PHYLOGENETIC AND GENOMIC CHARACTERIZATION OF PORCINE ENTEROTOXIGENIC *ESCHERICHIA COLI*

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

SARA MARIA SHEPARD

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

DR. TIMOTHY JOHNSON, ADVISOR

August 2011

Acknowledgements

I extend many thanks to my advisor, Dr. Timothy Johnson, for his guidance, mentorship, and support throughout this project. I also thank my committee members, Dr. Richard Isaacson and Dr. Michael Murtaugh, for their suggestions and contributions.

Thank you to the many wonderful people I have worked with in Dr. Johnson's lab. I especially appreciate assistance from Jessica Danzeisen and Kevin Lang on this project.

It has been a pleasure to learn alongside the other students and faculty in the Comparative and Molecular Biosciences and the Veterinary Medicine graduate programs at the University of Minnesota.

I would also like to acknowledge the National Pork Board and the U.S. Department of Agriculture for funding my work on this project.

Abstract

Porcine enterotoxigenic *Escherichia coli* (ETEC) is a significant pathogen of young pigs, causing considerable morbidity and mortality. While the known plasmid-encoded virulence factors related to porcine ETEC infection have been well-studied, the chromosomal background of porcine ETEC has been largely understudied. Since chromosomal backgrounds and any chromosome-encoded virulence factors may directly or indirectly influence ETEC pathogenesis, they deserve attention.

In this study, we utilized the first completed genome sequences of porcine ETEC to better understand porcine ETEC chromosomal content. We first examined the porcine ETEC chromosome by performing multilocus sequence analysis on 80 different porcine ETEC isolates implicated in neonatal and post-weaning diarrhea. We found that the porcine ETEC examined clustered into several specific lineages, suggesting the acquisition of porcine ETEC virulence plasmids into different *E. coli* chromosomal lineages on multiple occasions. These results also suggest that only certain chromosomal backgrounds support successful ETEC-related plasmid carriage. Patterns in resistance and virulence plasmid carriage were less clear, with plasmids of interest distributed widely among the isolates. Additionally, we used predictive software to identify putative surface-expressed proteins with predicted high antigenicity from the completed genome of UMNK88, a K88-positive porcine ETEC. The prevalence of these genes was examined in porcine ETEC strains and in commensal *E. coli* from healthy pigs. Genes found in ETEC significantly more often than in commensal *E. coli* include Antigen 43 precursor protein, *tatD*, and several other putative outer membrane or exported proteins.

Overall, this study showed that porcine ETEC are polyphyletic, with most isolates falling in clusters within phylotypes A or B1, that genetic content varies between ETEC and commensal porcine *E. coli*, and that virulence and resistance plasmids are widely distributed among porcine ETEC.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction, Specific Aims, and Literature Review

Introduction

Traditionally, the study of porcine enterotoxigenic *Escherichia coli* (ETEC), an important swine pathogen, has focused primarily on plasmid-encoded virulence factors, despite the fact that chromosomal factors are known to play a role in ETEC pathogenesis in humans.[1] Recent studies have established that non-ETEC *E. coli*, such as enteropathogenic, enterohemorrhagic, and extraintestinal *E. coli*, exhibit chromosomal sequence-based similarities that suggest a distinct evolutionary history within each group, despite the fact that many virulence factors are plasmid-mediated.[2] Studies of human ETEC revealed a wider variety of chromosomal backgrounds, suggesting the acquisition of horizontally-transferred ETEC virulence genes by multiple lineages of *E. coli*.[3]

To date, little chromosomal analysis of porcine ETEC has been conducted. It is unknown whether a particular chromosomal background containing additional virulence factors or uniquely suited to support virulence plasmids is necessary for a strain to become a porcine ETEC, or whether any *E. coli* is equally likely to acquire plasmids and become a virulent porcine ETEC. In an attempt to begin to answer these questions, we will sequence several housekeeping genes in porcine ETEC samples and compare them to previously sequenced *E. coli* strains. We will perform multilocus sequence analysis (MLSA) based on the gene sequences and construct a phylogenetic tree, which is an established method of determining relationships between *E. coli*.[4]

Based on the discovery of chromosomally-mediated virulence factors in human ETEC, and because of increasing evidence that chromosome-plasmid interactions occur

and may affect a bacterium's ability to carry a particular plasmid [5 6 7 8], we predict that porcine ETEC strains are limited to subsets of phylogenetically-related strains, compared to other types of *E. coli*. As a result, we hypothesize that they will group together on the phylogenetic tree, based on the analysis of evolutionarily conserved housekeeping genes. We expect that this group will consist of multiple clusters, with isolates grouping by fimbrial type.

Further, we expect that not only plasmid-encoded, but also chromosomally-encoded genes are differentially distributed throughout porcine ETEC and porcine commensal *E. coli* strains. We will sequence two prototypical ETEC strains and their plasmids and identify novel genes of interest based on predicted protein description, location, and structure. Subsequently, we will screen porcine ETEC samples and fecal *E. coli* from healthy pigs for those genes. Identification of porcine-ETEC-specific genes could be useful in expanding our understanding of ETEC, finding ETEC-specific antigens, and in future attempts at controlling the disease. We will also detect plasmid-encoded genes to infer information about the distribution of select resistance and virulence plasmid carriage in porcine ETEC isolates.

Specific Aims and Hypotheses

- **Specific Aim 1 –** Describe the phylogenetic distribution of porcine ETEC using multilocus sequence analysis (MLSA).
- **Specific Aim 2 –** Starting with a porcine ETEC genomic sequence, use predictive programs to identify genes in porcine ETEC that are less frequently found in commensal *E. coli*.

- **Specific Aim 3 –** Describe the distribution of select resistance and virulence plasmid carriage in porcine ETEC.

Hypotheses: Porcine ETEC strains share a common chromosomal background and will group together on a phylogenetic tree of diverse *E. coli*. This grouping will likely consist of several clusters of isolates belonging to the same phylotypes and with similar plasmid content. Chromosomal and plasmid-encoded genes that are more common in porcine ETEC than in porcine commensal *E. coli* will be identified.

Literature Review

Enterotoxigenic *Escherichia coli* (ETEC)

Escherichia coli are gram-negative, facultative anaerobic, flagellated bacilli. Discovered by Theodore Escherich in the late nineteenth century, *Escherichia coli* is one of the most-studied microbial organisms, and as such we know a great deal about its genetics, molecular biology, and population dynamics.[9] The primary habitat of *E. coli* is the vertebrate gastrointestinal tract, and it is also found in the environment, surviving after excretion from the host.[10] As important members of the gut microbiota, they are beneficial to their mammalian hosts by further breaking down digestive contents and preventing the overgrowth of harmful microorganisms.[11] While most *E. coli* are non-pathogenic, they may act as opportunistic pathogens, and diverse disease-causing pathotypes also exist. These include extraintestinal pathogens, such as uropathogenic *E. coli* that cause urinary tract infections, and neonatal meningitis strains that cause disease in human infants. Several diarrheagenic pathotypes exist, including enteroaggregative *E. coli* (EAEC), enteropathogenic *E. coli* (EPEC), enterohemorrhagic *E. coli* (EHEC), enteroinvasive *E. coli* (EIEC), and enterotoxigenic *E. coli* (ETEC).[12] In most disease-causing *E. coli*, pathogenicity is conferred by virulence genes encoded by plasmids, bacteriophages, or on chromosomal pathogenicity islands.[13]

ETEC, the focus of this work, are characterized by their possession of colonization factors that allow the bacteria to adhere to intestinal epithelial cells, and by the production of specific toxins that result in profuse, watery diarrhea. In humans, ETEC is the leading cause of traveler's diarrhea worldwide and can be fatal

to infants and young children.[14] As an example, a recent study of disease among young children in Bangladesh identified ETEC as the most common cause of diarrheal disease. Further, children diagnosed with ETEC diarrhea remained underweight and had stunted growth at the end of the 2-year follow-up period. [15] ETEC strains also have the ability to cause illness, poor growth, and death in many animal species.

ETEC in Swine: Significance to the Swine Industry

Diarrhea, particularly in young pigs, is a major cause of financial and death losses to the swine industry. There are many causes of diarrheagenic disease in pigs, including bacterial diseases, such as *Lawsonia intracellularis* and *Clostridium perfringens*; viral diseases, such as transmissible gastroenteritis virus (a coronavirus) and rotavirus; and parasitic infections, such as coccidiosis. Importantly, ETEC is the most common cause of diarrhea in young pigs, resulting in a disease known as enteric colibacillosis and causing an estimated 42% of piglet death losses in the United States.[16] Even when other pathogens are responsible for the primary disease process, porcine ETEC are likely to play a secondary role, contributing to morbidity and mortality.[17] ETEC are generally well-controlled in neonatal piglets due to vaccines administered to the sow before farrowing, with protective antibodies passed to the piglets through colostrum and milk.[18] ETEC remain a challenging problem at the time of weaning, when a variety of factors, which will be discussed later in this work, make the piglets particularly vulnerable to infection. In the United States, disease caused by ETEC in weaned piglets alone is estimated to cost $90 million annually, due to mortality, morbidity, poor growth, and the cost of treatment.[19] Post-weaning enteric colibacillosis is a problem worldwide. It

appears that the incidence and severity of disease have increased over the past decade, which may be due to changing management practices or to the emergence of new, more virulent clones.[20] The possibility that new strains and virulence genes have recently emerged points to the need for a better understanding of porcine ETEC and new prevention measures to mitigate the impact of this disease.

Pathogenesis

ETEC are spread via the fecal-oral route. They colonize the small intestine by binding to receptors on intestinal epithelial cells. In the intestine, the bacteria reproduce rapidly, typically reaching approximately 10^9 ETEC cells per gram of intestine in the mid-jejunum and ileum.[13] ETEC produce toxins that cause electrolyte and water loss from the small intestine, manifesting as profuse, watery diarrhea. Fluid loss may lead to dehydration, depression, metabolic acidosis, and death.[18] In pigs, reactions to lipopolysaccharide released by ETEC in acute infections can result in shock and death before diarrhea is even evident.[13] ETEC infection can also lead to secondary septicemia, icterus, petechial hemorrhage, and splenomegaly, likely passing through mesenteric lymph nodes before entering the bloodstream and causing systemic infection.[21] On histopathology, ETEC can be seen adhering to the jejunum and ileum, approximately 1 micrometer from the microvilli, with that distance being spanned by the fimbriae. There are no microscopic lesions, except in cases of shock or septicemia, which can include microscopic hemorrhage, fibrinous thrombi, and necrotic villi throughout the stomach, small intestine, and colon.[13]

Neonatal ETEC Infection

Newborn pigs are colonized with *E. coli*, potentially including ETEC, through oral ingestion from their environment and particularly from the sow's teats. The gastrointestinal tract of newborn pigs provides a favorable environment for ETEC, as the pH of the stomach and duodenum is less acidic and the production of digestive enzymes is low when compared with older pigs. ETEC diarrhea in very young pigs (up to 2 weeks of age) may be less watery than diarrhea in older pigs, but dehydration is still a major health risk.[13] A maternal ETEC vaccine is commonly administered to sows and gilts before farrowing, as will be discussed in more detail later. Neonatal pigs are generally protected from ETEC infection by antibodies in the colostrum and milk of the dam, but they are more susceptible to ETEC infection if concurrent disease processes are present and cause immunosuppression. For example, one study on a Japanese farm experiencing an ETEC outbreak found that only piglets also infected with porcine reproductive and respiratory syndrome (PRRS) virus experienced diarrhea or septicemia, although ETEC was also isolated from asymptomatic piglets that were not infected with PRRS virus.[22]

Post-weaning ETEC infection

While neonatal ETEC diarrhea is generally well-controlled, post-weaning enteric colibacillosis is a more persistent problem. While the disease is ultimately caused by ETEC, it is mediated by a variety of factors, such as the stress of weaning, the cessation of antibodies ingested in the sow's milk, changing receptors on the intestinal epithelium as the pigs age, and change in diet. Diarrhea typically starts within the first week of weaning, lasts up to a week, and causes weight loss.[20] The diarrhea typically has a yellow or gray appearance and is very fluid. Diarrhea caused by ETEC is alkaline, as

opposed to the acidic diarrhea caused by other etiologies, which may be useful for on-farm diagnosis.[18] In affected herds, piglet mortality can be greater than 25%, in the absence of treatment.[13]

Key ETEC Virulence Factors

ETEC are characterized by the production of specific adhesins and enterotoxins, which will be described below.

Adhesins

Adhesins, or colonization factors, allow ETEC to persist in the gastrointestinal tract. Most of the adhesins are fimbriae, long proteinaceous structures produced by the bacteria that attach to receptors on the intestinal epithelial cell surface. The adhesins most common in pigs are fimbrial adhesins K88 (also called F4), K99 (F5), 987P (F6), F41, and F18, as well as the non-fimbrial Adhesin Involved in Diffuse Adherence-I (AIDA-I).[23 24] Additional, as-yet unclassified fimbriae may also exist in porcine ETEC. Recently, *E. coli* isolated from pigs in Vietnam were found to contain enterotoxin genes but lacked known fimbriae. Further investigation showed that these isolates produced an unidentified type of fimbriae *in vitro* and caused acute diarrhea when inoculated in 1-day old pigs.[25] There is some geographic variation in the predominant colonization factors present in porcine ETEC, but in the United States, K88 and F18 are identified in most cases of post-weaning enteric colibacillosis.[20]

K88 (also called F4) fimbriae enable bacterial colonization of the intestinal epithelium in pigs of any age, from neonatal, to post-weaning, to finisher pigs. Until recent years, K88-positive strains have classically been responsible for post-weaning diarrhea, though F18 and other fimbrial types appear to be emerging causes of post-

weaning diarrhea.[26] K88 fimbriae are encoded by the *fae* locus on a plasmid, and the fimbrial major subunit is the actual adhesin. There are three variants of K88, based on antigenic and sequence variation: ab, ac, and ad, with K88ac being the most common. K88-positive ETEC are almost universally hemolytic.

K88 adheres to the intestinal mucin-type sialoglycoprotein (IMTGP) receptor located on the intestinal epithelial cell brush border. Several proteins in porcine milk, such as lactadherin, lipoprotein lipase, and beta-casein, have been shown to bind K88 fimbriae and likely provide additional protection to nursing piglets. These proteins have been shown to prevent K88ac ETEC attachment to porcine intestinal cells *in vitro*.[27]

K99 is plasmid-encoded. These fimbriae bind to a receptor located in the posterior small intestine of neonatal pigs, calves, and lambs.[13] K99 ETEC is generally well-controlled in neonatal pigs due to protection from a maternal vaccine.

987P can be chromosomally or plasmid-encoded. This thick, rod-shaped fimbriae attaches in the posterior small intestine of neonatal pigs to cause disease. It is less of a risk in older pigs, where it preferentially binds to receptors on intestinal mucus, rather than the intestinal epithelium itself. In nursing neonates it is controlled by maternal vaccine, and it is rarely, if ever, a problem in older pigs.[13]

F18 fimbriae are long and flexible, with a characteristic zigzag pattern.[28] They are plasmid-encoded by the *fed* locus, and binding is mediated by the FedF adhesin. There are two variants, based on serological and sequence differences: ab and ac. F18ab is relatively poorly expressed (in vitro) and associated with Shiga toxin Stx2e-producing strains that cause edema disease. F18ac is more efficiently expressed and associated with

post-weaning enteric colibacillosis. Like K88-positive ETEC, F18-positive ETEC are hemolytic.[13]

Intestinal epithelial receptors for F18 are absent from neonatal pigs and increase as pigs age. Receptors begin to be expressed prior to weaning, which typically occurs between 3 to 4 weeks of age.[18]

Porcine ETEC are also known to possess afimbrial adhesins. For example, the AIDA-1 afimbrial adhesin is often encoded on the same plasmid as F18. AIDA-I is an auto-transported protein that mediates diffuse attachment to the intestinal epithelial cells. It is usually associated with F18 and/or toxin Stx2e, as well as enteroaggregative heat-stable enterotoxin *astA*.[2930]

Enterotoxins

Porcine ETEC enterotoxins are plasmid-encoded. While they disturb intestinal fluid metabolism and can cause serious disease and death, they rarely produce lesions or morphological changes in the intestinal mucosa.

Heat -stable enterotoxins are low molecular-weight toxins that are resistant to temperatures of 100°C for 15 minutes. They are poorly antigenic. There are three different kinds: STa, STb, and EAST1.[13]

STa (encoded by gene *estA)* is a small protein (18-19 amino, about 2 kDa) that is produced as a pre-pro-peptide, transported across the inner membrane, folded in the periplasm, and secreted through outer membrane channel TolC. There are only minor differences in the porcine form and the predominant human form of STa, and the toxin appears to be able to cause disease in either species, assuming successful colonization of

the intestine.[31] STa binds to a guanylate cyclase C glycoprotein receptor on the brush border of intestinal villous and crypt epithelial cells. These receptors are principally localized in the posterior jejunum. STa binding activates guanylate cyclase, stimulating the production of cyclic GMP.[32] cGMP elevation in the cell activates a protein kinase which phosphorylates a membrane chloride channel (cystic fibrosis transmembrane conductance regulator, or CFTR), causing the secretion of Cl- and HCO^{3-} , as well as inhibiting uptake of Na+. Hypersecretion causes acute diarrhea. Receptors in neonatal animals have a higher affinity for STa than those in older animals, so this is particularly a problem in young animals.[13]

STb (encoded by gene *estB)* is larger than STa (48 amino acids, 5 kDa) and has a different composition and mechanism of action.[24] The protein precursor is released into periplasm and converted to the active form, containing 2 disulfide bonds. It is transported across the membrane through TolC. STb binds to a sulfatide epithelial cell receptor, causing Ca^{2+} uptake by the cell.[33] This activates protein kinase C, which activates the CFTR. Elevated calcium also stimulates secretagogues prostaglandin E2 and 5-hydroxytryptamine. This results in the secretion of water and electrolytes in the duodenum and jejunum via unknown mechanisms. Ultimately, STb makes epithelial cells permeable, but does not kill them. Compared with STa, STb is more important in causing disease in older animals, from a few weeks of age to adult.[13]

Another toxin called EAST1 (encoded by gene *astA)* is related to STa. It is commonly found in porcine ETEC, but unlike the other heat-stable toxins, it can also be found in other types of *E. coli*, including EPEC and STEC. EAST1 is most often observed in F18 (with or without Stx2e toxin) edema disease-causing strains and K88

diarrheagenic strains. It shares a similar structure with the toxic portion of STa and appears to interact with STa receptor guanylate cyclase C, likely acting by the same mechanism.[34]

Heat-labile enterotoxins (LT) become inactivated within 15 minutes at 60°C. They are highly antigenic, and may be altered and used as adjuvants to stimulate the mucosal and systemic immune systems.[35] In pigs, LT-positive ETEC are usually K88-positive and also produce STb.[13] Two types of LT exist, LT-I and LT-II. These toxins are similar except for slightly different binding specificity, and LT-I is usually the form found in porcine ETEC. LT-I is a high-molecular-weight toxin complex (~84 kDa) consisting of a biologically active A subunit and five B subunits that bind GM1 ganglioside receptors on the surface of intestinal epithelial cells. LT-I can also bind other receptors, such as GD1b, asialo GM1, GM2, and numerous galactoproteins and galactose-containing glycolipids. LT-I is similar in structure and function to cholera toxin, though differences in the A1 portion of the A subunit are thought to result in the lower toxicity of LT compared to cholera toxin.[36]

LT-I is transported across the outer membrane via a type II secretion pathway, and then binds to lipopolysacchariade on the outer surface of the cell. When ETEC attaches to the intestinal epithelium, LT-I B subunits bind to receptors on the epithelial cells and the toxin is taken into the epithelial cells via receptor-mediated endocytosis. It is transported to the Golgi and the endoplasmic reticulum, where it dissociates. A fragment of the A subunit translocates into the cytoplasm and transfers an ADP-ribose moiety from NAD to the alpha subunit of the adenyl cyclase regulatory G-protein located in the basement membrane of the cell. This causes adenyl cyclase to be permanently

activated, and the resulting high levels of cAMP in the cell activate the CFTR membrane channel through phospohorylation of protein kinase A.[37] Again, Cl- and HCO^{3-} ions are secreted, and Na+ absorption is decreased. LT-I permanently alters the cell, which will continue to hypersecrete until it dies.

A 2003 study of *E. coli* from 170 diarrheic and 120 non-diarrheic pigs examined the prevalence of colonization and virulence factors. The most commonly identified combination of virulence factors was STb plus AIDA-1, found in 20.5% of the isolates from diarrheic piglets. EAST1, which was found in many isolates from diarrheic piglets, was also found in isolates from non-diarrheic piglets and experimentally did not cause diarrhea in colostrum-deprived newborn pigs. Therefore, EAST1, while often found in ETEC, is not a reliable marker for diarrheic disease in pigs.[38] Indeed, another 2007 survey of the prevalence of virulence genes in 304 post-weaning diarrhea isolates found that all toxin genes (including LT, STa, STb, STx2e) except EAST1 were associated with K88- or F18-positive isolates.[39]

Comparison of porcine and human ETEC virulence factors

Colonization factors and their corresponding intestinal epithelial receptors vary between human and porcine ETEC. In contrast, heat-stable and heat-labile enterotoxins produced by human and porcine ETEC are highly similar.[40], [41] A recent study by Zhang *et al.* demonstrated that pigs are a good model for human disease. Enterotoxin genes from human or porcine ETEC were inserted into *E. coli* strains containing porcine colonization factors. The toxins from human and porcine ETEC produced the same clinical signs of illness in piglets.[31]

While most colonization factors and enterotoxins are encoded on plasmids, additional chromosomal factors are known to play a role in human ETEC pathogenesis. These include TibA, an outer membrane glycoprotein adhesin; Tia, an invasion locus; and LeoA, which mediates toxin secretion.[42],[43],[44]

Diagnosis of porcine disease

ETEC diagnosis is done using fecal samples or by swabbing the ileum of a necropsied pig and submitting the sample to a diagnostic laboratory. The laboratory can detect the presence of *E. coli* and ETEC-specific toxins, as well as provide antimicrobial susceptibility information. Usually, ETEC can be identified by the characteristic profuse, watery diarrhea. Porcine transmissible gastroenteritis (TGE) is the main differential as it causes a similar diarrheal disease, but it is far less common than ETEC. ETEC may also be recovered from pigs with concurrent disease processes, in which case it may be acting as an opportunistic pathogen in immunocompromised animals.

Prevention & Treatment

When ETEC occurs on a farm, outbreaks tend to be severe, with widespread morbidity and mortality. Prevention of the disease is the primary goal, since treatment is challenging. Prevention efforts focus on a variety of management practices and are subject to limited vaccination options. Antibiotics are the mainstay of treatment, though their efficacy is limited by widespread and ever-increasing antimicrobial resistance in ETEC.[45]

While ETEC is generally well-controlled in neonatal pigs, due to the usage of vaccines administered to dams, post-weaning diarrhea remains a much bigger problem.[28] At the time of weaning, factors such as stress (due to separation from the dam, a changing

environment, and transport), differential expression of intestinal receptors for ETEC toxins, the lack of maternal antibodies ingested in milk, and adjusting to a new diet make the pigs especially vulnerable to infection. As such, the management practices described here will focus on those related to post-weaning, rather than neonatal, ETEC.

Management Practices

As with many other pathogens, biosecurity is an important part of ETEC prevention and can help keep the organism off the farm. Once the organism is found in an operation, it is typically very persistent. ETEC have been shown to survive in the environment for up to 6 months, when protected by manure.[46] Even drastic measures, such as using flame-throwers or formalin to disinfect facilities, have been unsuccessful in eradicating the pathogen from infected operations. Factors associated with ETEC infection include poor hygiene, ineffective disinfection, continuous farrowing systems, low environmental temperature (below 25 deg C), inadequate bedding, and air drafts.[13]

Diet modification

Diet can greatly influence the course of the disease in weaned pigs**.**[47] A diet high in milk products and other high-energy products from animal sources—particularly spray-dried plasma--encourages feed intake and can reduce disease incidence and severity. [46] Glutamine supplementation has protective effects and has recently been shown to increase the expression of genes related to prevention of oxidative damage in porcine intestinal epithelial cells and to boost growth rates in weaned pigs.[48] A recent study demonstrated reduced cytokine response and increased tight-junction (claudin-1 and occludin) protein expression in the intestines of pigs fed a glutamine-supplemented diet, compared to pigs fed a control diet.[49] Adding zinc to the feed, at levels so high that

they have antimicrobial effects, is helpful.[50] This method is not without complications, since persistently high levels of zinc in the manure can contaminate soil, making waste disposal difficult. Another feed additive that can have a beneficial effect is glutamine.

While dietary fiber is useful in reducing diarrhea in older pigs, it does not ameliorate ETEC diarrhea in young pigs. Also, soy in the diet appears to promote ETEC infection. Soy proteins cause inflammation that deepens intestinal crypts, increasing ETEC persistence in the intestine.[47] Additionally, electrolytes may be added to water to encourage water intake and promote optimal hydration, for preventive and palliative effects.

Weaning time

Weaning typically occurs at 3 to 3.5 weeks of age. Delaying weaning to 4 weeks or more has been found to decrease ETEC incidence and the severity of the disease.[51]

Treatment with antibiotics

When ETEC outbreaks do occur, treatment options are usually very limited, as ETEC demonstrate high levels of antimicrobial resistance. Multidrug antimicrobial resistance may be driven by the prophylactic use of antibiotics in swine operations. Indeed, the withdrawal of growth-promoter antibiotics in Scandinavia resulted, at least initially, in increases in post-weaning diarrhea due to ETEC.[52] Due to the high levels of resistance, susceptibility information, as determined by a diagnostic lab, should drive therapeutic drug choice. Antibiotics may be administered intravenously or orally (in water or feed). Even after ETEC infections clear, pigs often continue to grow poorly and are increasingly susceptible to other diseases. Therefore, even pigs that recover from ETEC-related illness may eventually be euthanized.

Existing Vaccines

Pre-farrowing, sows and gilts are vaccinated with an inactivated vaccine containing toxins and fimbrial antigens. The Porcine Porcoli® vaccine, made by Merck Animal Health, is one of several commercial vaccines and contains the ETEC LT toxoid and fimbrial antigens K88 (ab and ac), K99, and 987P.

Neonatal nursing pigs are protected from ETEC by antibodies passed from the dam through the colostrum (primarily IgG) and milk (primarily IgA). Only post-natal antibodies transfer in pigs. IgG and IgM antibodies in the colostrum are actively taken up by intestinal epithelial cells via pinocytosis. This process is very effective and neonates usually attain serum antibody levels near those found in adults.[53] Subsequently, immunoglobins in milk are not systemically absorbed, but they provide local protection by binding to pathogens such as ETEC, preventing their attachment to the mucosa. Secretory IgA is especially effective, as the secretory component enhances its stability in mucosal secretions by increasing resistance to proteolytic enzymes and gastric acid.[54]

Selective breeding

Both K88 and F18 intestinal epithelial receptors are encoded by single loci and inherited as simple Mendelian, dominant traits, so breeding homozygous recessive animals is being attempted to avoid the disease.[13] There are risks associated with this approach. If the sow is resistant (homozygous recessive) but the piglets are heterozygous, the sow will not produce antibodies for K88, leaving the piglets very susceptible to infection. The selective breeding strategy is complicated by the fact that only the chromosomal regions, and not the specific genes, for the receptors has been identified.[55]

Additional vaccine attempts

On farms with ETEC infections, an autologous vaccine can be made by inoculating hens with ETEC fimbriae and subsequently adding egg yolk antibodies from these hens to the pig feed to prevent infection.[56] This method is effective, but it is expensive and difficult to do on a large-scale.

It has been observed that the *faeG* major subunit in K88-positive strains and the *fedF* gene in the fimbrial operon of F18-positive strains are well conserved, and there are ongoing efforts to design vaccines around these antigens.[57] [58] Recently, *fedF* was successfully introduced into genetically engineered tobacco plant chloroplasts, and the resultant protein was found to be immunogenic in mice.[59] Similarly, FaeG has been produced by barley plants and was found to stimulate the production of anti-K88 antibodies in mice. These antibodies blocked K88-positive ETEC strains from attaching to porcine intestinal enterocytes *in vitro*.[60] In a promising recent study, portions of LT and STa toxins were embedded in the ETEC K88ac fimbrial FaeG protein. These chimeric fimbriae were used to inoculate rabbits, which subsequently produced antibodies against K88ac, LT, and STa. The antibodies were shown to prevent fimbrial attachment to porcine enterocytes and neutralized LT and STa toxins *in vitro.*[61]

Development of a porcine ETEC vaccine faces several challenges. To be effective, the vaccine needs to stimulate mucosal immunity, blocking the bacteria in the lumen of the intestine to prevent adhesion and toxin effects. Additionally, while passive immunity protects neonatal piglets from disease, maternal antibodies also interfere with successful vaccination pre-weaning.

In summary, porcine ETEC causes serious swine morbidity and mortality, with many challenges to prevention and management. A better understanding of this organism is needed to inform control and treatment efforts.

Chapter 2: Phylogeny of Porcine ETEC

Introduction

Enterotoxigenic *E. coli* are characterized by the ability to produce adhesins that allow colonization of the intestine and production of enterotoxins that cause profuse diarrhea. Disease caused by ETEC is a major problem in young pigs, causing major losses to the swine industry each year. Often acting in concert with other disease processes, ETEC are responsible for poor growth, morbidity, and mortality in young pigs.[13]

While most known ETEC virulence factors are plasmid-encoded, it is also important to examine the bacterial chromosome. In light of increasing evidence of chromosomal background and plasmid interaction and co-evolution, it is particularly important to consider the whole organism.[62 63] Since bacterial strains in the same lineage often inherit and carry many of the same traits, identifying and describing porcine ETEC lineages may improve our understanding of ETEC populations and epidemiology. Examining similarities between strains may also allow us to make inferences about the evolution of virulence and may facilitate the identification of lineage-specific antigens that can be used in vaccines.[64]

Phylogenetic analysis, using sequence data to construct dendrograms, or "trees," is an established method of evaluating relatedness between strains and identifying bacterial lineages.[10] This method relies on nucleotide sequences from conserved 'housekeeping' genes that act as molecular chronometers.[65] These sequences are organized according to similarity and used to infer relative relatedness between strains. It should be remembered that phylogenetic trees only estimate the structure of phylogenies

and should be viewed as possible explanations of evolutionary history, as they are based on indirect evidence.[65]

Recent analyses of human ETEC have revealed that they are polyphyletic. While many other *E. coli* pathotypes group tightly in phylogenetic trees, human ETEC strains appear to be more diverse. A recent study of 1, 019 human ETEC isolates and 8 porcine isolates from several countries revealed 42 different multilocus sequence typing (MLST) lineages, with evidence of virulence factor exchange between lineages. Human and porcine strains were sometimes found to fall in the same lineages. Based on the diversity identified and the patterns in lineages, the authors concluded that human ETEC likely emerged on at least three separate occasions.[64] Examination of a fully-sequenced prototypical human ETEC strain (H10407) revealed that the chromosome was similar to sequenced *E. coli* commensal strains, indicating that virulence was simply the result of plasmid-encoded virulence factor acquisition.[66] Another recent study examined entire genomes of five ETEC isolates from children in Guinea-Bissau. These isolates were selected based on their representation of diverse clonal groups. While no genomic regions were completely conserved across all ETEC strains, there was more genomic content shared between ETEC genomes than between the sequenced ETECs and other pathotypes, suggesting the presence of an ETEC genomic core.[67]

To date, the chromosome of porcine ETEC has been largely understudied, and thus there is a need for better understanding of the chromosomal background of these important pathogens. To that end, we will sequence portions of several housekeeping genes in porcine ETEC samples and use multilocus sequence analysis (MLSA) to infer information about strain evolution and relatedness. Based on the fact that a genomic core

has been identified in human ETEC, and because of increasing evidence that chromosomal background affects plasmid acquisition, carriage, or expression (and ETEC virulence factors are known to be plasmid-encoded), we hypothesize that porcine ETEC strains are related and will group into a limited number of clusters near each other on a phylogenetic tree of diverse *E. coli*.

Materials and Methods

Sample Selection. A total of 89 porcine *E. coli* isolates were obtained from the University of Minnesota College of Veterinary Medicine's Veterinary Diagnostic Laboratory (Table 1). Samples were isolated from fecal or intestinal samples from sick young pigs and were originally submitted for diagnostic purposes. Clinical signs usually included diarrhea. Colonization factor and toxin genes in the samples were detected by multiplex polymerase chain reaction (PCR) assay as part of normal diagnostic testing.[68] In an effort to obtain the most representative samples possible, we selected samples from a variety of U.S. states, from both neonatal and post-weaning aged pigs, and a comparable number of K88- and F18-positive samples. We also included 1 K99-positive sample and 7 samples lacking K88/F18/K99 colonization factors but containing toxin genes associated with ETEC; either ETEC-specific heat-labile or heat-stable enterotoxins, or EAST1 or AIDA, which are commonly, but not exclusively, found in ETEC. Boiled lysis preparations were made using bacteria from 1 mL overnight broth cultures. Bacteria were pelleted by centrifugation and resuspended in 1/5 volume of sterile water, then

boiled for 10 minutes, recentrifuged, and the supernatant (2μl per reaction) was used as template DNA for PCR.[65]

A note about sample names: Most of our sample names start with 'PWD' (an abbreviation for 'post-weaning diarrhea') or 'ND' (an abbreviation for 'neonatal diarrhea') based on the pig age listed in a preliminary report provided by the diagnostic laboratory. Upon closer examination of the clinical records for each case, it became evident that age information in the record was sometimes different from the summary report used to generate our sample names. Names of strains were not changed, although PWD and ND designations are not necessarily reflective of actual age of the animal. Ages are provided in the table. As expected based on the epidemiology of this disease, most ETEC strains came from weaned piglets, and that includes some samples starting with "ND." Weaning age for pigs generally ranges from 14 to 28 days.[69]

Table 1. Porcine ETEC samples used for MLSA analysis.

Strains were selected from diagnostic samples submitted to the University of Minnesota, College of Veterinary Medicine Diagnostic Laboratory between 2007 and 2010. Selection was based on the presence of ETEC virulence factors, with an effort to minimize concurrent disease processes that cause diarrheal disease, when possible. Strains excluded from our final analysis due to poor MLSA sequence data (9 strains) have been excluded from this list. This table contains 38 K88-positive strains, 34 F18-positive strains, 1 K99-positive strain, and 7 strains without any of these colonization factors. All strains listed here carry at least one ETEC-related virulence factor gene.

Sample Name	State	Year Collected	Pig Age (days)	Diarrhea/ Enteritis in Clinical Signs	K88	F18	StA	StB	LT	Stx2e	AIDA	EAST1
ND2-1	MN	2010	42	N	+	-	+	+	+	-	-	+
ND2-11	MN	2010	neonatal	Y	+	-	-	+	+	-	-	+
ND2-16	MO	2010	post-weaning	Y	+	-	-	+	+	-	-	+
ND2-19	MN	2010	unspecified	Y	+	-	-	+	+	-	-	+
ND2-2	MN	2010	post-weaning	Y	-	+	-	+	+	+	-	+
ND2-20	MN	2010	post-weaning	Y	+	-	-	+	+	-	-	+
ND2-21	MN	2010	unspecified	unspecified	-	+	-	+	-	-	-	+
ND2-24	IL	2009	21	Y	+	-	-	+	+	-	-	+
ND2-4	MN	2010	post-weaning	N – weight loss post-weaning	-	+	-	+	+	-	-	+
ND2-7	MO	2009	unspecified	Y	+	-	-	+	+	-	-	+
ND2-8	WI	2009	28	Y	+	-	+	+	+	-	-	+
ND2-9	IN	2009	unspecified	Y	-	+	+	+	+	-	-	-
ND 10	MO	2007	28	Y	+	-	-	+	+	-	-	-
ND 11	MO	2007	21	Y	-	+	-	+	+	-	-	-

Sample Name	State	Year Collected	Pig Age (days)	Diarrhea/ Enteritis in Clinical Signs	K88	F18	StA	StB	LT	Stx2e	AIDA	EAST1
ND 12	MN	2007	21	Y	+	-	+	+	+	-	-	-
ND 13	MO	2007	unspecified	Y	-	+	-	-	-	+	-	-
ND 14	NC	2007	21	Y	+	-	-	+	+	-	-	-
ND 17	MN	2007	21	Y	+	-	+	+	+	-	-	-
ND 18	MN	2007	21	Y	+	-	-	+	+	-	-	-
ND 36	MN	2008	21	Y	-	+	+	+	-	+	+	-
ND 39	IL	2008	10	Y	+	-	-	+	+	-	-	+
ND 4	MN	2007	21	Y	+	-	+	+	+	-	-	-
ND 54	MN	2008	14	N - sudden death, purple skin	+	-	-	+	+	-	-	+
ND 89	NE	2008	post-weaning	Y	-	+	-	+	+	-	-	-
PWD2-10	OK	2009	21	unspecified	-	+	-	-	-	-	+	-
PWD2-11	OH	2009	post-weaning	Y	+	-	-	+	+	-	-	+
PWD2-12	IL	2009	35	Y	+	-	-	+	+	-	-	+
PWD2-13	IL	2009	35	Y	+	-	+	+	+	-	-	-
PWD2-14	IL	2009	unspecified	Y	+	-	-	+	-	-	+	+
PWD2-17	IL	2010	28	Y	-	+	-	+	+	+	-	+
PWD2-18	IL	2010	28	Y	-	+	+	+	-	-	-	-
PWD2-2	OK	2010	21	Y	-	+	-	-	-	+	+	-
PWD2-22	MN	2009	post-weaning	Y	+	-	-	+	+	-	-	+
PWD2-23	MN	2009	28	Y	+	-	-	+	+	-	-	+
PWD2-24	MO	2009	35	Y	+	-	-	-	-	-	-	-
PWD2-25	MO	2009	unspecified	Y	-	+	-	+	+	-	-	+
PWD2-4	OK	2010	28	Y	-	+	-	-	-	+	+	-
PWD2-5	TX	2010	49-64	Y	-	+	-	-	-	-	+	-

Sample Name	State	Year Collected	Pig Age (days)	Diarrhea/ Enteritis in Clinical Signs	K88	F18	StA	StB	LT	Stx2e	AIDA	EAST1
PWD2-6	MN	2010	post-weaning	Y	-	+	+	-	+	-	-	+
PWD2-7	OK	2009	28	Y	-	+	+	+	-	-	-	-
PWD2-9	MN	2009	28	Y	-	+	-	+	+	-	-	+
PWD 1	MN	2007	28	Y	+	-	-	+	+	-	-	-
PWD 137	OK	2008	35	Y	-	-	-	-	-	-	-	+
PWD 139	MN	2008	42	Y	+	-	-	+	+	-	-	+
PWD 144	IL	2008	28	Y	+	-	+	+	+	-	-	+
PWD 148	NC	2008	28	Y	+	-	-	+	+	-	-	+
PWD 15	NC	2007	42	N - swollen eyes, off-feed, lame	-	+	-	+	-	+	-	-
PWD 16	AR	2007	49	N - depression, lethargy, coughing, death	-	+	+	+	-	+	-	-
PWD 18	NC	2007	35	N - neurological signs	-	+	-	+	-	+	-	-
PWD 19	MN	2007	post-weaning	Y	+	-	+	+	-	+	-	-
PWD 22	MN	2007	35	Y	+	-	-	+	+	-	-	-
PWD 30	IA	2007	49	Y	-	+	-	+	+	-	-	-
PWD 33	MO	2007	30	Y	+	-	-	+	-	-	-	-
PWD 35	NC	2007	35	Y	+	-	-	+	+	-	-	-
PWD 36	MN	2007	42	Y	+	-	-	+	+	-	-	-
PWD 37	MN	2007	28	Y	-	+	-	+	+	-	-	-
PWD 38	MO	2007	35	N - nasal discharge, dyspnea, rough coats	-	+	+	+	-	-	-	-
PWD 4	MO	2007	40	Y	-	+	+	+	-	-	-	-
PWD 41	NE	2007	28	Y	-	+	-	-	-	+	-	-
PWD 46	MN	2007	28	Y	-	+	+	+	-	-	-	-
PWD 47	NC	2007	30	Y	+	-	-	+	+	-	-	-
PWD 48	IL	2007	28	N - sudden death	-	+	-	+	-	-	-	-
PWD 5	MN	2007	post-weaning	N - dyspnea, sudden death	-	+	-	-	-	+	-	-

Sample Name	State	Year Collected	Pig Age (days)	Diarrhea/ Enteritis in Clinical Signs	K88	F18	StA	StB	LT	Stx2e	AIDA	EAST1
PWD 52	MN	2007	post-weaning	Y	-	+	-	+	+	-	-	-
PWD 56	MO	2007	35	N - history of influenza at farm	-	+	+	+	-	-	-	-
PWD 6	NC	2007	35	Y	-	+	-	+	+	-	-	-
PWD 60	NC	2007	42	Y	-	+	-	+	+	-	-	-
PWD 62	MO	2007	35	N - sudden death	-	+	+	+	-	-	-	-
PWD 64	MO	2007	49	Y	+	-	-	+	+	-	-	-
PWD 9	MN	2007	35	N - inappetance, thumping	-	+	-	-	-	-	-	-
PWD 92	IL	2007	35	N - sudden death	-	-	-	+	-	-	-	-
PWD 94	MO	2007	42	Y	-	-	-	+	+	-	-	-
K99	MN	2010	19-28	Y	-	-	+	-	-	-	-	-
UMN "K88"	MN	2007	post-weaning	Y	+	-	+	+	+	-	-	-
UMN "F18"	IA	2007	31	Y	-	+	+	+	-	+	-	-
D10-009410	OK	2010	7	Y	-	-	-	-	-	-	-	+
D10-012980	CO	2010	28	Y	-	-	-	-	-	-	+	-
D10-012084	KS	2010	14	unavailable	-	-	+	+	Un-known	-	-	-
D10-013308	NC	2010	1 to 8	Y	-	-	-	-	-	-	-	+

Multilocus Sequence Analysis. The standard housekeeping genes *adk, gyrB, icd, mdh, purA,* and *recA* were selected for sequencing. These are housekeeping genes traditionally used in the Achtman scheme of MLST for analysis of *E. coli*, described online at http://mlst.ucc.ie/.[70] Internal fragments of selected genes were amplified using the following primers from the MLST database (Figure 1 and Table 2).

Figure 1. Location of sequenced genes for MLSA analysis, and percent polymorphism per gene. Dark bars indicate nucleotide sequence and light bars indicate amino acid sequence.

Adapted from: Wirth, *et al*. Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol.* 2006 Jun;60(5):1136-51.

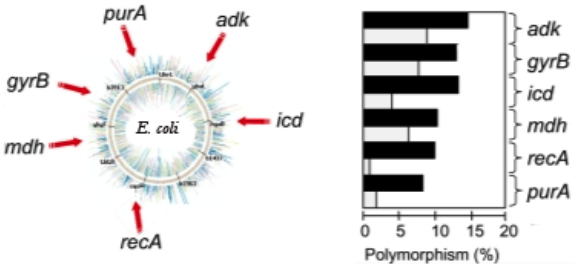


Table 2. MLSA sequencing product primers. Primers used in this study, from http://mlst.ucc.ie/.

Primer Sequence (F=forward; R=reverse)	Product Size	Annealing Temperature
adk F: 5'-ATTCTGCTTGGCGCTCCGGG-3' R: 5'-CCGTCAACTTTCGCGTATTT-3'	583 bp	54°C
gyrB F: 5'-TCGGCGACACGGATGACGGC-3' R: 5'-ATCAGGCCTTCACGCGCATC-3'	911 bp	60°C
icd F: 5'-ATGGAAAGTAAAGTAGTTGTTCCGGCACA-3' R: 5'-GGACGCAGCAGGATCTGTT-3'	878 bp	54°C
mdh F: 5'-ATGAAAGTCGCAGTCCTCGGCGCTGCTGGCGG-3' R: 5'-TTAACGAACTCCTGCCCCAGAGCGATATCTTTCTT-3'	932 bp	60°C
purA F: 5'-CGCGCTGATGAAAGAGATGA-3' R: 5'-CATACGGTAAGCCACGCAGA-3'	816 bp	54°C
recA F: 5'-CGCATTCGCTTTACCCTGACC-3' R: 5'-TCGTCGAAATCTACGGACCGGA-3'	780 bp	58°C

Genes were amplified according to the following PCR protocol: 10 minutes at 95°C; 25 cycles of: 94°C for 30 seconds, the appropriate annealing temperature for each primer for 30 seconds, 72°C for 45 seconds to 2 minutes; then 7 minutes at 72°C and holding at 4°C. The PCR reaction contained 2 μl boiled bacterial template, 1 μl of each forward and reverse primer (diluted to 3.2 μM concentration), 18.2 μl water, 2.5 μl PCR buffer, 2 μl dNTP, and 0.3 μl Amplitaq Gold® (Applied Biosystems) DNA polymerase, for 27 μl total volume per reaction. After amplification, samples were visualized on a gel to confirm the presence of the correctly sized product.

Samples were prepared according to the University of Minnesota BioMedical Genomics Center specifications (1 μl forward or reverse primer, 1 μl PCR primer, and 4 μl water) and submitted for Sanger sequencing. Returned forward and reverse sequence information was examined, edited into consensus sequences, and trimmed to a consistent length using DNASTAR Lasergene® SeqMan software. All six genes were successfully sequenced in sufficient length for 80 of the original 89 samples. Trimmed sequences were imported into MEGA4[71] and aligned using the ClustalW algorithm, along with corresponding sequences from 19 previously sequenced *E. coli* strains obtained from the NCBI database for the purpose of comparative analysis. The final concatenated sequence of six sequenced genes consisted of 3,618 nucleotide positions. A phylogenetic tree of porcine and reference strain *E. coli* was constructed in MEGA4 using the Neighbor-Joining method.[72]

Allele type comparison. Individual gene sequences were entered into the MLST database at http://mlst.ucc.ie/ to obtain allele types for each gene for each sample. These allele types were compared to each other and to human ETEC allele types in the online MLST database by constructing a dendrogram in MEGA4.

Phylotyping. Triplex PCR for two genes (*chuA* and *yjaA*) and a DNA fragment (TspE4C2) was performed, using the interpretive approach by Clermont, *et al*.[73] Phylotypes were determined based on the presence/absence of bands for each PCR product.

Results

MLSA gene products were visualized prior to sequencing (Figure 2). Single nucleotide polymorphisms were observed in the trimmed and concatenated gene sequences after alignment in MEGA4 (Figure 3).

Figure 2. MLSA PCR products. PCR products were visualized on a 1.5% agarose gel prior to submission for sequencing. A 100 kb size standard used to confirm correct product size.

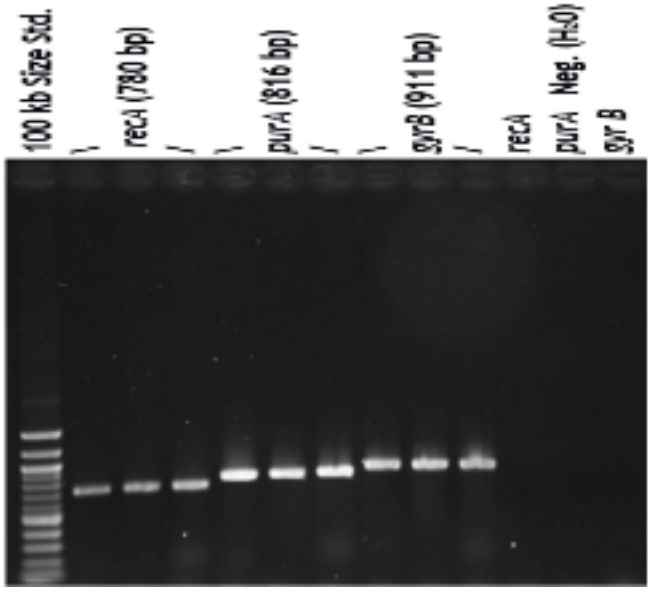
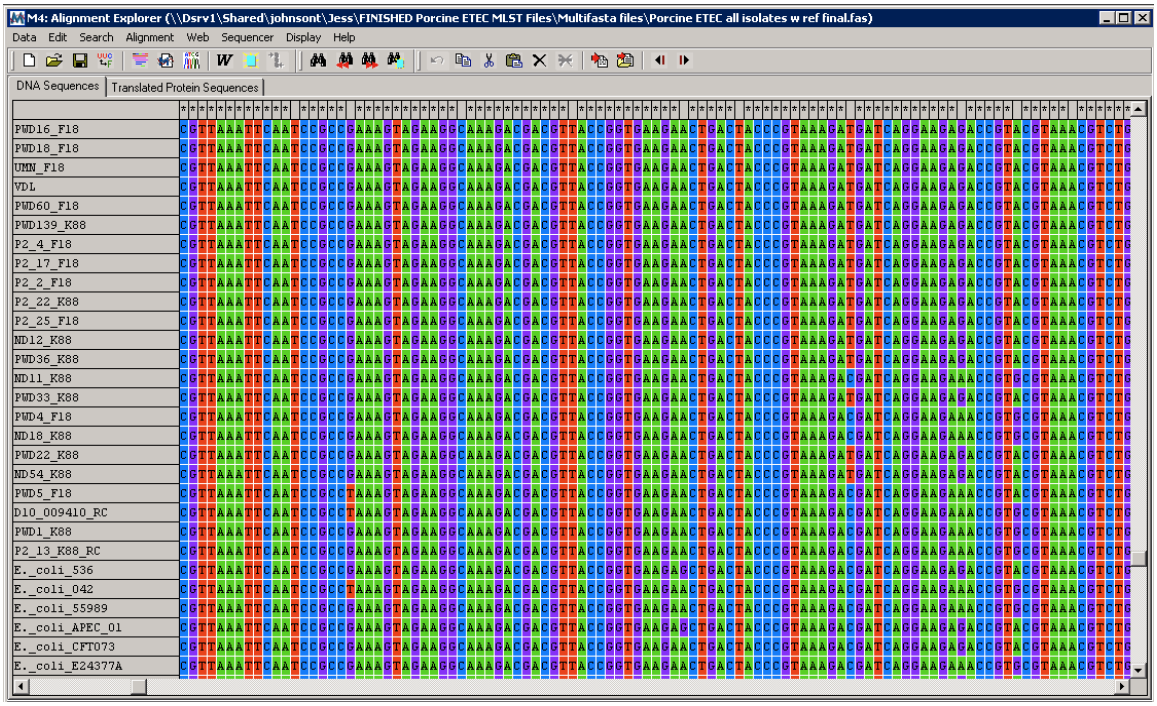


Figure 3. Screen shot of partial MLSA sequence alignment in MEGA4.[71] Single nucleotide polymorphisms are visible as nucleotide changes in vertical columns.



Sequence information from 19 published *E. coli* strains, including members from all 5 *E. coli* phylotypes, were included in our MLSA dendrogram (Figure 4). These 5 groups could be differentiated based on the tree, and there were alternating branches of A and B1 phylotype, which is not uncommon for *E. coli* since A and B1 are 'sister groups'.[74] The confirmation of samples grouping distinctly and according to phylotype supports the validity of our tree structure.

Figure 4. MLSA dendrogram of 101 *E. coli* isolates. Phylogenetic analysis was conducted in MEGA4. Evolutionary history was inferred using the Neighbor-Joining method, based on 3618 nucleotide positions per isolate. The bootstrap consensus tree was inferred from 500 replicates, with branches reproduced in greater than 50% of replicates indicated by a star. The tree is unrooted and drawn to scale. Horizontal distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site.

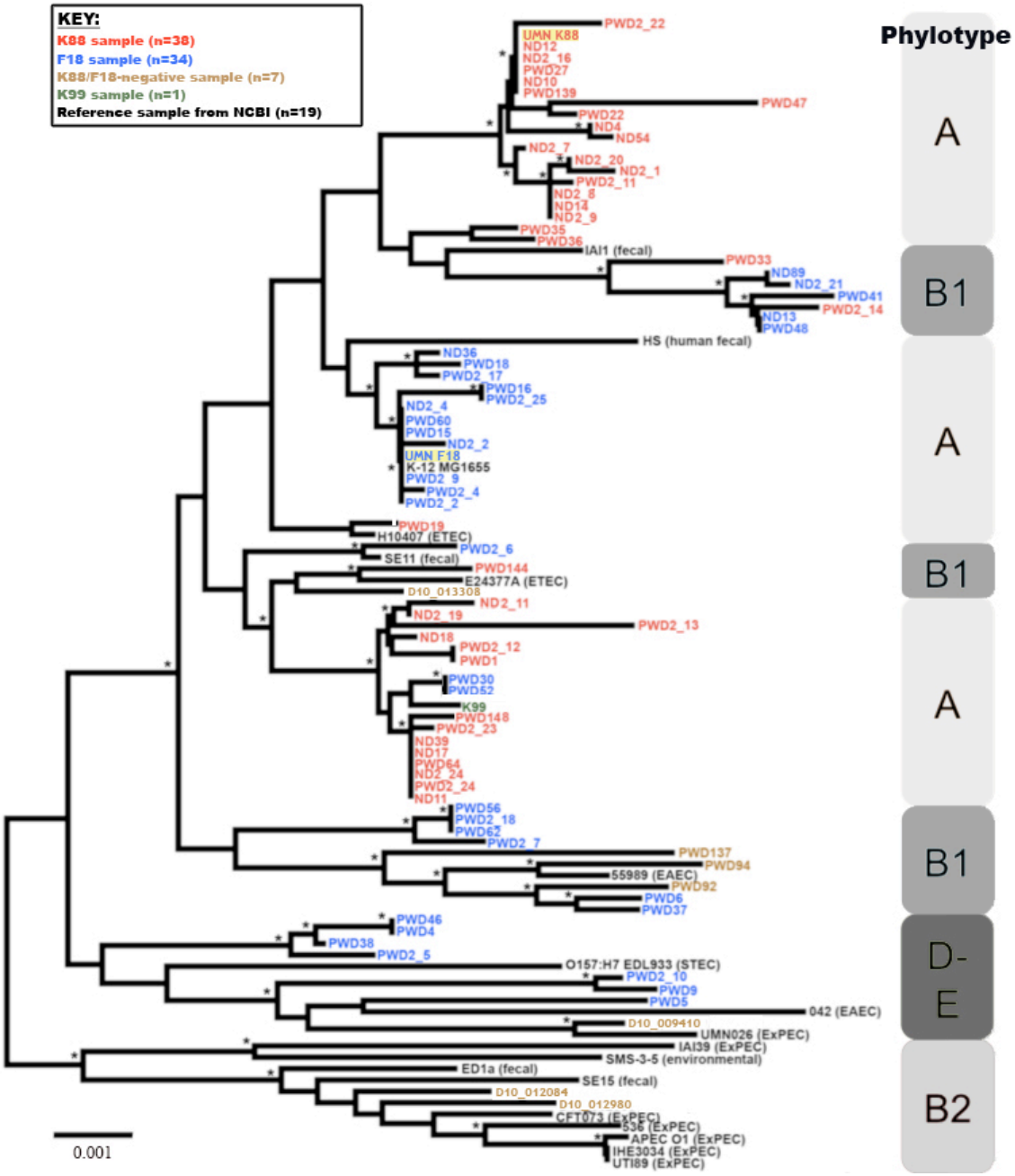
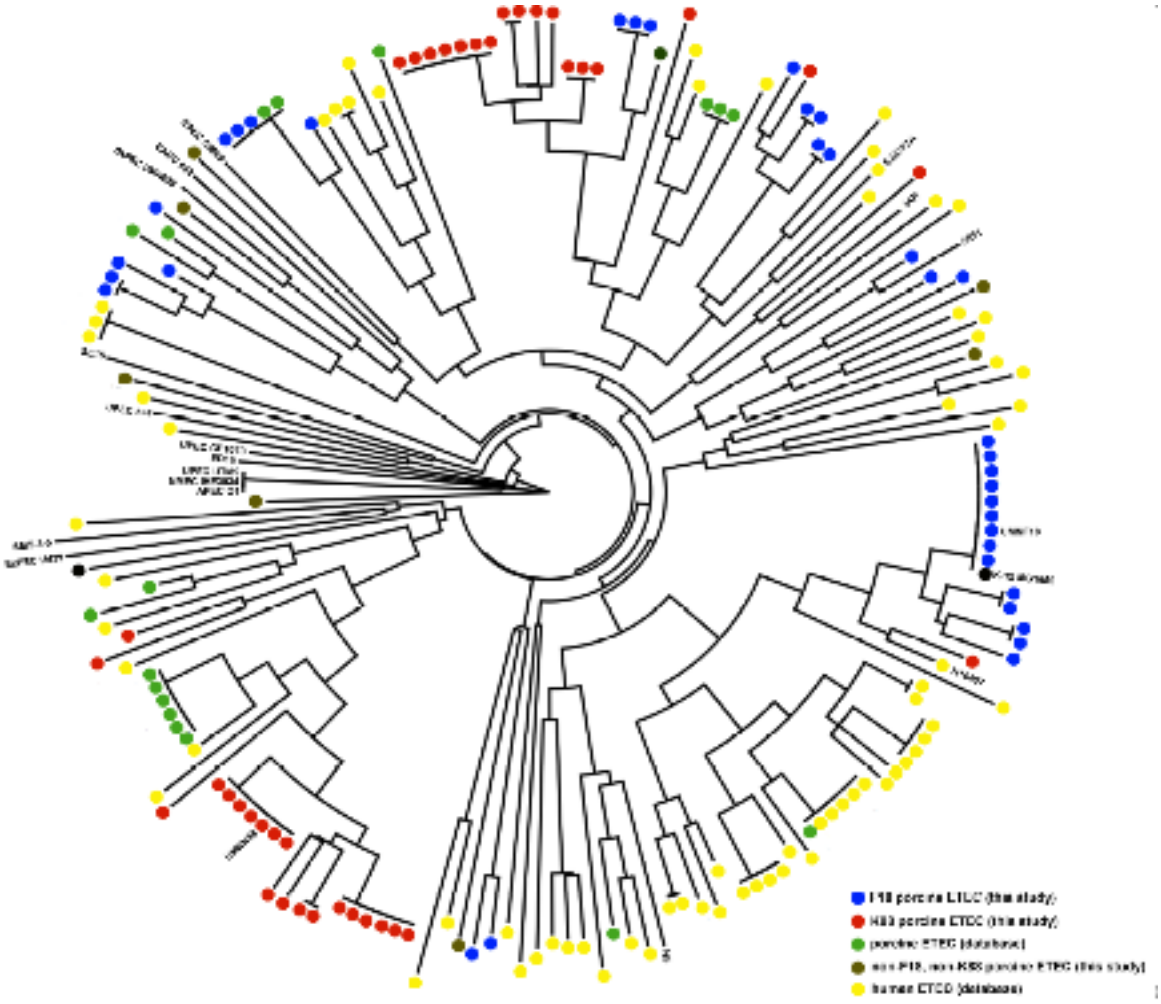


Figure 5. Allele-type dendrogram of porcine ETEC, human ETEC, and reference.

***E. coli*.** Allele types from isolates in this study and existing database isolates were used to construct a Neighbor-Joining dendrogram (http://pubmlst.org/analysis/). The dendrogram was visualized using Mega4.



EAEC = Enteroaggregative *E. coli*
UPEC = Uropathogenic *E. coli*
NMEC = Neonatal Meningitis *E. coli*
ExPEC = Extraintestinal Pathogenic *E. coli*
APEC = Avian Pathogenic *E. coli*
SMS-3-5 = environmental *E. coli*
IAI1, HS, SE11, ED1a = commensal (human fecal *E. coli*)

Porcine ETEC samples are polyphyletic, and most samples belong to phylotypes A and B1.

Based on our MLSA dendrogram, porcine ETEC samples cluster together on multiple tree branches. The majority (87.5%) of the porcine ETEC samples belong to phylotypes A and B1, which are classically associated with commensal *E. coli*.

The distribution of most porcine ETEC isolates in distinct branches within only two phylotypes suggests that limited chromosomal backgrounds are suitable for the acquisition and maintenance of plasmids encoding ETEC virulence factors. Virulence factors likely arose or were horizontally transferred separately into ancestral *E. coli* of these similar lineages. These virulence factors would have subsequently been passed on vertically, accounting for the clustering of samples on multiple branches that we observe here.

Human ETEC and commensal strains are interspersed with porcine ETEC. The two human ETEC reference strains included in the tree and the commensal/laboratory strain of *E. coli,* K-12 strain MG1655, all fall among our porcine strains. This suggests that strains of *E. coli* that are isolated from humans, given the acquisition of the appropriate horizontally-transferred virulence factors, have the potential to become porcine-specific pathogenic ETEC strains, and vice-versa. A published study of human ETEC strain H10407 found that its chromosome was highly similar to the laboratory strain K-12 MG1655, so the grouping of these two isolates in close proximity on our tree was to be expected.[66] These findings also suggest that commensal *E. coli*, as opposed to pathogenic *E. coli,* may have a more favorable chromosomal background for ETEC

plasmids. The circular allele-type dendrogram also shows porcine ETEC, human ETEC, and non-pathogenic *E. coli* samples interspersed throughout the figure, suggesting that their chromosomal background may be preferentially favorable to ETEC virulence plasmids, compared to the backgrounds of other pathotypes. In contrast, most extraintestinal pathogenic *E. coli* group separately on the dendrogram. Interestingly, some enteroaggregative *E. coli*, which frequently carry EAST1, grouped with some of our porcine samples that lacked characteristic ETEC fimbriae, indicating that these porcine samples may actually be enteroaggregative *E. coli.*

F18-positive strains appear to be more diverse than K88-positive strains. Our MLSA dendrogram shows that F18-positive samples are more widely distributed on the tree, compared to the K88-positive samples, which fall primarily into two distinct groups. Additionally, 7 F18-positive samples fall into phylotype D/E, while no K88-positive samples belong to those groups. (Phylotypes D and E are impossible to distinguish based on the phylotyping method used.) The circular allele-type dendrogram of porcine and human ETEC samples also shows that F18-positive samples are somewhat dispersed throughout the tree, while almost all of the K88-positive samples cluster closely in two groups.

Porcine ETEC samples appear to be less diverse than human samples. Based on the circular allele-type dendrogram, human ETEC samples represent more distinct combinations of allele-types and are more widely distributed than porcine samples. Some human ETEC samples fall into one large group, but they are also widely distributed

throughout the figure, suggesting more diversity in the human samples. In contrast, fewer clusters of F18 and K88 porcine samples are seen in this figure, indicating fewer chromosomal lineages of porcine ETEC implicated in disease.

K88- and F18-negative strains containing ETEC-associated toxins may not cluster with K88/F18-positive strains. This indicates that not only plasmid content but also chromosomal background is distinct in these strains. In contrast, the K99-positive strain was found to cluster with K88/F18-positive ETEC. This result must be interpreted with caution due to our limited sample numbers of K88/F18-negative samples.

It is also important to acknowledge that strains D10_009410, D10_012980, and D10_012084, which are phylotype B2 or D/E, may not truly be ETEC. These strains were isolated from pigs with diarrheal disease and contain EAST1 or AIDA genes, but no other exclusively ETEC-related colonization or virulence factors. EAST1 and AIDA can be found in EAEC, EHEC, and EPEC. Strains D10_013308 and PWD 137 also only contain EAST1, but these were found to group with other K88- or F18-positive ETEC strains in phylotype B1. The 5 isolates mentioned here were the only samples used in this study that did not contain genes for ETEC-specific colonization and virulence factors.

Discussion:

Studying the phylogeny of pathogenic bacteria is vital to understanding how pathogenesis evolves. Insights from phylogenetic analyses may unlock new strategies for

disease control, treatment, and prevention. As an increasing body of knowledge builds in this area, we will gain the resources and perspective to better interpret this kind of information.

Based on our examination, the chromosomal background of porcine ETEC is relatively conserved, compared to *E. coli* of other pathotypes. It is polyphyletic, but samples clearly cluster. These clustering samples are identical or highly similar, at least as far as the discriminatory power of our MLSA analysis, indicating genetic relatedness. Samples containing characteristic ETEC colonization factors (K88, F18, K99) are found predominantly in two phylogenetic groups, A and B1, with some F18-positive strains in group D/E. This distribution supports our hypothesis that porcine ETEC are related and share elements of a common chromosomal background, when compared to other types of *E. coli.* A particular chromosomal background may be necessary to support horizontally transferred virulence factors, possibly through acquisition, chromosome-plasmid interaction, or regulation of plasmid-encoded genes. Our results also indicate that the porcine ETEC chromosome is more closely related to commensal *E. coli* than to other pathotypes, such as ExPEC or other diarrheagenic pathotypes. Of course, it is also important to keep in mind that this study was done with clinical diagnostic samples, rather than randomly-selected samples. These samples came from a variety of different states, and we did not use more than one sample per farm, but they are only from certain regions of the United States and may not be representative of the global population of porcine ETEC. As additional porcine and human ETEC and other *E. coli* sequence information become available, this type of analysis will become more robust and will add to our understanding of this important swine pathogen.

Chapter 3: Using Genome Sequencing Towards the Genotyping of Porcine Enterotoxigenic *Escherichia coli*

Introduction.

In our phylogenetic analysis, we determined that porcine ETEC strains appear to have a similar chromosomal background to commensal *E. coli*, based on chromosomal sequence similarity. While this observation is interesting, it is important to remember that this analysis considered only the sequence of a core set of conserved housekeeping genes, representing less than 1% of the total genomic sequence of each organism. Despite the similarities in these regions, we hypothesize that other regions of porcine ETEC and commensal *E. coli* chromosomes are unique and that it will therefore be possible to identify genes that are present in porcine ETEC but absent from commensal porcine *E. coli*. In this chapter, we generate the first completed sequence of a K88-positive porcine ETEC, designated strain UMNK88. We will utilize this sequence to compare it with the published sequence of commensal *E. coli* K-12 strain MG1655, a strain typically used for the comparison of pathogenic versus commensal *E. coli* to identify distinguishing traits. We will place special emphasis on predicted surface-expressed proteins that could have relevance in future attempts at vaccine development. In addition to examining chromosomal genes, we will also examine plasmid sequences from our sequenced porcine ETEC strain for genes that may be uniquely expressed in porcine ETEC. Finally, we will screen ETEC samples from sick pigs and non-pathogenic *E. coli* from healthy

pigs to determine whether the chromosomal and plasmid-encoded genes of interest have different prevalence in pathogenic and commensal porcine *E. coli* populations.

Materials and Methods.

Bacterial strains. The 140 *E. coli* strains used in this study were acquired from porcine diagnostic samples sent to the University of Minnesota, College of Veterinary Medicine's Veterinary Diagnostic Laboratory. Many of these samples are the same as those previously used for MLSA analysis in Chapter 2. These diagnostic samples were obtained from sick young pigs with diarrhea, enteritis, or edema disease and are presumed to be predominantly porcine ETEC strains. An additional 86 *E. coli* strains were obtained from experimental research farm facilities at Iowa State University, Ames, Iowa. These were isolated from fecal samples of healthy pigs, and are presumed to be predominantly commensal *E. coli*. They will be referred to as 'Pig Fecal *E. coli*,' abbreviated PFEC.

Bacterial sequencing. A K88-positive strain was selected for whole-genome sequencing. This strain was submitted to the diagnostic laboratory after collection from a farm in Minnesota in 2007. It was isolated from a pig of post-weaning age (approximately 4 weeks old) whose clinical signs included enteritis. It was identified as an enterotoxigenic *E. coli* based on the presence of enterotoxin genes detected via multiplex PCR by the diagnostic laboratory. In addition to the K88ac adhesin, this strain

is positive for heat labile enterotoxin and heat stable enterotoxins A and B. It belongs to the O149 serogroup (determined by slide agglutination test, performed at Pennsylvania State University's *E. coli* Reference Center) which is the most frequently occurring ETEC serogroup in K88-positive cases of post-weaning diarrhea.[18] This strain is designated UMNK88.

The whole genome of UMNK88 was sequenced to a depth of 30x coverage using pyrosequencing on Roche 454 GS-FLX with Titanium chemistry. Two libraries were generated for the genome, consisting of a shotgun library and a 3-kb paired-end library. Sequencing was performed on a Titanium plate. Sequence reads were assembled into contigs with Newbler Assembler (454 Life Sciences) to resolve full chromosomal and plasmid sequences. Homopolymers were examined for sequence quality and manually edited in SeqMan (DNASTAR, Lasergene). The genome and plasmids were fully assembled and verified for accuracy using PCR.

Gene prediction and annotation

Annotation was transferred from previously annotated *E. coli* genomes to orthologous genes and manually curated. Annotation was carried out using the genome viewers RAST (Rapid Annotation using Subsystem Technology) [75] and Artemis.[76] Coding sequences were predicted using the gene prediction programs Orpheus[77], Glimmer2[78] and Glimmer3[79], then selectively checked using BLAST[80]. GenePRIMP 0.3 [81] was used to help identify problems in the automated annotation. Protein domains were marked up using Pfam (Wellcome Trust, Sanger Institute). Plasmids were also curated manually.

The annotated genome sequence of UMNK88 has been deposited in the public databases (accession numbers: CP002729 for the UMNK88 complete chromosome; CP002730, CP002731, CP002732, CP002733, and HQ023862 for the UMNK88 plasmids pUMNK88_K88, pUMNK88_IncI1, pUMNK88_Ent, pUMNK88_Hly, and pUMNK88_IncA/C.

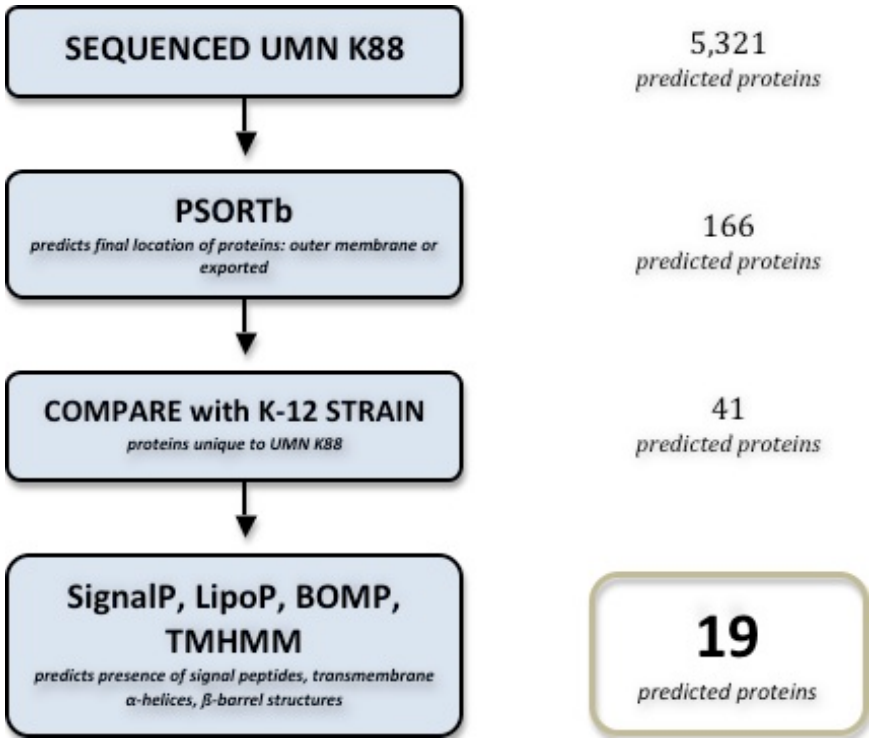
***In silico* analysis.** The 5,321 predicted proteins in UMNK88 were analyzed using the PSORTb 3.0[82] subcellular localization prediction program to identify putative outer membrane and exported proteins. This yielded 166 proteins, which were compared to the non-pathogenic, commensal lab K12 strain MG1655 using protein BLAST. Proteins that matched (>90% amino acid homology) between the strains were excluded from further analysis, resulting in 41 predicted outer membrane or exported proteins that were unique to UMNK88.

The 41 remaining proteins were analyzed according to predicted immunogenicity and suitability as vaccine targets. Using SignalP 3.0 [83] and LipoP [84], 15 of the 41 proteins were identified as containing predicted signal peptides. Predicted beta-barrel structures, which indicate an association with the bacterial outer membrane, were identified in 8 proteins by BOMP [85]. TMHMM (Center for Biological Sequence Analysis, Technical University of Denmark) was used to predict transmembrane alpha-helices. While the presence of alpha-helix structures suggests association with a cell membrane, which is favorable for vaccine candidate selection, proteins with multiple transmembrane regions have been demonstrated to pose problems in the cloning and expression steps necessary for vaccine development [86]. For this reason, we would have excluded any proteins with 3

or more predicted transmembrane alpha-helices, however none of the remaining proteins were found to have more than one predicted transmembrane spanning region.

Ultimately, we identified 19 putative outer membrane or exported proteins that were present in the UMNK88 ETEC strain but absent in the non-pathogenic strain and also met our other criteria: the presence of beta-barrels or predicted signal peptides, with no more than one transmembrane alpha-helix (Figure 6). Virulence plasmids Hly, Ent, and K88 plasmid sequences were analyzed using the same predictive programs to identify genes for surface-expressed proteins. Six genes for uncharacterized proteins were selected for further examination. Uncharacterized proteins were selected since known virulence factors have been examined in the past. This resulted in a total of 25 genes of interest. Genes and predicted proteins are described in Table 3.

Figure 6. Flow chart illustrating process of selecting genes for screening. From the 5,321 predicted proteins in the UMNK88 chromosome, 19 were ultimately selected for testing.



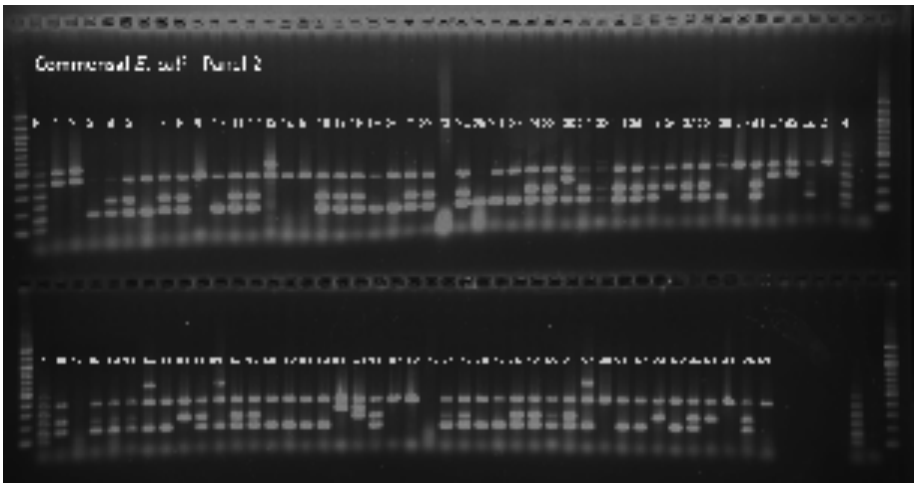
Primer design. Primers were designed for the 25 genes of interest using Primer 3 (version 0.4.0).[87] Four separate multiplex PCR panels were used to detect all 25 genes, and these panels were sometimes further split for better PCR product visualization. Primers are shown in Table 3.

Table 3. Proteins and primers used in this study. Gene letter used in this study is indicated in first column, followed by gene product description, forward and reverse primer sequences, size (in base pairs) of PCR product, start and stop coordinates of gene on chromosome or plasmid. Accession number for UMNK88 chromosome is CP002729, for K88 plasmid is CP002730, and for Hly plasmid is CP002733.

Gene Code Letter	Product	Primer: Forward	Primer: Reverse	Prod siz (bp)	Gene Coordinates (Start)	Gene Coordinates (Stop)
	Panel 1					
A	tonB-dependent vitamin B12 receptor BtuB	cctgactggacagcaacaga	tcccattcgaaaccataagc	567	4682095	4683978
B	putative outer membrane protein/invasin	agtacacaggccggtaatgc	gtctgtttgcttcccggata	428	4943734	4946982
C	putative exported protein	gcacggtgacagacccttat	gacttacgggtgccactgat	319	4911818	4913317
D	antigen 43 precursor protein	caacaccactgtcggacatc	gaggttggtggtggtcagtt	244	1263160	1266009
E	MltA-interacting protein MipA	ggagcagccagaaacagaac	cagttgctggcaaaggtgta	142	4894628	4895374
F	fimbrial usher protein	tgccaaaacacttcaccaaa	ggcgcaataatttgacgatt	501	20629	21774
G	general secretion pathway protein D GspD	ggctggactcagaaatggaa	aactgcattaatccggcatc	364	3590375	3592435
	Panel 2					
H	bifunctional enterobactin receptor/adhesin protein	ccggctatgagaaaaagctg	gtgactccgtgggataagga	600	1235100	1237190
I	DNase TatD	ctgaacaagccggttatcgt	ggcgaagttatcggtggtta	387	1337533	1338330
J	putative hemolysin activator protein	ctgatgcgttaccacactgg	atgtcccgtaaatccagcag	301	1248690	1249736
K	putative flagellin structural protein	cgacggtgatgttgataacg	agcatcacctacagcccaac	481	427503	430520
L	putative type IV secretion system effector protein	aggttgctcaacgcttgatt	ccgctttactggttcggtag	230	242532	243014
M	putative fimbrial protein	tccgattctcaacaacacca	gttccatccaggcaaagaaa	161	778011	778577
N	complement resistance protein TraT	tgatgattgcactggtcagc	caggtttcactcatctgggttt	109	1212774	1212902
	Panel 3					
O	putative phage tail fiber protein	agttcttcgctgaggtgcat	cttcgacggcttaacgctac	458	3006980	3009007
P	putative outer membrane protein	tagcaaatctgatcgcacca	acacggtcagcgtcaacata	339	2342017	2342355
Q	putative outer membrane protein	aagttctggcaatgctggtc	ctggcgtcgaggttgtattc	250	2341405	2341635
R	putative outer membrane precursor Lom	aacgcattacagcgataccc	cgttttcccccttgttgtcag	187	1908248	1908847
S	glucoside specific outer membrane porin BglH	tgaaagtggcgctcatagtg	tcacaccctgctgagttttg	134	1885673	1887346
	Plasmid Panel					
T	Hly plasmid: putative invasin	agtacacaggccggtaatgc	tttgcttcccggataaacac	423	12791	15403
U	K88 plasmid: putative fimbrial subunit	tcaacaattgctgtggcatt	ccttcaggggaagtgaaggt	278	53132	53611
V	K88 plasmid: putative fimbrial subunit	ggcggtgttgctaaaagtct	actcgcctgaccttcagaga	147	54246	54668
W	Hly plasmid: conserved hypothetical protein	ctttttatctgccggctcag	ctgctcccccaatcaactta	500	4912	5490
X	K88 plasmid: putative fimbrial usher protein	taggttactggcccggtatg	ctttgggaacgcctgttaaa	333	52500	53000
Y	K88 plasmid: putative fimbrial subunit	gctaatgctgcaacaagcaa	cccgcagtatctttcagctc	227	53639	54118

Gene prevalence determination. Gene prevalence studies were performed on a collection of 94 porcine ETEC associated with neonatal or post-weaning diarrhea from the Veterinary Diagnostic Laboratory, as well as the 86 PFEC samples from healthy pigs. Initially, runs with different annealing temperatures were completed to determine the ideal annealing time for each panel. PCR thermocycler protocol: 10 minutes at 95°C; 25 cycles of: 94°C for 30 seconds, 58-60°C 30 seconds, 72°C for 1 minute; then 7 minutes at 72°C and holding at 4°C. The PCR reaction contained 2 μl boiled bacterial template, 2 μl of pooled primer mix, 15.5 μl water, 2.5 μl 5x PCR buffer with MgCl2, 1 μl dNTP, and 2 μl Taq polymerase for 25 μl total volume used per reaction. After amplification, samples were electrophoresed in 2% agarose gels run at 150V for 40 min, stained with ethidium bromide, and visualized under ultraviolet light to detect PCR amplification products. A 100 bp size standard was used to confirm appropriate PCR product size. An example of multiplex PCR results is seen in Figure 7.

Figure 7. PCR multiplex panel 2 results for PFEC samples. Products are described in Table 3. Samples are visualized on 2% agarose gel with a 100kb size standard for reference.



Statistical analyses. Gene prevalence in porcine ETEC and PFEC samples was determined using PCR, as described above. To determine whether differences in gene prevalence between groups were significant, proportions of positive genes in ETEC and PFEC samples were first calculated using the adjusted Wald interval.[88] This method provides greater stringency when determining significance in data sets that include some values that are close to zero. In our data set, the prevalence of some genes was zero or close to zero, particularly in PFEC samples. The adjusted Wald interval involves adding two successes and two failures to each calculation of proportion, which also adds four to the sample size.

Adjusted Wald proportion = (x+2)/(n+4),

where x is the number of positive samples and n is the sample size.

Chi-square tests of independence, using the adjusted proportions, were also calculated. Results that yielded a $p < 0.05$ were considered significant.

Using gene prevalence data for our 80 ETEC MLSA samples described in the previous chapter, code was written in R (Wirtchaftsuniversitat, Vienna, Austria) to find combinations of genes such that at least one of the genes was found in the maximum number of ETEC isolates.[89] Genes with ≥20% positive samples in PFEC (7 genes) were excluded from this analysis, since the goal was to identify genes for proteins that are expressed in ETEC but not in commensal *E. coli*. The R code follows:

```
# functions to do the work
countfork <- function(k, m, most=5) {
b <- t(combn(ncol(m),k))
ans <- sapply(1:nrow(b), function(i) sum(rowSums(m[,b[i,],drop=FALSE])>0))
out <- matrix(NA, nrow=nrow(b), ncol=most+2)
out[,2:(ncol(b)+1)] <- b
out[,most+2] <- ans
out[,1] <- k
colnames(out) <- c("num", paste("in", 1:most, sep=""), "count")
out
}
countfor <- function(k, m) {
out <- as.data.frame(do.call(rbind, lapply(k, countfork, m=m, most=max(k))))
out[order(out$count, out$num, decreasing=TRUE),]
}
# read the data
d <- read.csv("MLST_RV_7_15_2011.csv", as.is=TRUE, row.names=1)
# check it
m0 <- as.matrix(d)
```

```
table(addNA(m0))
# convert p/n to TRUE/FALSE
m <- array(m0=="p", dim=dim(m0), dimnames=dimnames(m0))
# count on all data
out <- countfor(1:4, m)
lapply(split(out, out$num),head)[2:3]
# count only on genes 1 to 12
out2 <- countfor(1:4, m[,1:12])
lapply(split(out2, out2$num),head)[2:3]
# count only on genes 1 to 12 without 4
out3 <- countfor(1:4, m[,c(1:3,5:12)])
lapply(split(out3, out3$num),head)[2:3]
# write output files
write.csv(out, file="onAll.csv", row.names=FALSE)
write.csv(out2, file="on12.csv", row.names=FALSE)
write.csv(out3, file="on12wo4.csv", row.names=FALSE)
```

Results

All genes of interest were found in a greater proportion of the ETEC samples than in the PFEC samples, although some differences were statistically insignificant. The following genes were found significantly more frequently in ETEC than in PFEC based on the Chi-square test of independence ($p < 0.05$): B, C, D, E, G, H, I, K, L, N, O, P, Q, S, T, V, W, X, and Y.

Figure 8. Proportions of samples positive for each gene. The chart shows adjusted Wald proportions of positive samples for each gene of interest in PWD and PFEC collections. Asterisks under the sample names indicate significance ($p<0.05$), as determined by a Chi-square test of the adjusted proportions for each gene.

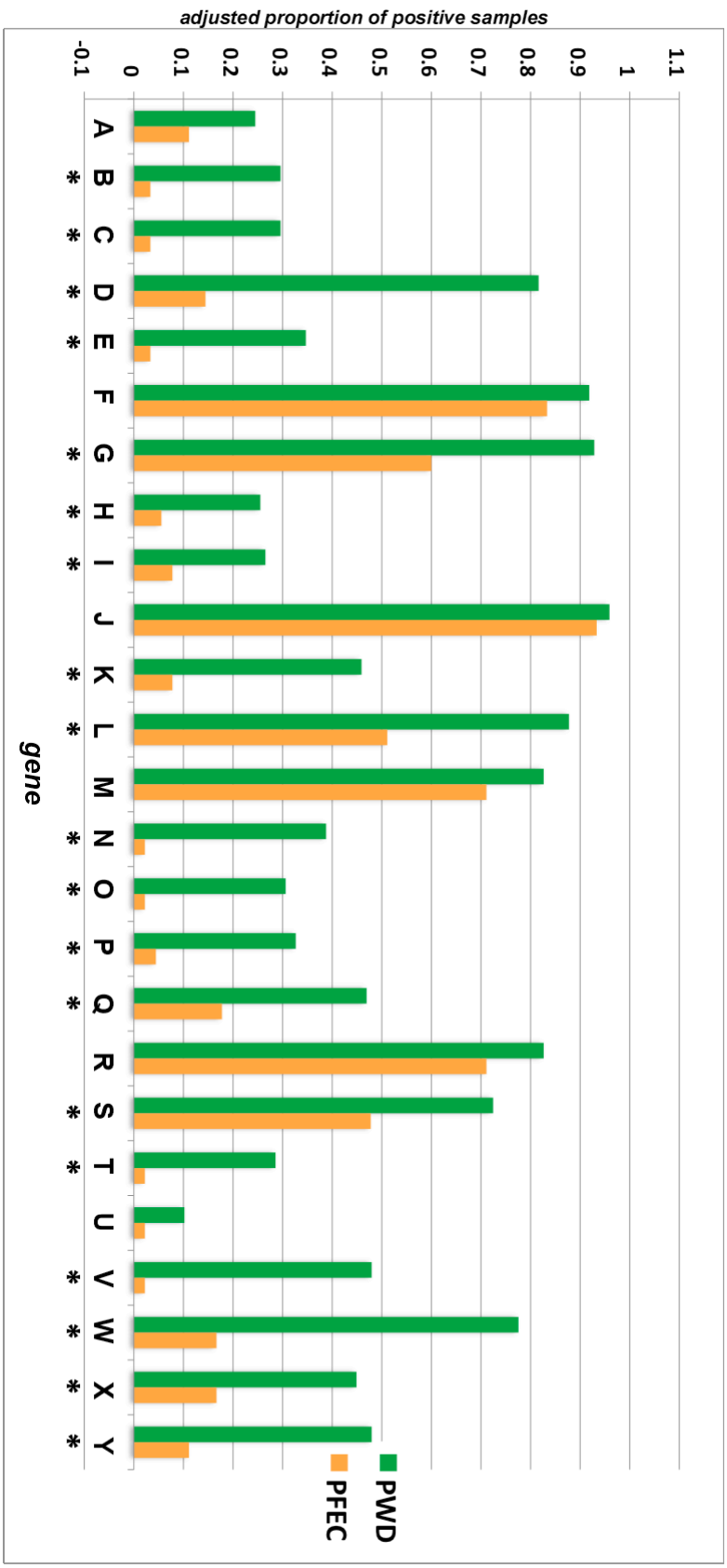
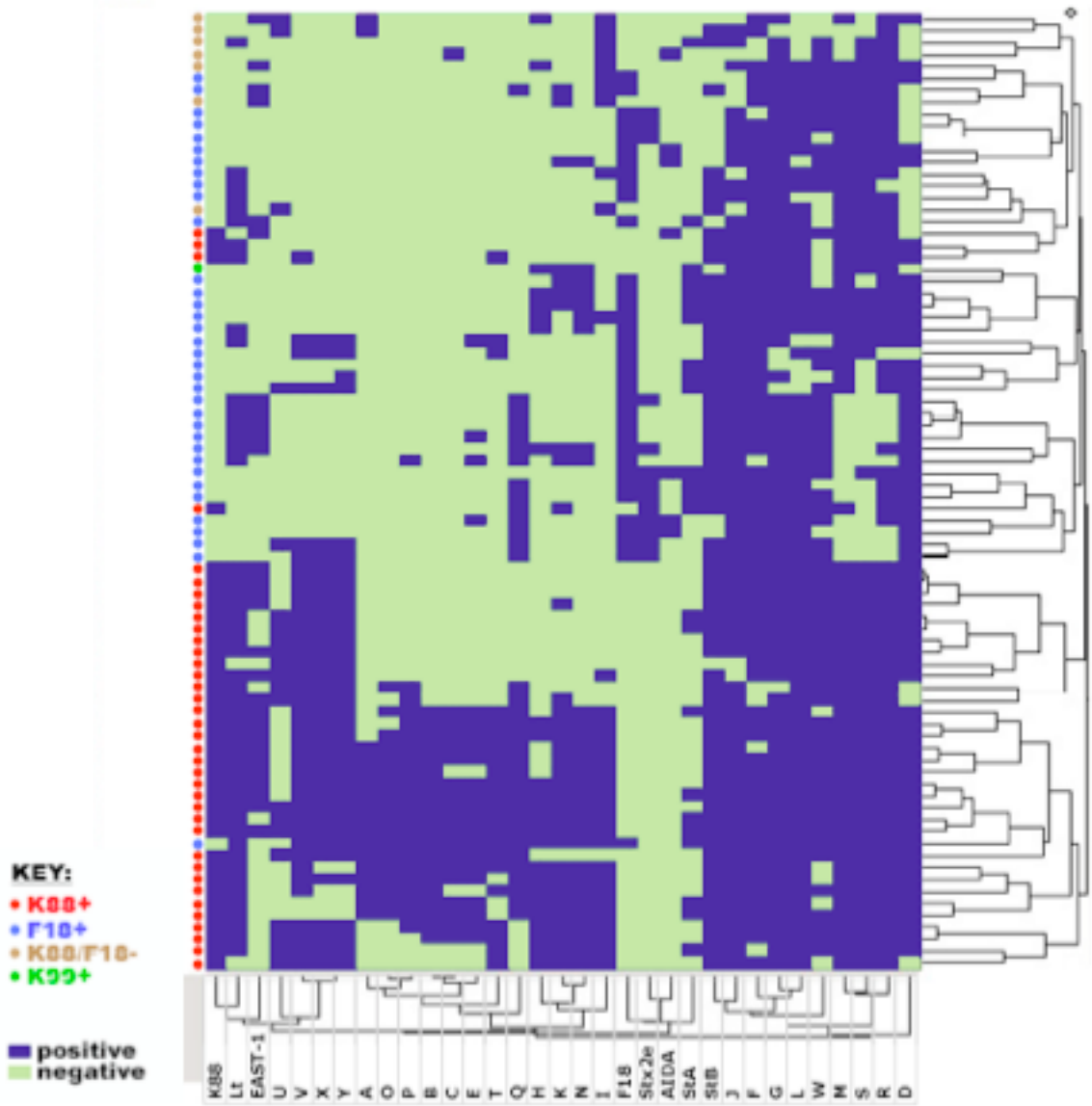

adjusted proportion of positive samples
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
gene
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
*
PFEC
PWD

Figure 9. Two-way hierarchical cluster analysis of porcine ETEC samples, based on ETEC virulence factors and the 25 genes of interest (A-Y). Dark blue bars indicate a positive PCR reaction, and green bars indicate a negative result. Gene letters correspond to the descriptions in Table 3. Color-coded samples are listed on the y-axis, with gene codes and virulence factors on the x-axis.



Based on our cluster analysis, K88-positive samples were associated with genes V, X, or Y (which is not surprising, since these genes are found on the K88 plasmid), as well as A, B, C, E, H, I, K, N, O, P, and T. Patterns in F18-positive strains were less clear, though this is not surprising, since the genes examined were selected from the K88-positive UMNK88 sequence. Many genes were found in both K88-positive and F18-positive strains, including T and W (genes found on the Hly plasmid in UMNK88, though they are hemolysin genes that are not plasmid-specific) as well as D, F, G, J, L, M, R, and S.

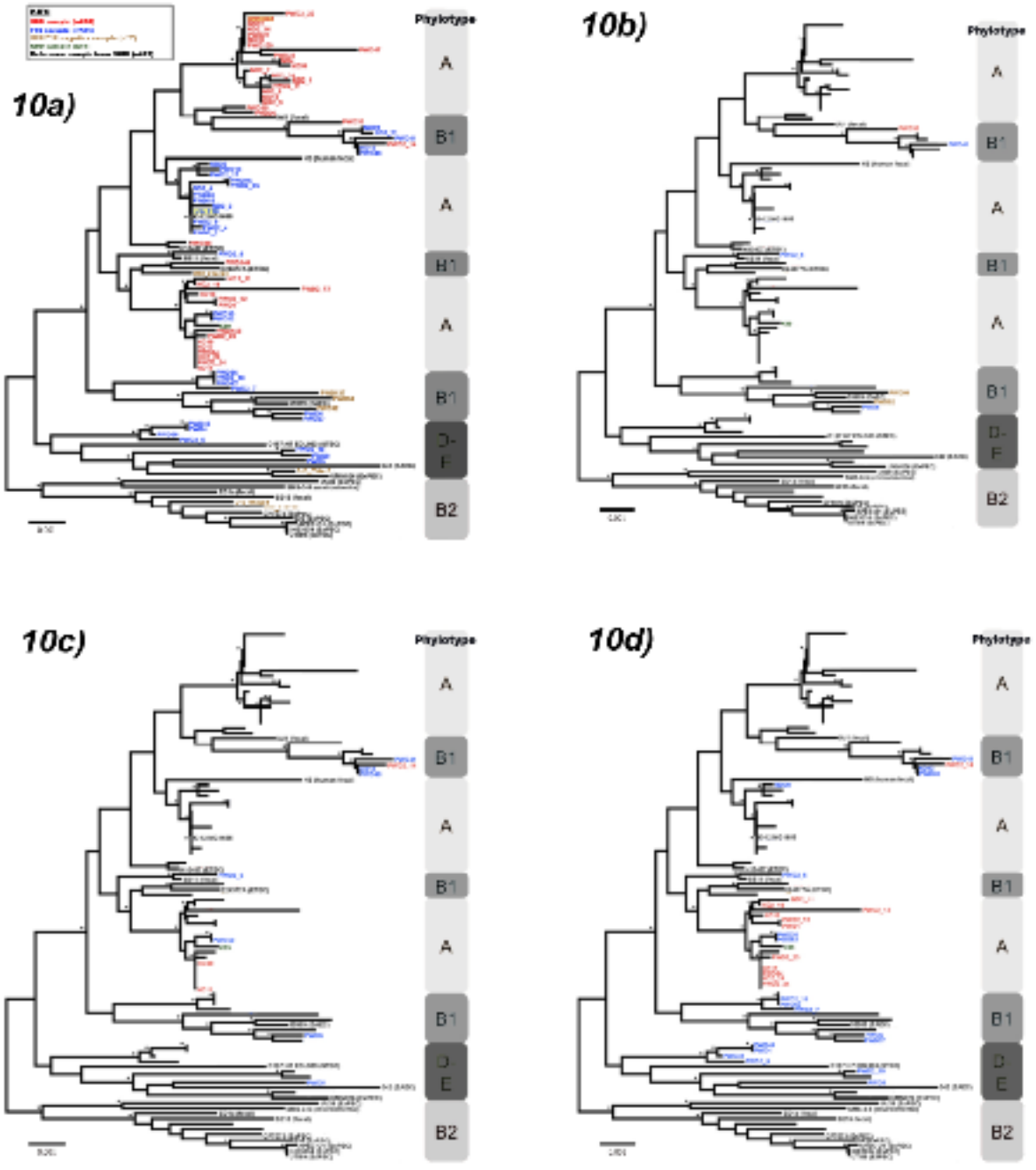
Gene I has an especially interesting pattern in the cluster analysis. It is present in many K88-positive and non-K88/F18 strains, but it is absent from most F18-positive strains. This gene is annotated in UMNK88 as "DNase TatD." This gene is part of the Tat operon, which exports proteins with double arginine signal peptides.[90] A BLASTp search revealed 99-100% identity matches to human ETEC strains (H100407 and E24377A), *Shigella spp.,* and EHEC strains, but there were even more hits to non-pathogenic commensal and lab strains such as *E. coli* DH10ß, a different K-12 strain, IAI1, SE11, and KO11 (an ethanologenic engineered strain). Because of this, it is somewhat surprising that the gene was not found more often in the PFEC samples.

Targeting proteins found in pathogenic bacteria but absent from commensal bacteria could prove effective in vaccine development. Vaccines or other treatments should not eliminate commensal organisms since they are important to the microbiome of the gastrointestinal tract. The loss of commensal organisms can also create an empty niche that may be filled by opportunistic pathogens. For these reasons, genes found in

20% or more of the PFEC samples were excluded from further analysis. Sample data organized according to the remaining 12 genes were analyzed in R to find the combinations of genes such that at least one of the genes would be found in the maximum number of samples.

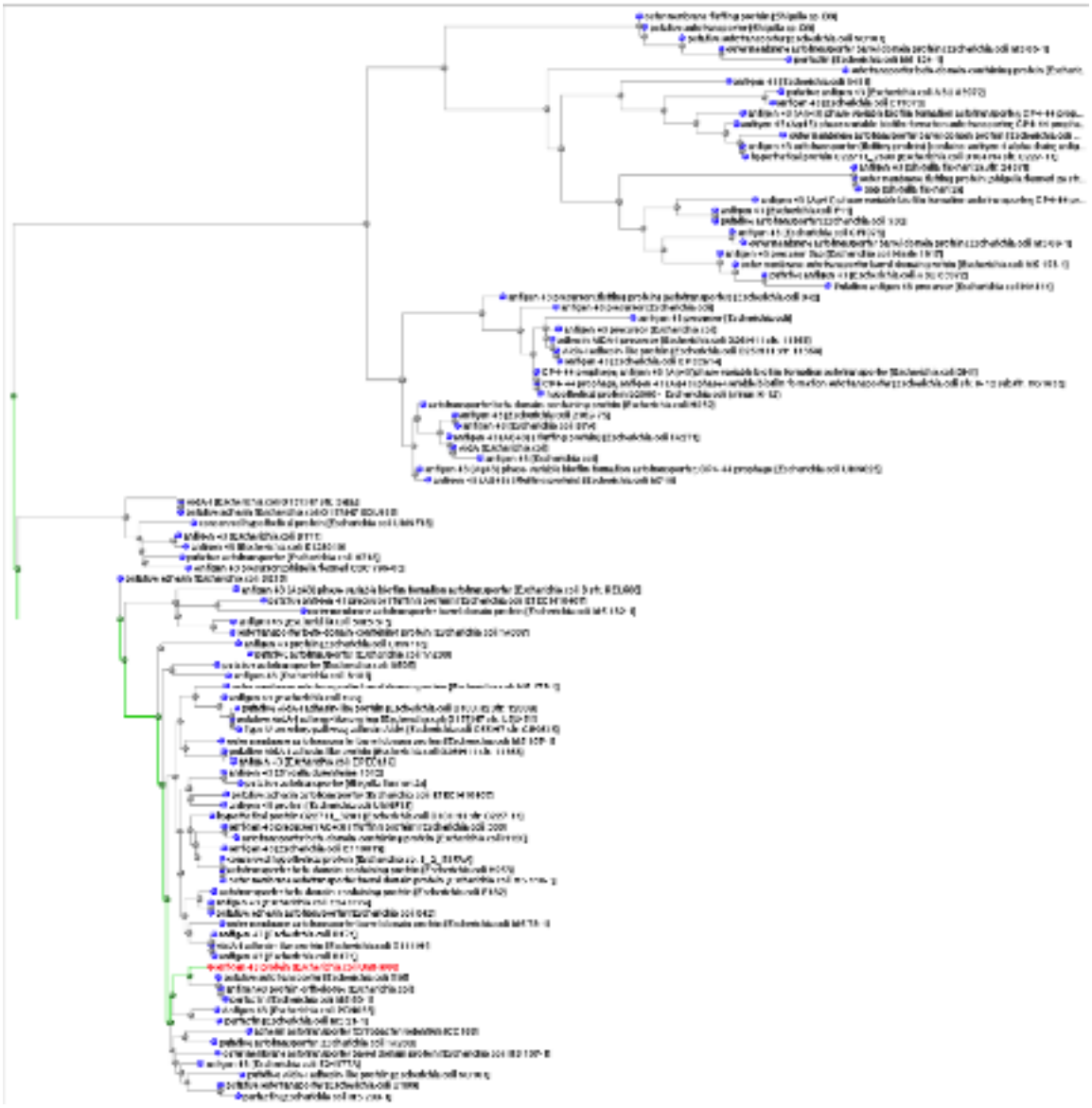
For all of the genes, D and W were found in the most samples: 73 of 80. This means that if a successful vaccine was made that targeted these proteins, disease caused by >90% of our porcine ETEC samples could potentially be prevented. Figure 10b shows the dendrogram from Chapter 2, with all samples positive for D and/or W removed, as we would estimate the coverage of a vaccine created from these two proteins. This combination seems particularly good at eliminating the branches of ETEC in phylotype A.

Figure 10. MLSA dendrogram of porcine ETEC and reference samples. **10a.** Original dendrogram showing all samples, including *E. coli* reference strains (not tested). **10b.** Dendrogram with porcine ETEC samples positive for genes D and/or W removed (73 of 80 porcine ETEC samples eliminated). **10c.** Dendrogram with samples positive for genes D and/or I removed (69 of 80). **10d.** Dendrogram with samples positive for genes I and/or Q removed (49 of 80).



Gene D, annotated in UMNK88 as 'antigen 43 precursor protein,' is also described as an 'autotransporter-like protein of the Type V secretion pathway' and a 'phase-variable biofilm formation protein.'[91] This protein has been implicated in biofilm formation and colonization of the bladder in a mouse model of urinary tract infection.[92] Results of a pBLAST in the NCBI database show 100% identity hits to *Shigella spp.*, several *E. coli* O157:H7 strains (including strain Sakai), and EHEC O26:H11 str.11368, as well as 97% identity hits to multiple EPEC strains. In contrast, the protein match is only 68% identity to strain K-12 MG1655, our non-pathogenic comparison strain. The genomic locations of these proteins in UMNK88 and K-12 do not correspond with each other, further indicating functional difference. All of this points to distinct structure and function of this protein in the pathogenic strains, compared with commensal strains.

Figure 11. Gene code D matches in pBLAST. Neighbor-joining tree of top pBLAST hits against gene D, UMNK88 antigen 43 protein, from the NCBI database.



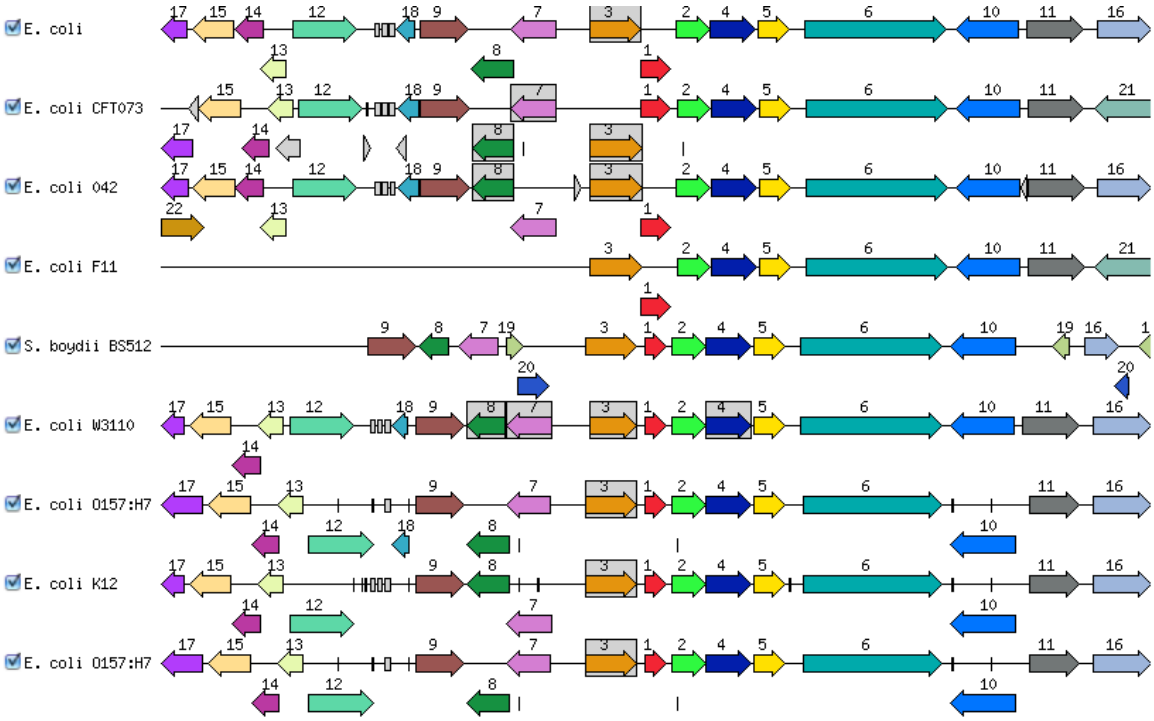
Gene W, annotated in UMNK88 on the hemolysin plasmid as a conserved hypothetical protein, matches many proteins described as IS elements or transposases. 98-100% identity matches are found in human ETEC, EPEC, EIEC, and UPEC strains, as

well as *Citrobacter rodentium*, *Shigella spp.,* and a human commensal fecal *E. coli* (strain SE11).

We also determined which two chromosomal genes would be found (one or both genes) in the ETEC samples, since plasmid carriage can be variable. Here the best combination was genes D and I, both of which have been described above. This combination would eliminate 69 of 80 samples, or about 85%, as seen in Figure 10c. More K88-positive samples were eliminated: 35 of 38, or 92%.

Finally, since antigen 43 does match, albeit weakly, K-12 MG1655, we decided to look for the next best combination of two chromosomal genes. These were genes Q and I, with positives in 49 of 80 samples (61%), as seen in Figure 10d. Gene I has been described above, and may be problematic since there are many strong pBLAST hits to commensal organisms. Gene Q corresponds to a putative inner membrane protein that was predicted by RAST. Later, annotation with Artemis did not call the gene the same way, and now Q (and also P) primers code for an intergenic region. The difference in the prediction by different programs reflects that this process is uncertain, and demonstrates a limitation in interpreting nucleotide sequences without additional confirmation. Nonetheless, a nucleotide BLAST of the Q primers matches putative outer membrane proteins in several *E. coli* strains. In RAST, the corresponding protein predicted in UMNK88 is a putative inner membrane protein, although PSORT predicted that it was an outer membrane protein. In RAST, top matches include a variety of pathogens (UPEC, STEC, *Shigella)* and a different (non-1655) commensal strain K-12.

Figure 12. Gene matches to gene codes P and Q from RAST. UMNK88 is the top line, called *E. coli*. Gene P is represented by the shaded orange arrows labelled '3' and gene Q is represented by the red arrows labelled '1.'



In the R program, code was also run to search for combinations of 3 or 4 genes that would result in more hits against our ETEC samples, but adding more genes to the pairs of genes described above never resulted in more than 1 or 2 additional positive samples. As such, it does not seem useful to elaborate on additional combinations of more genes. We have identified several gene pairs where most porcine ETEC samples are positive for at least one of the genes.

Discussion

As predicted, we were able to identify several genes in porcine ETEC that were differentially present or absent in porcine commensal *E. coli*. These genes were located on the chromosome and on plasmids. This points to diversity in chromosomal background and in plasmid carriage between porcine ETEC and commensal strains. Because we identified these genes based on the surface-accessible locations and structures of their predicted proteins, they may be useful in vaccine design. We have also shown how combinations of genes can be combined to maximize the number of samples that are positive for at least one gene in the combination. Potentially, proteins encoded by these genes could be combined in a multiple subunit vaccine to target the most possible ETEC strains. Before reaching this stage, additional investigations would need to be completed. Protein expression and antigenic potential of the proteins would have to be assessed. Vaccine delivery in a way that would effectively stimulate mucosal immunity, such as using a recombinant attenuated *Salmonella* vaccine vector or cloning the genes of interest into an ingredient in feed, would also need to be developed.[93]

Chapter 4: Resistance and Virulence Plasmid Distribution in Porcine ETEC

Introduction

Finally, we examined plasmid content of porcine ETEC isolates. Plasmids are vital to porcine ETEC for two primary reasons: they encode virulence factors and multidrug resistance. Virulence factors, consisting of colonization factors and enterotoxins, are what define porcine ETEC and give them their ability to cause disease. Antibiotics are the main treatment for ETEC infections, and the vast majority of samples submitted to the University of Minnesota Veterinary Diagnostic Laboratory contain isolates that are multidrug resistant. As such, it was not surprising to find that the K88-positive porcine ETEC strain that we sequenced (as described in the previous chapter) contained three virulence plasmids (Ent, Hly, and K88) and three additional resistance plasmids belonging to the IncA/C, IncI1 and IncN replicon types.

It has been shown *in vitro* that different *E. coli* strains readily acquire ETEC virulence plasmids.[94] However, we determined that a relatively narrow subset of *E. coli* lineages carried these plasmids. It could be thus inferred that something must make certain strains able to uniquely acquire, regulate, and maintain ETEC plasmids. Based on the patterns observed in the phylogenetic tree, indicating chromosomal similarity among ETEC strains, the something may be chromosomal background. To address this question, we hypothesize that ETEC samples belonging to the same phylotype are more likely to carry the same plasmids. In Chapter 2, we found that K88- and F18-positive samples cluster with samples carrying the same virulence (K88 or F18) plasmid type in a phylogenetic tree. Here, we will examine the prevalence of IncA/C and IncI1 resistance plasmids and Ent, Hly, and IncFV virulence plasmids in porcine ETEC samples. It

should be noted that the distinction between resistance and virulence plasmids is often blurred, as the same plasmids often carry both resistance and virulence genes.

The 'resistance' plasmids examined are IncA/C and IncI1. UMNK88 contains a 161-kb IncA/C plasmid whose sequence was previously described.[95] This plasmid type contains several "hot spot" accessory regions and commonly encodes multidrug resistance. The IncA/C plasmid in UMNK88 contains a number of resistance-associated genes, including those encoding resistance to phenicols, sulfonamides, aminoglycosides, and mercury. IncI1 plasmids were found in both UMNK88 (91-kb plasmid) and UMNF18 (70-kb plasmid) and have also been previously described.[96] IncI1 plasmids typically only include a single "hot spot" region, which may include antibiotic resistance modules. The IncI1 plasmid in UMNF18 contains four gene cassettes encoding streptothricin, aminoglycoside, and chloramphenicol resistance. IncI1 plasmids may also play a role in colonization, via a Type IV conjugation pilus that also adheres to epithelial cells and abiotic surfaces, and is thought to contribute to virulence.[97] IncA/C and IncI1 plasmids have been shown to be highly prevalent among disease-associated porcine ETEC, underscoring their importance in the dissemination of multidrug resistance among bacteria of production animals.

The 'virulence' plasmids examined are Ent, Hly, and IncFV. UMNK88 contains an 81-kb Ent plasmid containing heat labile and heat stabile enterotoxin-encoding genes. This is an IncFIB replicon-type plasmid. UMNK88 also contains a 65-kb Hly plasmid of unknown Inc type, which encodes hemolysin genes *hlyCABD*. UMNF18 contains a plasmid called IncFV, a 103-kb IncFIIA replicon-type plasmid that encodes genes encoding multiple colicins and also several antimicrobial resistance genes.

It is important to note that many of the traits and genes described here, which will be used as genetic indicators of plasmid presence, are not plasmid-specific. For instance, hemolysins can be found on various plasmids or may be chromosomally-encoded. For this reason, when using non-plasmid-specific genes, we used multiple genes from the same type of plasmid to infer information about plasmid content. Nevertheless, this information remains an estimation.

Materials and Methods

Strain selection. Porcine ETEC strains described in previous chapters were examined.

Selection of plasmid genes of interest. To investigate plasmids, porcine ETEC strain UMNK88 plasmid sequences described in the previous chapter were utilized. An additional F18-positive porcine ETEC strain, UMNF18, was also sequenced in the method described previously. This strain was isolated from a weaned pig on a farm in Iowa in 2006 and is positive for F18, STa, STb, and Stx2e. Like the K88-positive strain, its antimicrobial susceptibility test reflected that it is multidrug resistant. The F18-strain contained five plasmids. Draft sequence accession numbers for UMNF18 are CP002890 for the chromosome, and CP002891, CP002892, CP002893, CP002894, and CP002895 for UMNF18 plasmids pUMNF18_IncI1, pUMNF18_87 (F18 plasmid), pUMNF18_IncFV, pUMNF18_P7, and pUMNF18_32. From UMNK88 and UMNF18 sequences, the following plasmids types were selected for examination: IncA/C (in UMNK88), IncI1 (in UMNK88 and UMNF18), Ent (in UMNK88), Hly (in UMNK88), and IncFV (in UMNF18). Genes are listed in Table 4.

Primer Design. Primers were designed using PrimerSelect (Lasegene, Madison, WI) and are listed in Table 4.

Gene prevalence determination. Gene prevalence studies were performed on 78 porcine ETEC samples used for MLSA analysis. The PCR thermocycler protocol consisted of 10 minutes at 95°C; 30 cycles of 94°C for 30 seconds, 57°C for 90 seconds, and 72°C for 1 minute; then 7 minutes at 72°C, and holding at 4°C. The PCR reaction contained 2 μl boiled bacterial template, 1.35 μl pooled primer mix, 18.3 μl water, 2.5 μl 10x PCR buffer with MgCl2, 2 μl dNTP, and 0.2 μl Taq polymerase (Go Flexi®, by Promega) for 25 μl total volume used per reaction. After amplification, samples were electrophoresed in 2% agarose gel and run at 200V for 1h, stained with ethidium bromide, and visualized under ultraviolet light to detect PCR amplification products. A 100 bp size standard was used to confirm appropriate PCR product sizes.

Some of the genes examined were not plasmid-specific, therefore the following criteria were used to call an isolate "positive" for a particular plasmid type: a positive PCR reaction for the single gene in IncA/C or IncI1 plasmids; positive PCR reactions for 2 of 3 Ent or Hly plasmid genes; or positive PCR reactions for both IncFV plasmid genes.

Table 4. Plasmid-related genes and primer sequences.

Gene number	Plasmid	Primer: F	Primer: R	Prod Size (bp)	Gene/Product Description
1	IncA/C	GGTGAGTTTCGCAAGAAAGG	CTGTCTGCTGCTTACGCTTG	206	*repA* replication gene
2	IncI1	CGAAAGCCGGACGGCAGAA	TCGTCGTTCCGCCAAGTTCGT	140	*repA* replication gene
3a	Ent	TAAAACGTTCCGGAGGTCTTATGC	CTGCCGGAGCTATATTCAGATTTC	429	heat labile enterotoxin subunit A
3b	Ent	GGAGCTCCCCAGACTATTACAGAA	TTGGGGGTTTTATTATTCCATACA	284	heat labile enterotoxin subunit B
3c	Ent	GCAATCATATCCGTCTCAGGAACA	TGCCAGGCAGCTCCATATTTCTTT	376	putative MarR-family transcriptional regulator
4a	Hly	CTCAGGCGCTGGCAGATGGTCA	ACGCGGGGCGCACTTACTGTG	526	putative invasin
4b	Hly	GGCCGCCCTGCACATCAAAAG	ACATCGGGCTGCCGGGGTAAC	482	DNA primase SogL
4c	Hly	TCACGCCGTCACCGGTAGTGGA	CACGCCGCCAGGCAAACTGAC	270	putative mobilization protein MobA
5a	IncFV	CCGGGCCCTGACGTTCCTGTC	GCGCCCTTCTGCGATTTCACG	394	replication gene (protein RepA1)
5b	IncFV	CTGTTGCGCGGGGTTCAGGAG	GCGGCGTGCGGTACATCTTGTC	334	conjugal transfer protein

Statistical analyses. Adjusted Wald proportions and 95% confidence intervals (reflecting two standard errors above and below the proportion) were determined, as described in Chapter 3. Non-overlapping confidence intervals were interpreted to indicate significant results.

Results

Overall, the plasmids examined were widely distributed throughout our porcine ETEC samples of different lineages. The exception was that Hly and Ent plasmids were not found in any of the phylotype D/E samples. Samples in phylotype A, particularly in the top cluster of the dendrogram, appeared to be more likely to carry the three or more types of the plasmids examined than isolates in phylotype B1 or D/E (with the exception of plasmid IncFV), but most of the differences observed were not statistically significant. Samples in phylotype A were significantly more likely to carry Ent plasmids, as shown in Figure 14. Due to limited sample numbers, particularly in phylotypes B1 and D/E, additional samples would need to be tested in order to attain significant results for other plasmid types.

Figure 13. Resistance and virulence plasmid content. Plasmid content was inferred by PCR results regarding the detection of specific genes, shown in Table 4. Only porcine ETEC isolates (colored names) were evaluated. Strains with names in black are included for reference in the dendrogram structure but were not evaluated for plasmid content.

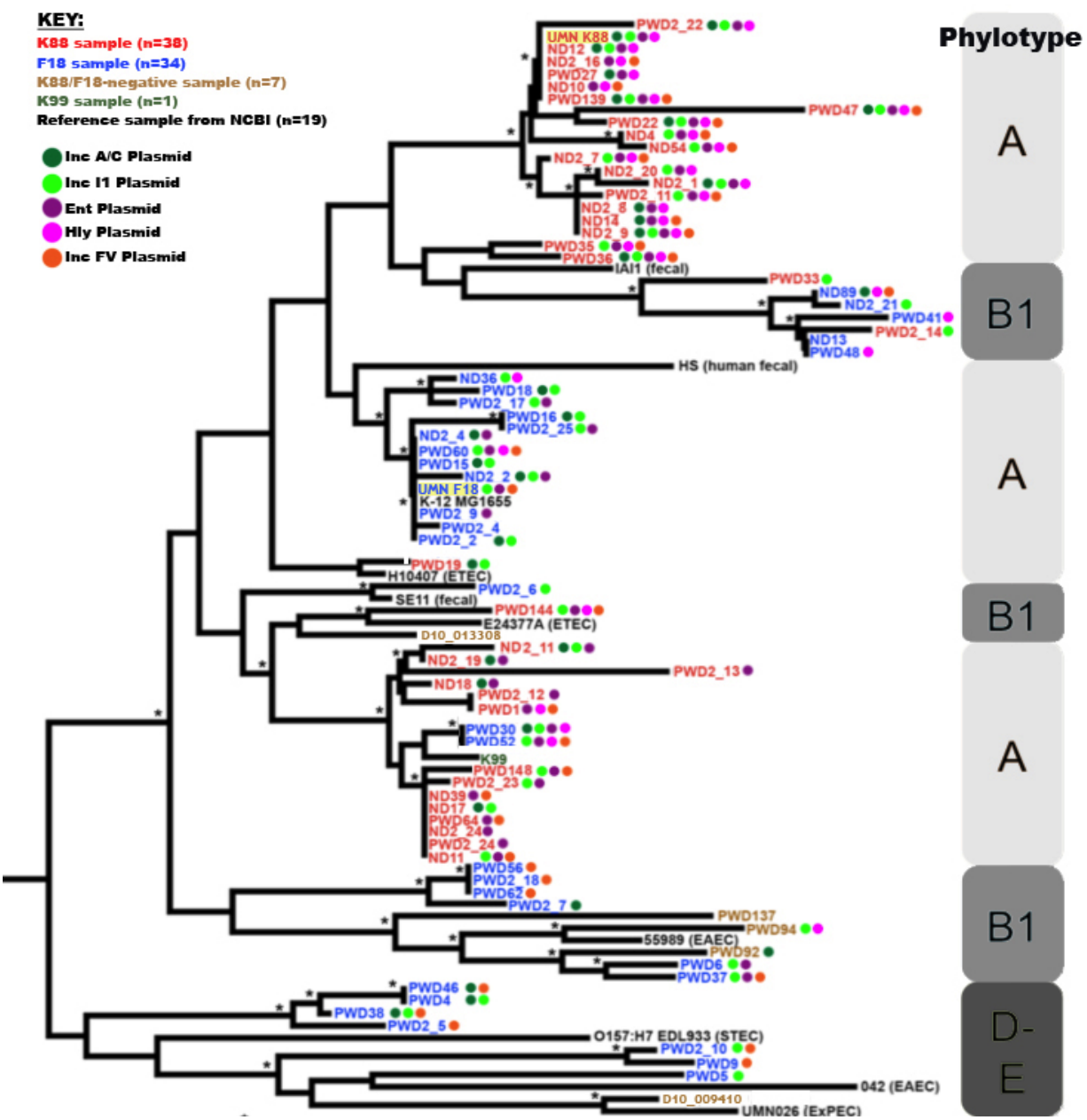
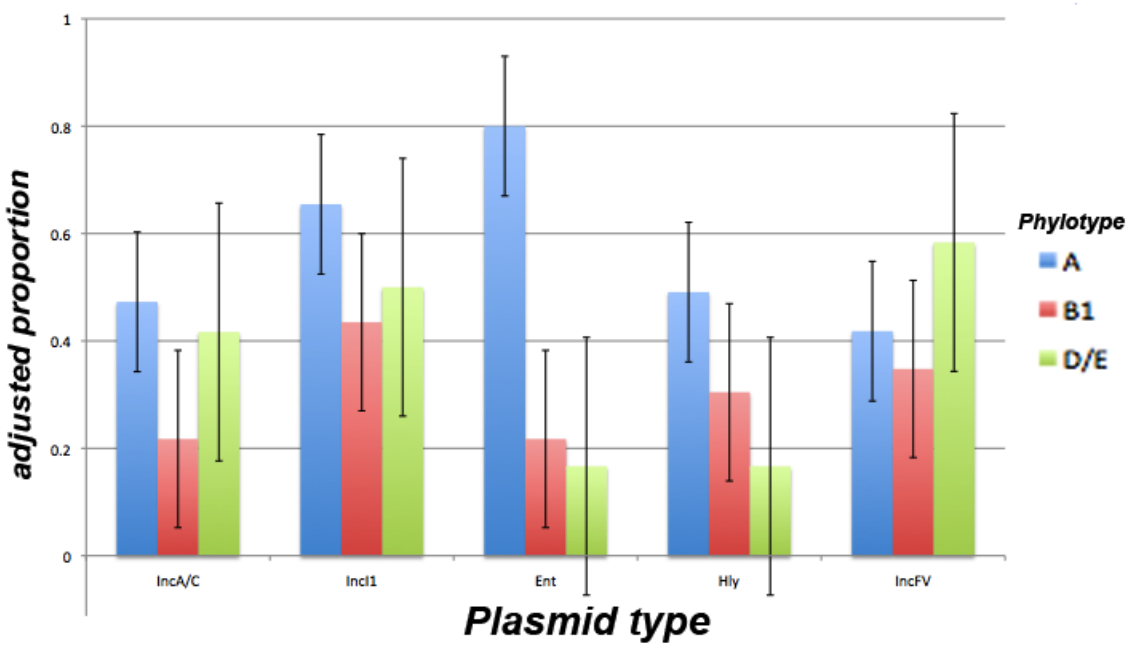


Figure 14. Inferred plasmid content in porcine ETEC isolates. 78 isolates were examined, 51 in phylotype A, 19 in phylotype B, and 8 in phylotype D/E. Plasmid content was inferred based on the presence of 1 to 3 genes per plasmid, as detected by PCR reaction. 95% confidence intervals are indicated by error bars.



Discussion

Based on plasmid-related gene prevalence, the five resistance and virulence plasmid types examined appear to be found in a variety of different porcine ETEC isolates in different phylotypes. Notably, no isolates in phylotype D/E contained Ent or Hly plasmids. These plasmids were sequenced in our UMNK88 strain. They may be associated with K88-positive isolates, in which case their absence from porcine ETEC isolates in phylotype D/E is not surprising since none of these strains are K88-positive.

Due to limited sample size, particularly in phylotypes B1 and D/E, our only statistically significant conclusion regarding specific plasmid distribution is that Ent plasmids are more common in isolates in phylotype A than in the other groups.

The wide distribution of plasmids encoding virulence and resistance traits is interesting, because it appears to go against the prevailing dogma that there is a virulence-resistance trade-off. Specifically, this idea states that virulent isolates are less likely to be resistant, and vice-versa, suggesting that the dangerous combination of highly virulent and resistant bacteria is uncommon.[10] While this appears to be true in certain kinds of *E. coli*, particularly ExPEC isolates, porcine ETEC strains clearly have the ability to harbor multiple virulence- and resistance-encoding plasmids.[98,99] This has serious practical implications, making these potent, disease-causing strains very challenging to treat.

Chapter 5: Conclusions

As hypothesized, porcine ETEC isolates do appear to group together in a phylogenetic tree containing different pathotypes of *E. coli*, though they are still polyphyletic, with multiple clusters representing multiple related lineages. Human ETEC and commensal strains are interspersed with porcine ETEC, which primarily belong to phylotypes A and B1. This suggests that the chromosomes of most porcine ETEC strains more closely resemble the chromosomes of commensal strains than the chromosomes of other *E. coli* pathotypes, and that commensal *E. coli* may have a greater propensity to acquire and maintain ETEC virulence plasmids or might possess a more suitable chromosomal background for virulence to occur. F18-positive strains cluster less tightly than K88-positive strains, and all porcine ETEC found in phylotype D/E were F18-positive, indicating more diversity in F18-positive than K88-positive ETEC strains.

We also identified genes that were present in most porcine ETEC and absent from most commensal porcine *E. coli*. We identified pairs of genes that were present (one or both genes) in up to 92.5% of our porcine ETEC isolates. Due to the way we mined porcine ETEC genomic data to select genes for predicted proteins that are surface-expressed, these predicted proteins have potential for use in a multiple subunit vaccine. Similar "reverse vaccinology" approaches have been successfully used to develop vaccines against *Neisseria* and *Streptococcus spp.*[100101]

We inferred plasmid content in porcine ETEC samples and found that selected virulence and resistance plasmid types appeared to be widespread throughout porcine ETEC samples. Some differences were noted between groups, with porcine ETEC in phylotype A usually containing more plasmids than porcine ETEC in other phylotypes,

but the significance of our results was limited by the small sample size of some groups. Contrary to our hypothesis, we were unable to identify meaningful patterns specific to groups of isolates on this scale.

There are several limitations to the present study. One limitation that was already mentioned was the number of isolates used. Additional patterns or stronger significance would likely occur if more isolates were examined. Also, no commensal strains of *E. coli* isolated from pigs have been sequenced. Comparing the pathogenic porcine ETEC genome to commensal porcine *E. coli* would be more practical than comparing it to a commensal lab strain original isolated from a human, such as K-12, even though the K-12 strain is commonly used for this purpose. Additionally, all of our commensal PFEC strains came from a single site. Screening samples from different sites would provide more reliable results. Finally, our method of investigating plasmid content was imprecise, since many genes detected are not strictly specific to a particular plasmid type. The idea of examining plasmid types may be limited as a whole, due to the dynamic nature of the genetic content of plasmids, and simply focusing on individual plasmid-encoded genes may be more informative.

Going forward, the ETEC-specific gene identification process, using predictive software to identify putative surface-exposed proteins, should be repeated for an F18-positive porcine ETEC strain. As described in Chapter 4, our lab has already sequenced an F18-positive strain that could be used for this purpose (UMNF18 GenBank: AEJ57294.1). The next step in designing a subunit vaccine based on predicted proteins that are found in porcine ETEC would involve investigating protein expression. Then, purified proteins should be investigated for immunogenicity. Sera from pigs that have

recovered from ETEC infection could be used for antigenicity assays by Western blot. Finally, a method of delivery that would effectively stimulate mucosal immunity, such as a recombinant attenuated Salmonella vaccine, would need to be developed.

As more organisms are sequenced, genomic analysis will become an increasingly robust and powerful tool that will continue to elucidate our understanding of porcine ETEC and other pathogens.

References

1. Turner, S.M., Scott-Tucker, A., Cooper, L.M. & Henderson, I.R. Weapons of mass destruction: virulence factors of the global killer enterotoxigenic Escherichia coli. *FEMS Microbiol. Lett* **263**, 10-20 (2006).
2. Beutin, L., Kaulfuss, S., Herold, S., Oswald, E. & Schmidt, H. Genetic Analysis of Enteropathogenic and Enterohemorrhagic Escherichia coli Serogroup O103 Strains by Molecular Typing of Virulence and Housekeeping Genes and Pulsed-Field Gel Electrophoresis. *J Clin Microbiol* **43**, 1552-1563 (2005).
3. Turner, S.M. et al. Phylogenetic Comparisons Reveal Multiple Acquisitions of the Toxin Genes by Enterotoxigenic Escherichia coli Strains of Different Evolutionary Lineages. *J. Clin. Microbiol.* **44**, 4528-4536 (2006).
4. Moura, R.A. et al. Clonal Relationship among Atypical Enteropathogenic Escherichia coli Strains Isolated from Different Animal Species and Humans. *Appl. Environ. Microbiol.* **75**, 7399-7408 (2009).
5. Ramos, J.L., Marqués, S. & Timmis, K.N. TRANSCRIPTIONAL CONTROL OF THE *PSEUDOMONAS* TOL PLASMID CATABOLIC OPERONS IS ACHIEVED THROUGH AN INTERPLAY OF HOST FACTORS AND PLASMID-ENCODED REGULATORS. *Annu. Rev. Microbiol.* **51**, 341-373 (1997).
6. Yun, C.-S. et al. Pmr, a Histone-Like Protein H1 (H-NS) Family Protein Encoded by the IncP-7 Plasmid pCAR1, Is a Key Global Regulator That Alters Host Function. *J. Bacteriol.* **192**, 4720-4731 (2010).
7. Sherley, M., Gordon, D.M. & Collignon, P.J. Species differences in plasmid carriage in the Enterobacteriaceae. *Plasmid* **49**, 79-85 (2003).
8. Fleckenstein, J.M., Lindler, L.E., Elsinghorst, E.A. & Dale, J.B. Identification of a gene within a pathogenicity island of enterotoxigenic Escherichia coli H10407 required for maximal secretion of the heat-labile enterotoxin. *Infect. Immun* **68**, 2766-2774 (2000).
9. M, L., J., T, D., J. A. & E, B., S. *The Dictionary of cell biology*. (Academic Press: London; San Diego, 1995).
10. Robinson, D.A., Feil, E.J. & Falush, D. *Bacterial Population Genetics in Infectious Disease*. (Wiley-Blackwell: 2010).
11. Bruce, A. *Molecular biology of the cell*. (Garland Science: New York, 2002).
12. Kaper, J.B., Nataro, J.P. & Mobley, H.L. Pathogenic Escherichia coli. *Nat. Rev. Microbiol* **2**, 123-140 (2004).
13. Gyles, C.L. *Pathogenesis of bacterial infections in animals*. (Wiley-Blackwell: Ames, Iowa, 2010).
14. Disease Listing, Enterotoxigenic Escherichia coli, Gen | CDC Bacterial, Mycotic Diseases. at <http://www.cdc.gov/ncidod/dbmd/diseaseinfo/etec_g.htm>
15. Qadri, F. et al. Disease burden due to enterotoxigenic Escherichia coli in the first 2 years of life in an urban community in Bangladesh. *Infect. Immun* **75**, 3961-3968 (2007).
16. Novartis Animal Health US, Inc. at <http://www.livestock.novartis.com/diseases_enteric_swine.html>

17. The Merck Veterinary Manual -Enteric Colibacillosis. at <http://www.merckvetmanual.com/mvm/index.jsp?cfile=htm/bc/22305.htm&word=colibacillosis>
18. Straw, B. *Diseases of swine*. (Iowa State University Press: Ames, Iowa, 1999).
19. Studies are trying to answers for Postweaning Diarrhea caused by by enterotoxigenic Escherichia coli (ETEC). at <http://nationalhogfarmer.com/health-diseases/disease-prevention/answers-postweaning-diarrhea/index.html>
20. Fairbrother, J.M., Nadeau, E. & Gyles, C.L. Escherichia coli in postweaning diarrhea in pigs: an update on bacterial types, pathogenesis, and prevention strategies. *Anim Health Res Rev* **6**, 17-39 (2005).
21. Ngeleka, M., Martineau-Doizé, B. & Fairbrother, J.M. Septicemia-inducing Escherichia coli O115:K"V165"F165(1) resists killing by porcine polymorphonuclear leukocytes in vitro: role of F165(1) fimbriae and K"V165" O-antigen capsule. *Infect. Immun* **62**, 398-404 (1994).
22. Nakamine, M. et al. Dual infection with enterotoxigenic Escherichia coli and porcine reproductive and respiratory syndrome virus observed in weaning pigs that died suddenly. *J. Vet. Med. Sci* **60**, 555-561 (1998).
23. Isaacson, R.E., Fusco, P.C., Brinton, C.C. & Moon, H.W. In vitro adhesion of Escherichia coli to porcine small intestinal epithelial cells: pili as adhesive factors. *Infect. Immun* **21**, 392-397 (1978).
24. Dubreuil, J.D. Escherichia coli STb toxin and colibacillosis: knowing is half the battle. *FEMS Microbiol. Lett* **278**, 137-145 (2008).
25. Do, T.N., Wilkie, I., Driesen, S.J., Fahy, V.A. & Trott, D.J. Pathogenicity of Vietnamese enterotoxigenic Escherichia coli strains in colostrum-deprived one-day-old piglets. *Vet. Pathol* **43**, 150-160 (2006).
26. Johnson, T.J. & Nolan, L.K. Pathogenomics of the virulence plasmids of Escherichia coli. *Microbiol. Mol. Biol. Rev* **73**, 750-774 (2009).
27. Shahriar, F., Ngeleka, M., Gordon, J.R. & Simko, E. Identification by mass spectroscopy of F4ac-fimbrial-binding proteins in porcine milk and characterization of lactadherin as an inhibitor of F4ac-positive Escherichia coli attachment to intestinal villi in vitro. *Developmental & Comparative Immunology* **30**, 723-734 (2006).
28. Nagy, B. & Fekete, P.Z. Enterotoxigenic Escherichia coli (ETEC) in farm animals. *Vet. Res* **30**, 259-284 (1999).
29. Mainil, J.G., Jacquemin, E., Pohl, P., Kaeckenbeeck, A. & Benz, I. DNA sequences coding for the F18 fimbriae and AIDA adhesin are localised on the same plasmid in Escherichia coli isolates from piglets. *Vet. Microbiol* **86**, 303-311 (2002).
30. Zhao, L., Chen, X., Xu, X., Song, G. & Liu, X. Analysis of the AIDA-I gene sequence and prevalence in Escherichia coli isolates from pigs with post-weaning diarrhoea and oedema disease. *Vet. J* **180**, 124-129 (2009).
31. Zhang, W. et al. Escherichia coli constructs expressing human or porcine enterotoxins induce identical diarrheal diseases in a piglet infection model. *Appl. Environ. Microbiol* **74**, 5832-5837 (2008).

32. Al-Majali, A.M., Ababneh, M.M., Shorman, M. & Saeed, A.M. Interaction of Escherichia coli heat-stable enterotoxin (STa) with its putative receptor on the intestinal tract of newborn kids. *FEMS Immunol. Med. Microbiol* **49**, 35-40 (2007).
33. Beausoleil, H.E., Labrie, V. & Dubreuil, J.D. Is Escherichia coli STb enterotoxin sufficient to cause pig diarrhea? *Vet. Microbiol* **70**, 281-285 (1999).
34. Ménard, L.-P. & Dubreuil, J.D. Enteroaggregative Escherichia coli heat-stable enterotoxin 1 (EAST1): a new toxin with an old twist. *Crit. Rev. Microbiol* **28**, 43-60 (2002).
35. de Haan, L. et al. Role of GM1 binding in the mucosal immunogenicity and adjuvant activity of the Escherichia coli heat-labile enterotoxin and its B subunit. *Immunology* **94**, 424-430 (1998).
36. Rodighiero, C. et al. Structural basis for the differential toxicity of cholera toxin and Escherichia coli heat-labile enterotoxin. Construction of hybrid toxins identifies the A2-domain as the determinant of differential toxicity. *J. Biol. Chem* **274**, 3962-3969 (1999).
37. Viswanathan, V.K., Hodges, K. & Hecht, G. Enteric infection meets intestinal function: how bacterial pathogens cause diarrhoea. *Nat. Rev. Microbiol* **7**, 110-119 (2009).
38. Ngeleka, M., Pritchard, J., Appleyard, G., Middleton, D.M. & Fairbrother, J.M. Isolation and association of Escherichia coli AIDA-I/STb, rather than EAST1 pathotype, with diarrhea in piglets and antibiotic sensitivity of isolates. *J. Vet. Diagn. Invest* **15**, 242-252 (2003).
39. Zhang, W., Zhao, M., Ruesch, L., Omot, A. & Francis, D. Prevalence of virulence genes in Escherichia coli strains recently isolated from young pigs with diarrhea in the US. *Vet. Microbiol* **123**, 145-152 (2007).
40. Dallas, W.S. Conformity between heat-labile toxin genes from human and porcine enterotoxigenic Escherichia coli. *Infect. Immun* **40**, 647-652 (1983).
41. Moseley, S.L., Hardy, J.W., Hug, M.I., Echeverria, P. & Falkow, S. Isolation and nucleotide sequence determination of a gene encoding a heat-stable enterotoxin of Escherichia coli. *Infect. Immun* **39**, 1167-1174 (1983).
42. Lindenthal, C. & Elsinghorst, E.A. Enterotoxigenic Escherichia coli TibA glycoprotein adheres to human intestine epithelial cells. *Infect. Immun* **69**, 52-57 (2001).
43. Fleckenstein, J.M., Kopecko, D.J., Warren, R.L. & Elsinghorst, E.A. Molecular characterization of the tia invasion locus from enterotoxigenic Escherichia coli. *Infect. Immun* **64**, 2256-2265 (1996).
44. Brown, E.A. & Hardwidge, P.R. Biochemical characterization of the enterotoxigenic Escherichia coli LeoA protein. *Microbiology* **153**, 3776 -3784 (2007).
45. Aarestrup, F.M. *Antimicrobial resistance in bacteria of animal origin*. (ASM Press: 2006).
46. van Beers-Schreurs, H.M., Vellenga, L., Wensing, T. & Breukink, H.J. The pathogenesis of the post-weaning syndrome in weaned piglets: a review. *Vet Q* **14**, 29-34 (1992).
47. Lallès, J.-P., Bosi, P., Smidt, H. & Stokes, C.R. Nutritional management of gut health in pigs around weaning. *Proc Nutr Soc* **66**, 260-268 (2007).

48. Wang, J. et al. Gene expression is altered in piglet small intestine by weaning and dietary glutamine supplementation. *J. Nutr* **138**, 1025-1032 (2008).
49. Ewaschuk, J.B., Murdoch, G.K., Johnson, I.R., Madsen, K.L. & Field, C.J. Glutamine supplementation improves intestinal barrier function in a weaned piglet model of Escherichia coli infection. *Br J Nutr* 1-8 (2011).doi:10.1017/S0007114511001152
50. Owusu-Asiedu, A., Nyachoti, C.M. & Marquardt, R.R. Response of early-weaned pigs to an enterotoxigenic Escherichia coli (K88) challenge when fed diets containing spray-dried porcine plasma or pea protein isolate plus egg yolk antibody, zinc oxide, fumaric acid, or antibiotic. *J. Anim. Sci* **81**, 1790-1798 (2003).
51. Wellock, I.J., Fortomaris, P.D., Houdijk, J.G.M. & Kyriazakis, I. Effect of weaning age, protein nutrition and enterotoxigenic Escherichia coli challenge on the health of newly weaned piglets. *Livestock Science* **108**, 102-105 (2007).
52. Casewell, M., Friis, C., Marco, E., McMullin, P. & Phillips, I. The European ban on growth-promoting antibiotics and emerging consequences for human and animal health. *J. Antimicrob. Chemother* **52**, 159-161 (2003).
53. R, T., Ian *An introduction to veterinary immunology*. (Saunders: Philadelphia, 1982).
54. Porter, P., Noakes, D.E. & Allen, W.D. Secretory IgA and antibodies to Escherichia coli in porcine colostrum and milk and their significance in the alimentary tract of the young pig. *Immunology* **18**, 245-257 (1970).
55. Jacobsen, M. Abstract | Characterisation of five candidate genes within the ETEC F4ab/ac candidate region in pigs. at <http://www.biomedcentral.com/1756-0500/4/225>
56. Zúñiga, A., Yokoyama, H., Albicker-Rippinger, P., Eggenberger, E. & Bertschinger, H.U. Reduced intestinal colonisation with F18-positive enterotoxigenic Escherichia coli in weaned pigs fed chicken egg antibody against the fimbriae. *FEMS Immunol. Med. Microbiol* **18**, 153-161 (1997).
57. Verdonck, F. et al. Conserved regions in the sequence of the F4 (K88) fimbrial adhesin FaeG suggest a donor strand mechanism in F4 assembly. *Vet. Microbiol* **102**, 215-225 (2004).
58. Smeds, A. et al. Mapping the binding domain of the F18 fimbrial adhesin. *Infect. Immun* **71**, 2163-2172 (2003).
59. Shen, H. et al. Immunogenicity of recombinant F4 (K88) fimbrial adhesin FaeG expressed in tobacco chloroplast. *Acta Biochim. Biophys. Sin. (Shanghai)* **42**, 558-567 (2010).
60. Joensuu, J.J. et al. Glycosylated F4 (K88) fimbrial adhesin FaeG expressed in barley endosperm induces ETEC-neutralizing antibodies in mice. *Transgenic Res* **15**, 359-373 (2006).
61. Zhang, C. & Zhang, W. Escherichia coli K88ac fimbriae expressing heat-labile and heat-stable (STa) toxin epitopes elicit antibodies that neutralize cholera toxin and STa toxin and inhibit adherence of K88ac fimbrial E. coli. *Clin. Vaccine Immunol* **17**, 1859-1867 (2010).
62. Escobar-Páramo, P. et al. A specific genetic background is required for acquisition and expression of virulence factors in Escherichia coli. *Mol. Biol. Evol* **21**, 1085-1094 (2004).

63. De Gelder, L., Williams, J.J., Ponciano, J.M., Sota, M. & Top, E.M. Adaptive plasmid evolution results in host-range expansion of a broad-host-range plasmid. *Genetics* **178**, 2179-2190 (2008).
64. Steinsland, H., Lacher, D.W., Sommerfelt, H. & Whittam, T.S. Ancestral lineages of human enterotoxigenic Escherichia coli. *J. Clin. Microbiol* **48**, 2916-2924 (2010).
65. Owen, R.J. Bacterial taxonomics: finding the wood through the phylogenetic trees. *Methods Mol. Biol* **266**, 353-383 (2004).
66. Crossman, L.C. et al. A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic Escherichia coli strain H10407. *J. Bacteriol* **192**, 5822-5831 (2010).
67. Sahl, J.W. et al. A comparative genomic analysis of diverse clonal types of enterotoxigenic Escherichia coli reveals pathovar-specific conservation. *Infect. Immun* **79**, 950-960 (2011).
68. Bacterial PCR - CVM - VDL, University of Minnesota. at <http://www.vdl.umn.edu/ourservices/molecbacti/index.htm#ecoli>
69. Maximum Productivity - Managing Pig Health and Treating Pig Dieases on ThePigSite.com. at <http://www.thepigsite.com/pighealth/article/306/maximum-productivity>
70. Wirth, T. et al. Sex and virulence in Escherichia coli: an evolutionary perspective. *Mol. Microbiol* **60**, 1136-1151 (2006).
71. Kumar, S., Nei, M., Dudley, J. & Tamura, K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinformatics* **9**, 299-306 (2008).
72. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol* **4**, 406-425 (1987).
73. Clermont, O., Bonacorsi, S. & Bingen, E. Rapid and simple determination of the Escherichia coli phylogenetic group. *Appl. Environ. Microbiol* **66**, 4555-4558 (2000).
74. Carlos, C. et al. Escherichia coli phylogenetic group determination and its application in the identification of the major animal source of fecal contamination. *BMC Microbiol* **10**, 161 (2010).
75. Aziz, R.K. et al. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**, 75 (2008).
76. Rutherford, K. et al. Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944-945 (2000).
77. Frishman, D., Mironov, A., Mewes, H.W. & Gelfand, M. Combining diverse evidence for gene recognition in completely sequenced bacterial genomes. *Nucleic Acids Res* **26**, 2941-2947 (1998).
78. Delcher, A.L., Harmon, D., Kasif, S., White, O. & Salzberg, S.L. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* **27**, 4636-4641 (1999).
79. Delcher, A.L., Bratke, K.A., Powers, E.C. & Salzberg, S.L. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**, 673-679 (2007).
80. BLAST: Basic Local Alignment Search Tool. at <http://blast.ncbi.nlm.nih.gov/Blast.cgi>
81. Pati, A. et al. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat. Methods* **7**, 455-457 (2010).

82. Yu, N.Y. et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**, 1608-1615 (2010).
83. Bendtsen, J.D., Nielsen, H., von Heijne, G. & Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol* **340**, 783-795 (2004).
84. Rahman, O., Cummings, S.P., Harrington, D.J. & Sutcliffe, I.C. Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria. *World J Microbiol Biotechnol* **24**, 2377-2382 (2008).
85. Berven, F.S., Flikka, K., Jensen, H.B. & Eidhammer, I. BOMP: a program to predict integral -barrel outer membrane proteins encoded within genomes of Gram-negative bacteria. *Nucleic Acids Research* **32**, W394-W399 (2004).
86. Serruto, D., Rappuoli, R. & Pizza, M. Meningococcus B: From Genome to Vaccine. 183-204doi:10.1002/0470012536.ch8
87. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol* **132**, 365-386 (2000).
88. Agresti, A. & Coull, B.A. Approximate Is Better than "Exact" for Interval Estimation of Binomial Proportions. *The American Statistician* **52**, 119-126 (1998).
89. The R Project for Statistical Computing. at <http://www.r-project.org/>
90. Wexler, M. et al. TatD Is a Cytoplasmic Protein with DNase Activity. *Journal of Biological Chemistry* **275**, 16717 -16722 (2000).
91. van der Woude, M.W. & Henderson, I.R. Regulation and function of Ag43 (flu). *Annu. Rev. Microbiol* **62**, 153-169 (2008).
92. Ulett, G.C. et al. Functional analysis of antigen 43 in uropathogenic Escherichia coli reveals a role in long-term persistence in the urinary tract. *Infect. Immun* **75**, 3233-3244 (2007).
93. Gu, J. et al. Vaccination of attenuated EIS-producing Salmonella induces protective immunity against enterohemorrhagic Escherichia coli in mice. *Vaccine* (2011).doi:10.1016/j.vaccine.2011.07.069
94. Smith, H.W. & Linggood, M.A. Observations on the pathogenic properties of the K88, Hly and Ent plasmids of Escherichia coli with particular reference to porcine diarrhoea. *J. Med. Microbiol* **4**, 467-485 (1971).
95. Fernandez-Alarcon, C., Singer, R.S. & Johnson, T.J. Comparative genomics of multidrug resistance-encoding IncA/C plasmids from commensal and pathogenic Escherichia coli from multiple animal sources. PLOS ONE, in press.
96. Johnson, T.J. et al. Comparative Genomics and Phylogeny of the IncI1 Plasmids, a Common Plasmid Type Among Porcine Enterotoxigenic Escherichia coli. *Plasmid, in press.*
97. Dudley, E.G. et al. An IncI1 Plasmid Contributes to the Adherence of the Atypical Enteroaggregative Escherichia coli Strain C1096 to Cultured Cells and Abiotic Surfaces. *Infect. Immun.* **74**, 2102-2114 (2006).
98. Horcajada, J.P. et al. Quinolone-resistant uropathogenic Escherichia coli strains from phylogenetic group B2 have fewer virulence factors than their susceptible counterparts. *J. Clin. Microbiol* **43**, 2962-2964 (2005).
99. Soto, S.M., Jimenez de Anta, M.T. & Vila, J. Quinolones induce partial or total loss of pathogenicity islands in uropathogenic Escherichia coli by SOS-dependent or -

independent pathways, respectively. *Antimicrob. Agents Chemother* **50**, 649-653 (2006).
100. Moriel, D.G. et al. Identification of protective and broadly conserved vaccine antigens from the genome of extraintestinal pathogenic Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A* **107**, 9072-9077 (2010).
101. Musser, J.M. The next chapter in reverse vaccinology. *Nat. Biotechnol* **24**, 157-158 (2006).