glockshuber, R. (2008). Infinite kinetic stability against dissociation of supramolecular protein complexes through donor strand complementation. Structure *16*, 631–642.

Rangel, J.M., Sparling, P.H., Crowe, C., Griffin, P.M., and Swerdlow, D.L. (2005). Epidemiology of Escherichia coli O157: H7 outbreaks, United States, 1982–2002. Emerg. Infect. Dis. *11*, 603.

Reisner, A., Krogfelt, K.A., Klein, B.M., Zechner, E.L., and Molin, S. (2006). In vitro biofilm formation of commensal and pathogenic Escherichia coli strains: impact of environmental and genetic factors. J Bacteriol *188*, 3572–3581.

Remaut, H., Rose, R.J., Hannan, T.J., Hultgren, S.J., Radford, S.E., Ashcroft, A.E., and Waksman, G. (2006). Donor-Strand Exchange in Chaperone-Assisted Pilus Assembly Proceeds through a Concerted β Strand Displacement Mechanism. Mol Cell *22*, 831–842.

Rendón, M.A., Saldaña, Z., Erdem, A.L., Monteiro-Neto, V., Vázquez, A., Kaper, J.B., Puente, J.L., and Giron, J.A. (2007). Commensal and pathogenic Escherichia coli use a common pilus adherence factor for epithelial cell colonization. Proc Natl Acad Sci USA *104*, 10637–10642.

Roberts, F.J., Geere, I.W., and Coldman, A. (1991). A three-year study of positive blood cultures, with emphasis on prognosis. Rev. Infect. Dis. *13*, 34–46.

Rodriguez, V.B., Kidd, B.A., Interlandi, G., Tchesnokova, V., Sokurenko, E.V., and Thomas, W.E. (2013). Allosteric coupling in the bacterial adhesive protein FimH. Journal of Biological Chemistry.

Roe, A.J., Currie, C., Smith, D.G., and Gally, D.L. (2001). Analysis of type 1 fimbriae expression in verotoxigenic Escherichia coli: a comparison between serotypes O157 and O26. Microbiology (Reading, Engl) *147*, 145–152.

Roesch, P.L., and Blomfield, I.C. (1998). Leucine alters the interaction of the leucine-responsive regulatory protein (Lrp) with the fim switch to stimulate site-specific recombination in Escherichia coli. Mol Microbiol *27*, 751–761.

Rogers, C.S., Stoltz, D.A., Meyerholz, D.K., Ostedgaard, L.S., Rokhlina, T., Taft, P.J., Rogan, M.P., Pezzulo, A.A., Karp, P.H., Itani, O.A., et al. (2008). Disruption of the CFTR Gene Produces a Model of Cystic Fibrosis in Newborn Pigs. Science *321*, 1837–1841.

Roggenkamp, A., Sing, A., Hornef, M., Brunner, U., Autenrieth, I.B., and Heesemann, J. (1998). Chronic prosthetic hip infection caused by a small-colony variant of Escherichia coli. Journal of Clinical Microbiology *36*, 2530–2534.

Rolland, K., Lambert-Zechovsky, N., Picard, B., and Denamur, E. (1998) Shigella and enteroinvasive Escherichia coli strains are derived from distinct ancestral species of E. coli. Microbiology *144*, 2667-2672.

Ronald, L.S., Yakovenko, O., Yazvenko, N., Chattopadhyay, S., Aprikian, P., Thomas, W.E., and Sokurenko, E.V. (2008). Adaptive mutations in the signal peptide of the type 1 fimbrial adhesin of uropathogenic Escherichia coli. Proc Natl Acad Sci USA *105*, 10937–10942.

Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X., and Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics *19*, 2496–2497.

Russell, P.W., and Orndorff, P.E. (1992). Lesions in two Escherichia coli type 1 pilus genes alter pilus number and length without affecting receptor binding. J Bacteriol *174*, 5923–5935.

Russo, T.A., and James, J.R. (2003). Medical and economic impact of extraintestinal infections due to Escherichia coli: focus on an increasingly important epidemic problem. Micorbes Infect *5*: 449-456.

Russo, T.A., and Singh, G. (1993). An extraintestinal, pathogenic isolate of Escherichia coli (O4/K54/H5) can produce a group 1 capsule which is divergently regulated from its constitutively produced group 2, K54 capsular polysaccharide. J Bacteriol *175*, 7617–7623.

Ryu, J.H., Kim, H., Frank, J.F., and Beuchat, L.R. (2004). Attachment and biofilm formation on stainless steel by Escherichia coli O157:H7 as affected by curli production. Lett Appl Microbiol *39*, 359–362.

Sauer, F.G., Pinkner, J.S., Waksman, G., and Hultgren, S.J. (2002). Chaperone priming of pilus subunits facilitates a topological transition that drives fiber formation. Cell *111*, 543–551.

Schembri, M.A., Kjaergaard, K., Sokurenko, E.V., and Klemm, P. (2001). Molecular characterization of the Escherichia coli FimH adhesin. J Infect Dis *183 Suppl 1*, S28–S31.

Schierack, P., Walk, N., Ewers, C., Wilking, H., Steinrück, H., Filter, M., and Wieler, L.H. (2008). ExPEC-typical virulence-associated genes correlate with successful colonization by intestinal E. coli in a small piglet group. Environmental Microbiology *10*, 1742–1751.

Schroeder, G.N., and Hilbi, H. (2008). Molecular pathogenesis of *Shigella* spp.: Controlling host cell signaling, invasion, and death by Type 3 Secretion. Clin Microbiol Rev *21*, 134-156.

Schwan, W.R., Shibata, S., Aizawa, S.-I., and Wolfe, A.J. (2007). The two-component response

regulator RcsB regulates type 1 piliation in Escherichia coli. J Bacteriol *189*, 7159–7163.

Sears, H.J., Brownlee, I., and Uchiyama, J.K. (1950). Persistence of individual strains of Escherichia coli in the intestinal tract of man. J Bacteriol *59*, 293–301.

Sears, H.J., Janes, H., Saloum, R., Brownlee, I., and Lamoreaux, L.F. (1956). Persistence of individual strains of Escherichia coli in man and dog under varying conditions. J Bacteriol *71*, 370.

Segata, N., and Huttenhower, C. (2011). Toward an Efficient Method of Identifying Core Genes for Evolutionary and Functional Microbial Phylogenies. PLoS ONE *6*, e24704.

Seo, S.W., Yang, J., and Jung, G.Y. (2009). Quantitative correlation between mRNA secondary structure around the region downstream of the initiation codon and translational efficiency in Escherichia coli. Biotechnol. Bioeng. *104*, 611–616.

Shaikh, N., Holt, N.J., Johnson, J.R., and Tarr, P.I. (2007). Fim operon variation in the emergence of Enterohemorrhagic Escherichia coli: an evolutionary and functional analysis. FEMS Microbiology Letters *273*, 58–63.

Sibley, C.D., and Surette, M.G. (2011). The polymicrobial nature of airway infections in cystic fibrosis: Cangene Gold Medal Lecture. Can J Microbiol *57*, 69–77.

Skorupski, K., Tomaschewski, J., Rüger, W., and Simon, L.D. (1988). A bacteriophage T4 gene which functions to inhibit Escherichia coli Lon protease. J Bacteriol *170*, 3016–3024.

Slack, M.P.E., and Nichols, W.W. (1981). The penetration of antibiotics through sodium alginate and through the exopolysaccharide of a mucoid strain of Pseudomonas aeruginosa. The Lancet *318*, 502–503.

Sledjeski, D.D., and Gottesman, S. (1996). Osmotic shock induction of capsule synthesis in Escherichia coli K-12. J Bacteriol *178*, 1204–1206.

Smith, J.L., Fratamico, P.M., and Gunther, N.W. (2007). Extraintestinal pathogenic Escherichia coli. Foodborne Pathogens and Disease *4*, 134–163.

Snyder, J.A., Lloyd, A.L., Lockatell, C.V., Johnson, D.E., and Mobley, H.L.T. (2006). Role of phase variation of type 1 fimbriae in a uropathogenic Escherichia coli cystitis isolate during urinary tract infection. Infect Immun *74*, 1387–1393.

Sokurenko, E.V., Chesnokova, V., Doyle, R.J., and Hasty, D.L. (1997). Diversity of the Escherichia coli type 1 fimbrial lectin. Differential binding to mannosides and uroepithelial cells. J Biol Chem *272*, 17880–17886.

Sokurenko, E.V., Chesnokova, V., Dykhuizen, D.E., Ofek, I., Wu, X.R., Krogfelt, K.A., Struve, C., Schembri, M.A., and Hasty, D.L. (1998). Pathogenic adaptation of Escherichia coli by natural variation of the FimH adhesin. Proc Natl Acad Sci USA *95*, 8922–8926.

Sokurenko, E.V., Courtney, H.S., Maslow, J., Siitonen, A., and Hasty, D.L. (1995). Quantitative differences in adhesiveness of type 1 fimbriated Escherichia coli due to structural differences in fimH genes. J Bacteriol *177*, 3680–3686.

Sokurenko, E.V., Courtney, H.S., Ohman, D.E., Klemm, P., and Hasty, D.L. (1994). FimH family of type 1 fimbrial adhesins: functional heterogeneity due to minor sequence variations among fimH genes. J Bacteriol *176*, 748–755.

Sokurenko, E.V., Feldgarden, M., Trintchina, E., Weissman, S.J., Avagyan, S., Chattopadhyay, S., Johnson, J.R., and Dykhuizen, D.E. (2004). Selection footprint in the FimH adhesin shows pathoadaptive niche differentiation in Escherichia coli. Molecular Biology and Evolution *21*, 1373–1383.

Stamm, W.E. (1991). Catheter-associated urinary tract infections: epidemiology, pathogenesis, and prevention. Am J Med *91*, 65S–71S.

Stamm, W.E., Hooton, T.M., Johnson, J.R., Johnson, C., Stapleton, A., Roberts, P.L., Moseley, S.L., and Fihn, S.D. (1989). Urinary tract infections: from pathogenesis to treatment. J Infect Dis *159*, 400–406.

Steenbergen, S.M., Wrona, T.J., and Vimr, E.R. (1992). Functional analysis of the sialyltransferase complexes in Escherichia coli K1 and K92. J Bacteriol *174*, 1099–1108.

Stentebjerg-Olesen, B., Chakraborty, T., and Klemm, P. (2000). FimE-catalyzed off-to-on inversion of the type 1 fimbrial phase switch and insertion sequence recruitment in an Escherichia coli K-12 fimB strain. FEMS Microbiology Letters *182*, 319–325.

Stoll, B.J. (1997). The global impact of neonatal infection. Clin Perinatol *24*, 1–21.

Stoll, B.J., Hansen, N.I., Higgins, R.D., Fanaroff, A.A., Duara, S., Goldberg, R., Laptook, A., Walsh, M., Oh, W., and Hale, E. (2005). Very Low Birth Weight Preterm Infants With Early Onset Neonatal Sepsis. Pediatr Infect Dis J *24*, 635–639.

Stout, V., and Gottesman, S. (1990). RcsB and RcsC: a two-component regulator of capsule synthesis in Escherichia coli. J Bacteriol *172*, 659–669.

Sun, X., Sui, H., Fisher, J.T., Yan, Z., Liu, X., Cho, H.-J., Joo, N.S., Zhang, Y., Zhou, W., Yi, Y., et al. (2010). Disease phenotype of a ferret CFTR-knockout model of cystic fibrosis. J Clin Invest *120*, 3149–3160.

Swofford, D.L. (1991). PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1 Computer program distributed by the Illinois Natural History Survey, Champaign, Illinois.

Tamaki, S., Matsuzawa, H., and Matsuhashi, M. (1980). Cluster of mrdA and mrdB genes responsible for the rod shape and mecillinam sensitivity of Escherichia coli. J Bacteriol *141*, 52–57.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5:

Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Molecular Biology and Evolution *28*, 2731–2739.

Tchesnokova, V., Aprikian, P., Yakovenko, O., Larock, C., Kidd, B., Vogel, V., Thomas, W., and Sokurenko, E. (2008). Integrin-like allosteric properties of the catch bond-forming FimH adhesin of Escherichia coli. J Biol Chem *283*, 7823–7833.

Tenaillon, O., Skurnik, D., Picard, B., and Denamur, E. (2010). The population genetics of commensal Escherichia coli. Nat Rev Micro *8*, 207–217.

Thomas, W.E., Nilsson, L.M., Forero, M., Sokurenko, E.V., and Vogel, V. (2004). Shear-dependent "stick-and-roll" adhesion of type 1 fimbriated Escherichia coli. Mol Microbiol *53*, 1545–1557.

Thomas, W.E., Trintchina, E., Forero, M., Vogel, V., and Sokurenko, E.V. (2002). Bacterial adhesion to target cells enhanced by shear force. Cell *109*, 913–923.

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., Bouvet, O., et al. (2009). Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. PLoS Genet *5*, e1000344.

Uchiyama, I. (2008). Multiple genome alignment for identifying the core structure among moderately related microbial genomes. BMC Genomics *9*, 515.

Unhanand, M., Mustafa, M.M., McCracken, G.H., and Nelson, J.D. (1993). Gram-negative enteric bacillary meningitis: a twenty-one-year experience. The Journal of Pediatrics *122*, 15–21.

Valenski, M.L., Harris, S.L., Spears, P.A., Horton, J.R., and Orndorff, P.E. (2003). The Product of the fimI gene is necessary for Escherichia coli type 1 pilus biosynthesis. J Bacteriol *185*, 5007–5011.

Versalovic, J., Koeuth, T., and Lupski, J.R. (1991). Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. Nucleic Acids Res *19*, 6823–6831.

Wagenlehner, F.M.E., Pilatz, A., and Weidner, W. (2011). International Journal of Antimicrobial Agents. International Journal of Antimicrobial Agents *38*, 51–57.

Warren, J.W. (2001). Catheter-associated urinary tract infections. International Journal of Antimicrobial Agents *17*, 299–303.

Weissman, S.J., Johnson, J.R., Tchesnokova, V., Billig, M., Dykhuizen, D., Riddel, K., Rogers, P., Qin, X., Butler-Wu, S., Cookson, B.T., et al. (2012). High-Resolution Two-Locus Clonal Typing of Extraintestinal Pathogenic Escherichia coli. Appl Environ Microbiol.

Weissman, S.J., Beskhlebnaya, V., Chesnokova, V., Chattopadhyay, S., Stamm, W.E., Hooton, T.M., and Sokurenko, E.V. (2007). Differential stability and trade-off effects of pathoadaptive mutations in the Escherichia coli FimH adhesin. Infect Immun *75*, 3548–3555.

Weissman, S.J., Chattopadhyay, S., Aprikian, P., Obata-Yasuoka, M., Yarova-Yarovaya, Y., Stapleton, A., Ba-Thein, W., Dykhuizen, D., Johnson, J.R., and Sokurenko, E.V. (2006). Clonal analysis reveals high rate of structural mutations in fimbrial adhesins of extraintestinal pathogenic Escherichia coli. Mol Microbiol *59*, 975–988.

Whitfield, C. (2006). Biosynthesis and assembly of capsular polysaccharides in Escherichia coli. Annu. Rev. Biochem. *75*, 39–68.

Wirth, T. UCC MLST Information page. Mlst.Ucc.Ie.

Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L.H., Karch, H., Reeves, P.R., Maiden, M.C.J., Ochman, H., et al. (2006). Sex and virulence in Escherichia coli: an evolutionary perspective. Mol Microbiol *60*, 1136–1151.

Wold, A.E., Caugant, D.A., Lidin-Janson, G., de Man, P., and Svanborg, C. (1992). Resident colonic Escherichia coli strains frequently display uropathogenic characteristics. J Infect Dis *165*, 46–52.

Wu, W., Badrane, H., Arora, S., Baker, H.V., and Jin, S. (2004). MucA-Mediated Coordination of Type III Secretion and Alginate Synthesis in Pseudomonas aeruginosa. J Bacteriol *186*, 7575–7585.

Xia, Y., Gally, D., Forsman-Semb, K., and Uhlin, B.E. (2000). Regulatory cross-talk between adhesin operons in Escherichia coli: inhibition of type 1 fimbriae expression by the PapB protein. Embo J *19*, 1450–1457.

Xicohtencatl-Cortes, J., Monteiro-Neto, V., Saldaña, Z., Ledesma, M.A., Puente, J.L., and Girón, J.A. (2009). The type 4 pili of enterohemorrhagic Escherichia coli O157:H7 are multipurpose structures with pathogenic attributes. J Bacteriol *191*, 411–421.

Xie, Y., Yao, Y., Kolisnychenko, V., Teng, C.-H., and Kim, K.S. (2006). HbiF regulates type 1 fimbriation independently of FimB and FimE. Infect Immun *74*, 4039–4047.

Yakovenko, O., SHARMA, S., Forero, M., Tchesnokova, V., Aprikian, P., Kidd, B., Mach, A., Vogel, V., Sokurenko, E., and Thomas, W.E. (2008). FimH Forms Catch Bonds That Are Enhanced by Mechanical Force Due to Allosteric Regulation. Journal of Biological Chemistry *283*, 11596–11605.

Yamamoto, S., Tsukamoto, T., Terai, A., Kurazono, H., Takeda, Y., and Yoshida, O. (1995). Distribution of virulence factors in Escherichia coli isolated from urine of cystitis patients. Microbiol Immunol *39*, 401-404.

Yeh, J.Y., and Chen, J. (2004). Production of slime polysaccharide by EHEC and STEC as well as the influence of culture conditions on slime production in Escherichia coli O157:H7. Lett Appl Microbiol *38*, 488–492.

Zavialov, A.V., Tischenko, V.M., Fooks, L.J., Brandsdal, B.O., Aqvist, J., Zav'yalov, V.P., Macintyre, S., and Knight, S.D. (2005). Resolving the energy paradox of chaperone/usher-

mediated fibre assembly. Biochem J *389*, 685–694.

Zhang, W. (2006). Probing genomic diversity and evolution of Escherichia coli O157 by single nucleotide polymorphisms. Genome Research *16*, 757–767.

Zhou, Z., Li, X., Liu, B., Beutin, L., Xu, J., Ren, Y., Feng, L., Lan, R., Reeves, P.R., and Wang, L. (2010). Derivation of Escherichia coli O157:H7 from its O55:H7 precursor. PLoS ONE *5*, e8700.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res *31*, 3406–3415.