IDENTIFICATION OF GENOMIC DIFFERENCES BETWEEN *ESCHERICHIA COLI* STRAINS PATHOGENIC FOR POULTRY AND *E. COLI* K-12 MG1655

A Thesis

Submitted to the Faculty of Graduate Studies and Research

in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

in the Department of Veterinary Microbiology

University of Saskatchewan

By

Stacy Lynn Stocki

Saskatoon, Saskatchewan, Canada

Spring, 2003

In presenting this thesis in partial fulfillment of the requirements for a Degree of Doctor of Philosophy from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying, or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Veterinary Microbiology

University of Saskatchewan

Saskatoon, Saskatchewan (S7N 5B4)

ABSTRACT

Diseases of poultry caused by *Escherichia coli* result in significant economic losses every year. Specific virulence factors associated with *E. coli* strains pathogenic for poultry have been identified, but it is likely that others remain to be identified. The objective of this project was to identify novel virulence factors of avian *E. coli*. The signature-tagged mutagenesis technique was to be used for this purpose. Difficulties were encountered using that technique, and it was decided that suppression subtractive hybridization should be used to identify genomic differences between *E. coli* K-12 and avian *E. coli* pathogenic for poultry. A total of 62 fragments specific to two avian strains were identified. They were composed of sequences with homology to four types of fragments: plasmid sequences, phage sequences, sequences with known function and sequences without any currently known function. Two *E. coli* collections, a reference collection of diverse strains (ECOR) and a collection of 41 avian isolates, were screened for the presence of 25 of the 62 fragments. Eleven fragments were present in significantly more of the avian strains than of the ECOR strains. Some of the fragments were further characterized by sequencing the DNA flanking the fragments, and analyzing the sequences for the putative genes. One of the fragments was disrupted in the genome of an avian *E. coli* strain. The disruption mutant did not have reduced virulence in two animal models tested, which indicates that the disrupted gene may not be required for virulence. The role of the other fragment sequences in the virulence of avian *E. coli* has not yet been determined.

ACKNOWLEDGMENTS

First, I would like to thanks my supervisors, Drs. Brenda J. Allan and Lorne A. Babiuk who gave me the opportunity to work at VIDO. I will never forget their understanding and kindness. Without their guidance and tutelage this work would not have been possible.

I would like to thank Drs. J. Gordon, A. Potter, H. Deneer, and H. Goldie, for being my committee members and for providing valuable suggestions concerning this project. My thanks are also given to Dr. Carlton Gyles for being my external examiner.

I would also like to thank all of the people working at VIDO, for making my time at VIDO a pleasant experience and for all the help that they provided during the work on this degree. I would especially like to thank Dr. Kingsley Amoako, Tracy Prysliak, Neil Rawlyk, and Rachelle Buchannan for their friendship and expert help with this work. I would also like to thank all of the members of Bacteriology past and present, who made work an enjoyable experience.

I would like to thank my fellow graduate students for their help and support during this work, especially David Schneberger, Valaria Alcorn, Dorothy Huang, Sanipa Suradhat, Zsuzsa Papp and Cam Raggo for the much needed support, encouragement and "vent" sessions.

I would like to thank my parents. Without their love and support, this thesis would not have been possible.

This work was supported by the Canadian Bacterial Disease Network, and by the Natural Sciences and Engineering Research Council of Canada.

TABLE OF CONTENTS

List of Figures

List of Tables

List of Abbreviations

Amp	Ampicillin
BHI	Brain Heart Infusion broth
CFU	Colony forming units
CIP	Calf intestinal alkaline phosphatase
Cm	Chloramphenicol
CTAB	Hexadecyltrimethyl ammonium bromide
DAEC	Diffuse adherent *Escherichia. coli*
dATP	2'-Deoxyadenosine 5'-triphosphate
dCTP	2'-Deoxycytidine 5'-triphosphate
DFI	Differential fluorescence induction
dGTP	2`-Deoxyguanosine 5`-triphosphate
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
dTTP	Thymidine 5'-triphosphate
EAggEC	Enteroaggregative *Escherichia coli*
ECOR	*Escherichia coli* collection of reference
EDDA	ethylenediamine-di-(o-hydroxyphenyl acetic acid)
EHEC	Enterhemorrhagic *Escherichia coli*
EIEC	Enteroinvasive *Escherichia coli*
EPEC	Enteropathogenic *Escherichia coli*
ET	Electorphoretic type
ETEC	Enterotoxigenic *Escherichia coli*
FACS	Fluorescence-activated cell sorter
GAMBIT	Genomic analysis and mapping by *in vitro* transposition
HEV	Hemorrhagic enteritis virus
IVET	*In vivo* expression technology
Km	Kanamycin
LB	Luria-Bertani media
LEE	Locus of enterocyte effacement
LPS	Lippopolysaccharide
LT	Heat labile toxin
OPA	One-Phor-All plus buffer
ORF	Open reading frame
MLEE	Multilocus enzyme electrophoresis
NTPs	Deoxynucleotide triphospates
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
RDA	Representational differences analysis
Rif	Rifampacin
RIVET	Recombinase *In vivo* expression technology
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulfate
SSH	Suppression subtractive hybridization
ST	Heat stable toxin
SLT	Shiga-like toxins

List of Abbreviations Continued

STM	Signature-tagged mutagenesis
Tc	Tetracycline
TE	TRIS and EDTA buffer
TRIS	Tris(hydroxymethyl)aminomethane carbonate
UPEC	Uropathogenic *Escherichia coli*
UTI	Urinary tract infection

1 GENERAL INTRODUCTION AND OBJECTIVES

Escherichia coli is a diverse species. Some *E. coli* strains are part of the natural intestinal flora of healthy humans and animals, while others may cause disease. Diseases of poultry caused by *E. coli* result in millions of dollars of losses each year in the poultry industry. The *E. coli* strains that infect poultry are opportunistic pathogens that infect extraintestinal sites following exposure of birds to stress.

E. coli strains pathogenic for poultry are diverse with respect to their range of virulence. In avian pathogenic *E. coli* strains, similar to other bacterial pathogens, virulence is multifactorial and is associated with adherence factors (F1 and P-pili, and curli), the aerobactin iron-sequestering system, serum resistance, capsule production, and temperature sensitive hemagglutination (*tsh*). However, not all of these factors are found in all of the virulent *E. coli* strains. This fact suggests that there may be additional genes involved in virulence that have not yet been identified.

The goal of this work was to identify new virulence factors found in *E. coli* strains pathogenic for poultry. Two techniques were used in this study:

1) Signature-tagged mutagenesis, which was used to identify genes required for virulence.

2) Suppression subtractive hybridization, which was used to identify sequences found in *E. coli* pathogenic for poultry but not in *E. coli* K-12. The DNA sequences were analyzed to determine which may be potential virulence factors.

The use of the initial approach, signature-tagged mutagenesis, was not successful in identifying virulence-related factors in *E. coli* strains pathogenic for poultry.

We investigated the genomic differences between stains of *E. coli* capable of causing disease in poultry and non-pathogenic *E. coli* as an alternate approach. Suppression subtractive hybridization was used to compare the genomes of two avian *E. coli* strains with the genome of *E. coli* K-12. Sixty-two DNA fragments that were found in avian *E. coli* strains but not in *E. coli* K-12 were identified. As the fragments identified using the suppression subtractive hybridization technique were restriction fragments, and were likely to be portions of larger sequences found in the avian strains but not found in *E. coli* K-12. The sequences of these fragments were analyzed for homologies to sequences in the GenBank database, and four general types of sequence were found. These included: phage sequences, plasmid sequences, sequences with known function and sequences with unknown function. The distribution of 25 of the fragments in two *E. coli* collections was determined. The strain collections used were a group of isolates from extraintestinal infections of poultry and the *E. coli* collection of reference (ECOR). The ECOR collection was assembled to represent the genetic diversity of *E. coli* as a species. All of the fragments tested were in some strains of both collections. Of the 25 selected fragments, eleven of the fragments were found more often in the avian *E. coli* strains than in the ECOR strains, which suggested that they might function in the virulence of avian *E. coli* strains.

The regions flanking three of the 25 selected fragments were sequenced. The flanking sequences were analyzed for the presence of open reading frames (ORFs) and for homology to known sequences. One of the regions was part of a heme utilization gene cluster. Another was homologous to the locus of enterocyte attachment (LEE) family of pathogenicity islands. The function of the third region could not be predicted.

To further investigate the role of the fragments in the virulence of *E. coli* strains that were pathogenic for poultry, two of the selected 25 fragments were disrupted in the genome of an avian *E. coli* strain. One of the mutants had more than one copy of the fragment in the genome and only one of the copies was disrupted. That mutant, therefore, was not studied further. The ORF which was disrupted, was shown to be expressed *in vivo*. The virulence of the mutant strain was tested in two poultry models of infection, and was found to be unattenuated. This result indicated that the fragment did not have a role in the virulence of the avian *E. coli* strain in the models tested. Future directions for this project will be to identify the role of the other subtractive hybridization fragments in virulence of the *E. coli* strains pathogenic for poultry.

2 LITERATURE REVIEW

2.1 Characteristics and Classification of *Escherichia coli*

Escherichia coli is a member of the family Enterobacteriaceae, which are facultative anaerobes, exhibiting both fermentive and respiratory metabolism. *E. coli* are Gram-negative bacilli that grow singly or in pairs in liquid media. *Escherichia* species are found as part of the normal flora of the colon of warm-blooded animals and cockroaches (Holt, Krieg *et al.*, 1994), but can also be found in water, soil, or other environments because of fecal contamination (Sussman, 1997). They are 1.1-1.5 nm in width and 2.0-6.0 nm in length and can be either non-motile or motile with peritrichous flagella. *E. coli* strains can vary biochemically, so there is not one specific test that can be used to differentiate atypical *E. coli* strains from other Enterobacteriaceae (Holt, Krieg *et al.*, 1994). *E. coli* typically are negative for oxidase, urea hydrolysis, lipase, and H_2S production, but are positive for the catalase, citrate, and methyl red tests (Holt, Krieg *et al.*, 1994). Methyl red indicates the strain is capable of forming a stable acid after mixed acid glucose fermentation (Leboffe and Pierce, 1999).

E. coli strains can be classified by surface antigens or by pathotype. The surface antigens used for classification are somatic (O), capsular (K), or flagellar antigens (H) (Kauffmann, 1944). There are at least 167 different O antigens, 74 K antigens and 53 H

antigens identified for *E. coli* (Lior, 1994). Some serotypes have been strongly associated with disease. *E. coli* O157:H7 is one of the major serotypes of enterohemorrhagic *E. coli* (EHEC) (Burland, Shoa *et al.*, 1998). Most common serotypes associated with disease of poultry are O1, O2 and O78, although strains with at least 74 of the known O antigens have shown to be pathogenic for poultry (Gross, 1994). Many of the *E. coli* strains pathogenic for poultry are of unknown serotype and are classified as untypeable (Allan, van den Hurk *et al.*, 1993). K1 capsular antigens have been associated with highly virulent avian *E. coli* strains (Gross, 1994) and with *E. coli* which cause extraintestinal infections in humans (Nataro and Levine, 1994).

Pathogenic strains of *E. coli* have also been classified by pathotype, which is defined as "a classification of *E. coli* into groups that have a similar mode of pathogenesis and cause clinically similar forms of disease" (Donneneberg and Whittam, 2001). In human disease there are eight recognized pathotypes of *E. coli*: enterotoxigenic (ETEC), enteroinvasive (EIEC), enteropathogenic (EPEC), enterhemorrhagic (EHEC), enteroaggregative (EAggEC), diffuse adherent (DAEC), uropathogenic (UPEC) and meningitis-associated. Six of the pathotypes described cause diarrhea in humans, and two pathotypes cause extraintestinal infections. Some of the virulence factors used to classify *E. coli* strains as different pathotypes are adherence factors, toxins, lipopolysaccharide, and the ability to invade host cells.

E. coli also causes disease in other animals including diarrhea and systemic disease in piglets, lambs, and calves, mastitis in cattle, and a variety of diseases of poultry.

2.2 Diseases Caused by *E. coli*

2.2.1 Diseases of Poultry

E. coli can cause a number of disease syndromes in poultry. The general term for *E. coli* infection of poultry is colibacillosis (Barnes, 1994). Forms of colibacillosis include: respiratory disease, cellulitis, colisepticaemia, air sacculitis, pericarditis, synovitis, salpingitis, osteomelitis, peritonitis, omphalitis, and yolk sac infection. The form of colibacillosis caused by the *E. coli* infection depends on the route of infection, and the age of the bird.

E. coli infections of poultry cause large economic losses in the poultry industry. Two of the most economically important disease of poultry are respiratory colibacillosis and cellulitis. A study showed that the most common reason for broiler carcass rejection due to disease in 1992 was colisepticaemia (Yogaratnam, 1995), which can follow respiratory colibacillosis (Barnes, 1994). Another study determined that the second most reported disease by turkey producers was colibacillosis (Christiansen, Hird *et al.*, 1996). Economic losses due to cellulitis alone may exceed $40 million per year in the United States (Norton, 1997).

E. coli infections of poultry are generally secondary infections, a number of viruses, bacteria, and parasites are known to increase the susceptibility of the birds to *E. coli*. Other factors which increase the risk of infection by *E. coli* include environmental conditions such as poor litter quality, water contamination, low protein feed, overcrowding, dust, ammonia and temperature extremes (Barnes and Gross, 1997).

Vaccination with live attenuated viruses can also increase susceptibility of the birds to *E. coli* infection (Gross, 1994).

Respiratory colibacillosis is an example of a disease of poultry caused by *E. coli* as a secondary infection. Infection with Newcastle disease virus, infectious bronchitis virus, *Mycoplasma gallisepticum*, or a vaccine of either virus predisposes chickens to *E. coli* infection (Gross, 1994). Dust and ammonia can also cause respiratory damage and cause the birds to become susceptible to respiratory colibacillosis. *E. coli* strains that are pathogenic can be found in the intestinal tract of most poultry, and are shed into the environment, thus the source usually is inhalation of dust contaminated with feces (Gross, 1994). In broilers, mortality due to respiratory colibacillosis occurs most often between 4-9 weeks of age (Barnes, 1994). It is characterized by fever, air sacculitis, pericarditis, and other systemic lesions (Gross, 1994). Respiratory colibacillosis is a major cause of mortality, reduced feed conversion, and condemnation at time of slaughter for turkeys and broiler chickens (Gross, 1994). Respiratory colibacillosis can lead to other disease manifestations in poultry, including bacteraemia, and salpingitis.

Respiratory colibacillosis can also occur in turkeys, usually following infection with hemorrhagic enteritis virus (HEV) (Gross, 1994). Previous infection with HEV renders birds susceptible to a secondary infection by *E. coli*. *E. coli* from the lungs infect the pericardium and myocardium causing reduced blood pressure (Barnes and Gross, 1997), which in turn can lead to fibrinous perihepatitis and pericarditis. Arthritis may also occur (Gross, 1994).

Salpingitis which is an infection of the oviducts in hens can lead to swelling of the oviduct, preventing the passage of eggs (Barnes, 1994). Occasionally ovum pass into the peritoneum. If *E. coli* in the oviduct is carried with the ova it will infect the yolk

and cause severe peritonitis and death (Gross, 1994). *E. coli* infection of the peritoneum without the ovum is generally controlled and cleared. Hens can become infected when a respiratory infection spreads from the left greater abdominal air sac to the oviduct, or by contamination of the oviducts due to artificial insemination in turkeys (Gross, 1994). Salpingitis caused by *E. coli* can lead to infection of the yolk sac of eggs, and can cause mortality up to three weeks after the birds are hatched (Dho-Moulin and Fairbrother, 1999). Yolk sac infections can also be caused by fecal contamination of the eggs (Barnes and Gross, 1997). Mortality is especially high if the contamination occurs late in the incubation period. Omphalitis is commonly associated with yolk sac infection (Gross, 1994).

Mature chickens or turkeys may experience acute septicemia due to *E. coli* infection (Gross, 1994). The post mortem analysis of the birds indicates that they were in good condition and had been eating. The characteristic lesions are green livers, enlarged spleens, congested pectoral muscles, and necrotic foci on the liver (Gross, 1994). In some cases peritonitis and pericarditis are also seen (Barnes and Gross, 1997). This disease is more commonly seen in turkeys than in chickens (Gross, 1994).

Cellulitis, an economically important emerging disease of broiler chickens in North America was first reported in 1984 in Great Britain (Randall, Meakins *et al.*, 1984). It is an infection of the subcutaneous tissues of the skin, which leads to the production of fibrinous plaques (Gross, 1994). It has also been called infectious process, inflammatory process, or necrotic dermatitis (Barnes, ; Norton, 1997; Dho-Moulin and Fairbrother, 1999). It is generally associated with a scratch or skin break on the thighs or lower abdomen of broiler chickens (Norton, 1997), it has also been reported in turkeys (Olkowski, Kumor *et al.*, 1999). The infection is generally detected at the time

of slaughter (Gross, 1994). In Canada, between the years 1986 to 1998 the incidence has increased 15 fold, and now cellulitis is the leading cause of condemnation of broilers at time of slaughter (Canadian Food Inspection Agency, 1986-1998).

Swollen head disease is a form of cellulitis, which forms over the eye of the bird. The *E. coli* infection occurs after an infection with avian pneumovirus, infectious bronchitis virus or other virus, which increases the susceptibility of the birds (Barnes and Gross, 1997; Cook, 2000). Poor ventilation and high ammonia are also predisposing factors. It is believed that the infection occurs when *E. coli* gain access to subcutaneous tissues through inflamed conjunctival-associated lymphoid tissues (Barnes and Gross, 1997).

Enteritis caused by *E. coli* is considered rare but has been reported for chickens infected with an enterotoxigenic strain of *E. coli* strain (Joya, Tsuji *et al.*, 1990).

With growing restrictions on the use of antibiotics for veterinary use, control of *E. coli* diseases of poultry has become more challenging (Gross, 1994). The use of vaccines for bacterial diseases in the poultry industry has not become wide spread (Peighambari, Hunter *et al.*, 2002), so management practices have been modified to help control the incidence of *E. coli* disease (Barnes, 1994). The general strategy has been to maximize the resistance of the birds to the infection and to reduce exposure to *E. coli* strains. Proper nutrition, good ventilation, and the prevention of outbreaks of other disease reduces the incidence of colibacillosis. Since pathogenic strains are carried in the intestines of poultry it is extremely difficult to eliminate exposure to the strains, which makes other intervention strategies important.

2.2.2 Diseases of Humans

E. coli is a normal component of the flora of the large intestine of humans and animals. Some strains can cause disease in the intestines while others cause disease if they are able to colonize other tissues. There are two major groups of diseases of humans which are caused by *E. coli*: diarrhea and extraintestinal infection.

The diarrhea-causing strains have been grouped into six different pathotypes (Nataro and Kaper, 1998). Enterotoxigenic *E. coli* (ETEC) produce either heat labile toxin (LT), heat stable toxin (ST) or both. Enteroinvasive *E. coli* (EIEC) are able to invade epithelial cells in the colonic mucosa, which results in a strong inflammatory response. Enteropathogenic *E. coli* (EPEC) adhere initially by the bundle forming pili before forming attaching and effacing lesions. Enterhemorrhagic *E. coli* (EHEC) also produce attaching and effacing lesions and produce Shiga-like toxins (SLT). *E. coli* O157:H7 and O26:H11 are prototype strains of the EHEC group (Nataro and Levine, 1994). The enteroaggregative *E. coli* (EAggEC) have a distinctive binding pattern on HEp-2 cells and cause increased mucus secretion from the intestinal mucosa. The diffuse adherent *E. coli* (DAEC) show a distinctive pattern on HEp-2 cells and alter the morphology so that the bacteria are surrounded by the host cell membrane without being engulfed. The diarrheic *E. coli* strains are not normal flora of the human intestines and are thought to be spread by human-to-human contact or by fecal contamination of food or water.

The four types of extraintestinal infection of humans caused by *E. coli* strains are meningitis in newborns, urinary tract infections (UTI), soft tissue infections, and septicemia. The soft tissue infections and septicemia are generally associated with tissue

damage or immune compromised hosts (Sussman, 1997). The ability to colonize extraintestinal sites depends on three factors. The *E. coli* strain must have access to the extraintestinal site and have appropriate adherence factors to allow the strain to colonize the site. Finally, the *E. coli* strain must be able to grow at a faster rate than the host response is able to kill it. If any of these factors are altered, the ability of the strain to cause disease will change. For example, not all *E. coli* strains are capable of causing urinary tract infection (UTI). Virulence factors that function in UTI are more often found in uropathogenic *E. coli* (UPEC) strains than in *E. coli* strains isolated from feces, such as P-pili (Sussman, 1997). The P-pili allow the UPEC strains to bind to the uroepithelial cells which aids in the colonization of the urinary tract. UTI is much more prevalent in women than in men, due to the proximity of the urethra to the anus, demonstrating that the *E. coli* must have access to the site of infection before it can colonize and cause disease (Sussman, 1997). A comparison of the virulence factors in the UPEC strains isolated from immune compromised and immune competent hosts showed that the strains from the immune compromised hosts had fewer virulence factors than the strains from immune competent hosts (Tseng, Wu *et al.*, 2002). This observation illustrates the point that the bacteria must be able to grow more quickly than they are killed by the host immune system since fewer virulence factors are required if the host's ability to clear the infection is impaired.

2.3 Virulence Factors of *E. coli*

The ability of an *E. coli* strain to cause disease depends, in part, on the virulence factors that the strain produces. These virulence factors can be grouped according to the

role they play in the infection. There are four general steps to disease caused by bacterial infection: colonization, avoidance of the host immune system, growth of the pathogen, and tissue damage. Adherence factors like pili or fimbriae aid in the colonization of host tissues. These factors enable the bacteria to adhere to the host tissue and prevent mechanical clearance of the bacteria from the infection site. At the site of infection the bacteria must avoid killing by the host immune system. *E. coli* strains have different mechanisms to accomplish this, such as invasion of host cells, prevention of phagocytosis, modulation of surface antigens or alteration of host factors. The bacteria must also grow at a faster rate than the rate at which they are killed by the host response. In order to grow, the *E. coli* strains must be able to scavenge essential nutrients from host tissues. One of the limiting nutrients in host tissue is iron, and iron acquisition systems have been associated with virulence. Once the *E. coli* strain is established in the host tissue it must cause damage to the host to cause disease. The damage can be caused by the bacterial factors or by the host response. One of the ways *E. coli* strains can damage the tissue of the host is by the production of toxins. An example of this is Shiga-like toxins which causes damage to the vascular endothelium and platelet aggregation (Sussman, 1997).

Bacterial genomes have been described as having two parts: the core gene pool that is required for the strain to grow, and the flexible gene pool that codes additional functions (Hacker and Carniel, 2001). The flexible gene pool contains mobile elements that move in and out of bacterial genomes such as transposons, phage, plasmids, and genomic islands. The flexible gene pool encodes genes, which allow the bacteria to adapt to new environments, such as virulence factors that allow the bacterial strain to

survive the conditions of infection. Genes coding for virulence factors are often found in plasmids, in phage, or in specific genomic islands called pathogenicity islands.

Pathogenicity islands are a type of genomic island and are large regions that encode virulence factors. They tend to have a different GC content and codon usage than the rest of the chromosome (Hacker and Carniel, 2001) since they were acquired by horizontal gene transfer. Pathogenicity islands tend to be inserted near tRNA genes and often have phage, plasmid, or transposon-derived sequences which function in their horizontal movement (Donneneberg and Whittam, 2001). Horizontal transfer of DNA can be between strains of the same species or between unrelated organisms. The transfer of genomic islands can enable a strain of bacteria to adapt to a new environment quickly (Hacker and Carniel, 2001), and in the case of pathogens, the transfer of pathogenicity islands can enable a strains to invade different host tissues.

An example of the acquisition of mobile genetic elements affecting the pathogenicity of strains is the evolution of *E. coli* O157:H7. *E. coli* strain O55:H7 is an atypical EPEC strain, which is closely related to the EHEC strain *E. coli* O157:H7 (Donneneberg and Whittam, 2001). Work by Feng *et al.* (Feng, Lampel *et al.*, 1998) has led to a proposed model for the evolution of *E. coli* strains O157:H7 and O55:H7 from a common ancestor. This work shows the stepwise acquisition of γ-*eae*, *stx2*, the EHEC plasmid and *stx1*, followed by the loss of ability to ferment sorbitol and loss of β-glucuronidase activity, which led to the *E. coli* O157:H7 clone which has spread world wide (Feng, Lampel *et al.*, 1998). This evolutionary model suggests that the strain gained an adherence factor with the γ-*eae* gene, and two toxins with the *stx* genes, which altered its pathotype. Alternate evolutionary models have been proposed for other EHEC strains, which indicates that the EHEC strains were derived more than once (Kim,

Nietfeldt *et al.*, 2001). This example illustrates how the acquisition of some virulence genes can change the pathotype of an *E. coli* strain.

2.3.1 Adherence factors

The ability of the bacteria to colonize host tissues is critical for many bacterial diseases. The primary site for colonization by *E. coli* strains is mucosal tissue. The most common form of adherence to host cells is mediated by pili or fimbriae. Fimbriae are long thin structures on the surface of the bacteria. The tip of the fimbria binds to specific host cell structures, generally polysaccharides. The fimbriae are generally composed of a number of different proteins (Sauer, Mulvey *et al.*, 2000).

Type-1 fimbriae are very common among *E. coli* strains. Expression varies but the *fim* gene cluster is found in more than 95% of *E. coli* strains (Sauer, Mulvey *et al.*, 2000). The major structural protein in type-1 fimbriae is a 17 kDa protein called FimA, but FimD, FimF, FimG, and FimH are also part of the fimbrial structure. The FimC protein functions as a chaperon for the other Fim proteins. The expression of the *fim* gene cluster is controlled by FimB and FimE. The expression is phase dependent due to the orientation of the *fimA* promoter (Sussman, 1997).

Type-1 fimbriae bind to D-mannose, which is on the surface of many eukaryotic cells. This sugar is also common on soluble glycoproteins. It has been shown that fimbriated *E. coli* strains are able to colonize the urinary tract significantly more than afimbriated strains (Sussman, 1997). Type-1 fimbriae also bind to the surface of immune system cells that aids in the phagocytosis and killing of the bacteria. Type-1 pili have been shown to bind to the epithelial cells in the respiratory tract of poultry and

are believed to be associated with the virulence of avian *E. coli* strains (La Ragione, Cooley *et al.*, 2000), but P-pili and curli have also been associated with avian *E. coli* strains (Dho-Moulin and Fairbrother, 1999).

Attachment of *E. coli* strains to the surface of host cells is not always mediated by fimbriae. Attaching and effacing lesions are formed when EPEC and EHEC strains bind to the intestinal epithelium. These lesions are caused by the adherence of the protein intimin to its receptor Tir. Intimin is located on the surface of the bacterial cell, and Tir is located on the surface of the epithelial cells. Tir is a bacterial protein, which is transferred into the epithelial cell by a type III secretion system (DeVinney, Gauthier *et al.*, 1999). The attachment of the *E. coli* strains to the intestinal epithelium using intimin and Tir induces cytoskeletal rearrangements in the host cell which cause the collapse of the microvilli and the formation of a pedestal beneath the adherent bacteria (DeVinney, Stein *et al.*, 1999). The intimin/Tir system of attachment is an adherence mechanism that is not mediated by fimbriae.

2.3.2 Avoidance of the host immune system

The ability of the pathogen to adhere to host tissue is the first step in the disease process for many bacteria, successful colonization of the host also requires the ability to avoid or outgrow the host immune response. The host response to a bacterial infection uses all aspects of the immune system including the lethal effects of complement, antibody responses, and phagocytosis.

A number of genes found in *E. coli* strains are known to protect the bacteria from the killing effects of host serum factors. The presence of surface structures such as LPS,

capsules, and outer membrane proteins can prevent the deposition of complement factors near the surface of the *E. coli* strain (Nataro and Levine, 1994). Serum resistance in *E. coli* strains has been shown to be involved in the virulence of *E. coli* strains that cause extraintestinal infections in humans (Nataro and Levine, 1994), and poultry (Dozois, Fairbrother *et al.*, 1992).

K1 capsules have been associated with highly virulent avian *E. coli* strains (Gross, 1991), and *E. coli* strains that cause neonatal meningitis. K1 capsule is structurally similar to a eukaryotic surface antigen, which makes it poorly immunogenic. Capsules are also known to function in the prevention of phagocytosis by host's immune cells (Sussman, 1997).

Bacteria also avoid immune-mediated killing by invading host cells. Once in the host cell the bacteria must avoid killing mediated by fusion of the phagosome with lysozomes, either by preventing the fusion or by escaping the phagosome. When the bacteria are established in the host cell they are protected from the killing effects of antibody responses and complement (Salyers and Whitt, 1994). Enteroinvasive *E. coli* strains are able to mediate their entry into M-cells in the intestine and avoid killing by lysozomes. Once in the M-cells they grow and spread to neighboring epithelial cells and avoiding killing by the immune system (Sussman, 1997).

2.3.3 Iron Acquisition

Iron has been found to be a limiting nutrient *in vivo* and many extraintestinal pathogens have developed iron acquisition systems that allow them to overcome this limitation. Iron acquisition systems function in two general ways: by siderophore-

mediated up-take or by up-take of iron bound to host factors. Siderophores are high-affinity compounds that will remove iron from host factors and transport it back to the bacteria (Griffiths, 1994). Some bacteria can also bind iron-containing host factors such as transferrin, heme, hemoglobin, or lactoferrin, and release the iron for its own use (Ratledge and Dover, 2000).

Most enteric bacteria contain the enterobactin siderophore system (Ratledge and Dover, 2000). Enterobactin, also called enterochelin, is a low molecular weight molecule with a high affinity for iron. It has high enough affinity to remove ferric iron from transferrin, but not from heme, under physiological conditions (Ratledge and Dover, 2000). Enterobactin that is bound to iron is taken up by the bacterial cell and then cleaved to release the iron. This system enables *E. coli* to acquire iron when environmental iron levels are extremely low, however, it is an energy intensive system, as six gene products are required to synthesize enterobactin and another seven genes are used to transport the iron bound enterobactin into the bacterial cell (Ratledge and Dover, 2000). The genes for the enterobactin system are located on the chromosome of *E. coli* strains. The enterobactin molecules cannot be re-used by the cell since they are cleaved to release the bound iron (Sussman, 1997). This system is only expressed when the bacteria are grown under iron limiting conditions (Griffiths, 1994).

Many *E. coli* strains that are involved in extraintestinal infections also produce aerobactin (Sussman, 1997) (Griffiths, 1994). Aerobactin does not bind iron as avidly as enterobactin, but it can be recycled. It has also been shown that enterobactin is bound by serum albumin which could inhibit its function *in vivo*, whereas aerobactin is not bound by serum albumin (Griffiths, 1994). In one study of cellulitis isolates, 90% of the strains contained the aerobactin iron-chelation system (Peighambari, Vaillancourt *et al.*,

1995). Aerobactin genes can found on the chromosome of *E. coli* strains, or found on plasmids such as pColV (Griffiths, 1994).

E. coli strains can also use iron that is chelated by citrate. When *E. coli* are grown in iron limiting media containing citrate they express an iron-citrate transport system (Griffiths, 1994). The five genes that code for ferric citrate uptake are induced by ferric citrate (Angerer and Braun, 1998). It is not known how prevalent this system is in pathogenic *E. coli* strains (Griffiths, 1994).

There are over 500 different types of siderophores produced by different types of bacteria and fungi (Ratledge and Dover, 2000). Many *E. coli* strains express receptors for siderophores that are produced by other organisms (Sussman, 1997). This strategy for iron up-take enables *E. coli* to take advantage of the energy spent by other bacterial strains since *E. coli* manufactures only the receptor and not the enzymes required for the synthesis of the siderophores.

Some bacteria have receptors that enable them to scavenge iron directly from the host sources without the use of siderophores. Transferrin, lactoferrin, and heme are some of the host proteins that can be used as an iron source *in vivo*. Some *E. coli* strains have a homologue to the *Shigella shu* (*Shigella* heme utilization) gene cluster (Wyckoff, Duncan *et al.*, 1998), which is called *chu*. In a mouse model of uropathogenic *E. coli* infection of kidneys, the inactivation of the heme-uptake system reduced the ability of the mutant strain to compete with the strain it was derived from (Torres, Redford *et al.*, 2001), this suggests that heme uptake aids in the virulence of some *E. coli* strains.

2.3.4 Toxins

Tissue damage is mediated through secreted bacterial toxins in many diseases. There is a wide range of toxins produced by bacterial pathogens. A common structure for toxins is the 'A-B' subunit structure, in which the 'A' subunit is responsible for the alteration of a host factor that causes the tissue damage. The B subunit or subunits are responsible for recognizing specific cell types and mediating the entry of the 'A' subunit into the host cell. The heat-labile toxins (LT) of *E. coli* are an example of this group of toxins (Sussman, 1997). LT is made up one 'A' subunit and five 'B' subunits. The 'B' subunits bind to the host cell factor forming a pore, which inserts into the host cell membrane. The 'A' subunit is cleaved as it passes through the pore formed by the 'B' subunits and enters the host cell. The function of the 'A' subunit is to ADP-ribosylate the $G_{s\alpha}$ protein of the adenylate cyclase system, resulting in an increase of cyclic AMP in the host cell. The high level of cyclic AMP results in increased secretion of Cl^- and a loss of water from the intestinal epithelium cell, which leads to watery diarrhea (Gyles, 1994).

Other 'A-B' toxins have the same general structure of heat-labile toxin but different enzymatic functions. The vero toxins (Shiga-like toxins) have 'B' subunits which bind to the host cell and transport the 'A' subunit into the cell. Unlike the 'A' subunit of LT, the 'A' subunit of vero toxins cleaves the N-glycosidic bond in ribosomal RNA. This function, which inhibits the synthesis of proteins, is cytotoxic for the host cell (Gyles, 1994).

Another type of toxins, the heat-stable toxins (ST), are generally very small peptides that function by the modification of the host cell membrane. Heat stable toxin

'b' has been found in a number of *E. coli* strains including some avian *E. coli* strains (Dubreuil, 1997). The protein is known to cause modifications in the membrane of the target cell the mechanism of which has not yet been identified (Dubreuil, 1997).

2.3.5 Virulence factors of *E. coli* Pathogenic for Poultry

E. coli strains that are pathogenic for poultry have factors that enable them to colonize extraintestinal sites in poultry. Most of the factors that have been proposed to function in virulence have been identified by the association of the factors with avian *E. coli* strains. These factors include adherence factors (type-1 fimbriae, P-pili, and curli), flagella, aerobactin production, capsule production, serum resistance factors and temperature-sensitive hemagglutinin (Dho-Moulin and Fairbrother, 1999).

Type-1 fimbriae, P-pili, and curli have been associated with virulent avian *E. coli* strains (Dho-Moulin and Fairbrother, 1999). La Ragione *et al.* (La Ragione, Cooley *et al.*, 2000) have shown that type-1 fimbriae function in the adherence of avian *E. coli* strains to tracheal epithelium, and that curli function in the adherence of *E. coli* strains to the gut epithelium. Mutants defective in type-1 fimbriae and curli have also been shown to be less persistent and less invasive than the parent strains in a specific pathogen free day-old chicken model using an oral challenge (La Ragione, Sayers *et al.*, 2000). Less is known about the role of P-pili in virulence, but it has been shown that P-pili are not involved in binding to the tracheal epithelium of chickens (La Ragione, Sayers *et al.*, 2000).

Flagella are also involved in colonization. It has been shown that flagella of avian *E. coli* aid in the penetration of mucus, which improves adherence to mucosal

surfaces (La Ragione, Cooley *et al.*, 2000). A mutant that did not express flagella was less able to invade and colonize extraintestinal sites than the parent strains in a specific-pathogen free day-old chicken model using an oral challenge (La Ragione, Sayers *et al.*, 2000).

The production of aerobactin has been shown to be associated with virulence in avian *E. coli* strains (Lafont, Dho *et al.*, 1987) (Ngeleka, Brereton *et al.*, 2002). The genes that code for the aerobactin uptake system are generally located on ColV plasmids in avian *E. coli* but may also be found in the chromosome (Gross, 1994).

Serum resistance has been linked to virulence in avian *E. coli* (Gross, 1994). Serum resistance is mediated by LPS, capsule and other bacterial surface structures (Dho-Moulin and Fairbrother, 1999). The most common capsule types in avian *E. coli* are K1 and K80 (Gross, 1994). K1 capsules are poorly immunogenic and increase the serum resistance of *E. coli* strains (Dho-Moulin and Fairbrother, 1999), and strains that have these capsules are highly virulent (Ngeleka, Kwaga *et al.*, 1996). The TraT protein (Wooley, Nolan *et al.*, 1993), and the outer membrane protein Iss (Pfaff-McDonough, Horne *et al.*, 2000) have been shown to be more strongly associated with virulent avian *E. coli* than non-virulent *E. coli.*

Temperature-sensitive hemagglutinin (*tsh*) has been associated with virulent avian *E. coli* strains (Ngeleka, Brereton *et al.*, 2002), and has been found on ColV plasmids (Dozois, Dho-Moulin *et al.*, 2000). The protein Tsh is a secreted protein of the autotransporter type, the function of which is not yet known. The role of Tsh protein in virulence has been implicated in the development of lesions in the air sacs of chickens (Dozois, Dho-Moulin *et al.*, 2000) A protein (Hbp), which is identical to Tsh except for two amino acids, has been shown to cleave hemoglobin and to bind heme..

Even though type-1 fimbriae, P-pili, curli, flagella, aerobactin production, capsule production, serum resistance factors and temperature-sensitive hemagglutinin, have been identified as possible virulence factors of avian *E. coli* strains, it is believed that more factors have yet to be identified.

2.4 Techniques for the Identification of Virulence Factors

Many techniques can be used to identify virulence factors. Traditional methods for the identification of virulence factors include biochemical purification of bacterial factors and the identification of antigens using immunobloting. These techniques are based on the isolation of proteins that are then used to identify the genes that code for them. These techniques are labor intensive and have been largely replaced by recombinant DNA techniques including signature-tagged mutagenesis, *in vivo* expression technology and subtractive hybridization. These approaches use genome-scanning methods to identify genes coding for virulence factors and are not biased by the location or amount of protein produced by the virulence genes.

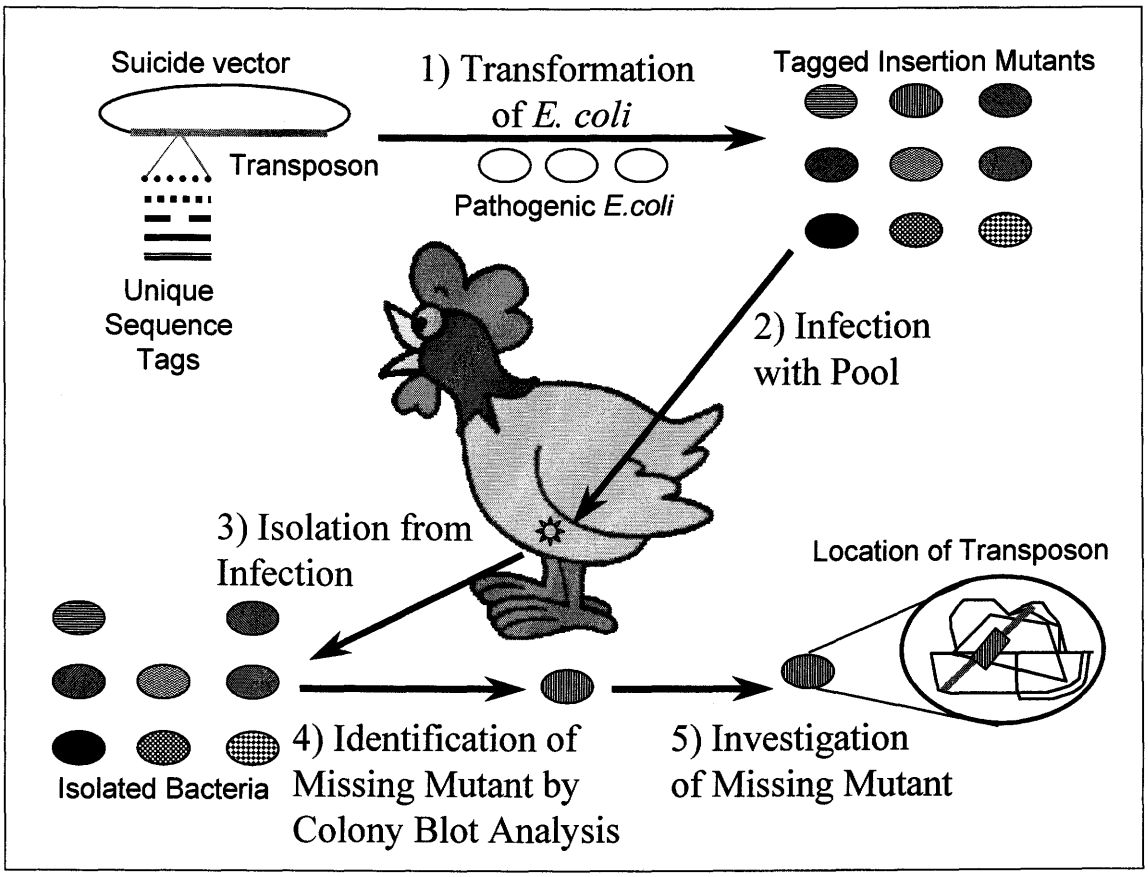
2.4.1 Signature-Tagged Mutagenesis

Signature tagged mutagenesis (STM) is a technique which uses negative selection to identify genes required for virulence (Hensel, Shea *et al.*, 1995). A virulent bacterial strain is transformed with a plasmid containing a tagged transposon, and the transposon is inserted into the chromosome to disrupt a gene. The transposon is tagged with a random sequence that can be amplified by PCR. The tag sequences allow the

detection of individual strains in a pool of tagged transposon mutants (Figure 2-1). The pooled strains are used to infect an animal for a period of time. The bacteria are isolated from the animal and used as a template to generate probes to the tags found. The original pool of bacteria is screened using probes from the isolated bacteria. The results of the screen reveal strains that did not survive the infection due to the presence of the tagged transposons disrupting genes potentially required for virulence. The transposon mutagenesis is random, so there is no bias with respect to the specificity of the genes detected using this technique. In this manner, STM can be used to screen large numbers of tagged mutant strains using few animals.

Signature tagged mutagenesis has been used in a number of systems to identify virulence genes. STM was first used with *Salmonella enterica* serotype Typhimurium. Several known virulence factors were identified, as well as genes with homology to virulence genes of other bacteria, and some novel genes, including a new type III secretion system (Hensel, Shea *et al.*, 1995). STM has been used to study other bacteria including *Vibro cholerae* (Chiang and Mekalanos, 1998), *Staphylococcus aureus* (Mei, Nourbakhsh *et al.*, 1997), *Streptococcus pneumoniae* (Polissi, Pontiggia *et al.*, 1998), *Legonella pneumophila* (Edelstein, Edelstein *et al.*, 1999), *Brucella suis* (Foulongne, Bourg *et al.*, 2000), *Yersinia enterocolitica* (Darwin and Miller, 1999), *Listeria monocytogenes* (Mei, Nourbakhsh *et al.*, 1997), and other bacterial species. This technique has also been used to study the parasite *Toxoplasma gondii* (Knoll, Furie *et al.*, 2001) and fungi such as *Candida glabrata* (Cormack, Ghori *et al.*, 1999), *Cryptococcus neoformans* (Nelson, Hua *et al.*, 2001) and *Aspergillus fumigatus* (Brown, Aufauvre-Brown *et al.*, 2000).

Figure 2-1: Signature-tagged mutagenesis.



A library of suicide vectors carrying mini transposons, which contain the unique tagging sequences are used for STM. 1) The suicide vector library is transformed into the pathogenic bacterial strain. A library of random insertion mutants, each containing a unique tag, is created. 2) A pool of the tagged mutants is used to infect an animal. 3) The tagged mutants are isolated from various organs in the animal, after the infection. 4) The isolated bacteria are compared to the input pool by colony blot analysis to determine which mutants did not survive the infection. 5) The mutants that did not survive the infection are investigated to determine which gene was inactivated by the transposon insertion and what role that gene may have in virulence.

Two factors required for using STM to study the virulence of a pathogen are a high frequency method to disrupt genes in the genome of the pathogen and a system to select for the attenuated transposon mutants. In the STM technique several methods have been used to disrupt sequences in the pathogen with a tagging sequence at high frequency. Initially, the method of disruption used was transposon mutagenesis, with transposons containing tagging sequences. In some bacterial systems, transposons are not suitable for mutagenesis, so other tools have been adapted for use with STM. In *Neisseria meningitidis*, a transposon has not been identified that inserts randomly into the genome at high frequency. A modified STM technique was therefore developed, called shuttle STM (Claus, Frosch *et al.*, 1998). In this technique *N. meningitidis* chromosomal DNA was cloned and then moved into *E. coli*, where transposon mutagenesis was done. The plasmids containing the *N. meningitidis* sequences with transposon insertions were used to transform the *N. meningitidis* strain, which was naturally competent. By homologous recombination, the transposons were inserted into the chromosome of *N. meningitides*.

A similar method was used with *S. pneumoniae*. Instead of using a tagged transposon to disrupt sequences, a tagged suicide vector was used (Polissi, Pontiggia *et al.*, 1998). Chromosomal DNA was cloned into a tagged suicide vector, which was transformed into the *S. pneumoniae* strain. The suicide plasmid was inserted into the chromosome by homologous recombination. The integration of the plasmid disrupted the chromosomal sequences and tagged the mutants.

Another modification of STM was used for the fungal pathogen *C. glabrata* (Cormack, Ghori *et al.*, 1999). In this case, a linearized *S. cerevisiae* plasmid containing the URA3 gene as a selectable marker and tags was used to transform *C. glabrata*. The STM library was created by the insertion of the plasmids into the *C. glabrata*

chromosome by non-homologous recombination. The plasmid used was also able to replicate in *E. coli,* hence sequence recovery could be used to identify the disrupted sequences in selected mutant strains.

STM also requires an appropriate method for selection of virulence traits. The selection of the attenuated mutants has generally been based on infection with the tagged mutants in an animal model. Tissue culture screening methods have also been used. For example, *C. glabrata* was screened using a HEp2 adherence assay, and *B. suis* was screened using an *in vitro* human macrophage infection model. Studies have shown that the type of model used will affect the virulence factors identified. Tagged mutants of *S. typhimurium* were tested in both a mouse and a calf model of Peyer's patch colonization. One of the mutants identified had a reduced rate of colonization in the murine model but normal levels of colonization in the calf model (Tsolis, Townsend *et al.*, 1999). In a murine pneumonia model, 126 mutants of *S. pneumoniae* were identified as attenuated, but 38 of them were not attenuated when the mutants were tested in a mouse septicemia model (Polissi, Pontiggia *et al.*, 1998).

The conditions used to infect the animal can also affect the mutants identified. The dose used for the infection has to be large enough to represent all of the strains in the pool screened so as to reduce the occurrence of false negatives and to insure that the infection is established. However, the dose should not overwhelm the animal, since that leads to very little selective pressure during infection. It has also been proposed that the duration of the infection can affect the mutants identified (Lehoux and Levesque, 2000). A long duration of infection would tend to select for mutations in genes required for persistent infection whereas a short duration of infection would select for the genes required for establishment of the infection.

There are some limitations associated with the STM technique for identification of virulence factors. Trans-acting factors such as toxins cannot be detected since toxins can be supplied by the other members of the pool. There is a detection threshold, which prevents the identification of weakly attenuating mutations. There is also a limit to the pool size that can be screened. The largest pool size that has been used successfully was 96. Originally, with *S. typhimurium*, a pool size of 192 was used but there was a high level of false negatives and the colony blots became less reproducible (Hensel, Shea *et al.*, 1995).

Recent modifications to STM have simplified the technique. Ninety-six tags known to generate strong signals on the colony blots have been used instead of random tags. This was first used with *S. aureus* (Mei, Nourbakhsh *et al.*, 1997). With random tags, the mutant strains had to be checked for false negatives before the library could be screened through an infection model. The 96 tags were checked for signal strength in the colony blots and checked for cross-reaction between the tags. Cross-reaction could reduce the number of attenuating mutants found by the screen. With the use of the 96 characterized tags, dot blots could be used instead of colony blots, which increased the sensitivity of the screen.

Another alteration of the STM technique was the use of defined sequences with PCR to screen the strains (Lehoux, Sanschagrin *et al.*, 1999). The tagging sequences were identified by PCR instead of using radiolabeled probes. Each tag required a different set of primers to amplify the tag sequence. To identify all of the tags in a pool, a PCR reaction had to be done using primers for each tag in the pool. The use of tags identified by PCR removes the need for either colony or dot blots. DNA from the bacteria isolated from the animal model and was used as the template for PCR. This

technique requires one PCR reaction for every tag used. Any strain lost in the infection was identified when the PCR reaction using the primer to its tag was negative.

A third modification of the STM technique involved using a high-density array to screen for the attenuated clones (Karlyshev, Oyston *et al.*, 2001). The array system allows the amount of each tag to be measured quantitatively. The relative amounts of tags in the input and output pools can be compared to determine if the differences are statistically significant. The quantitative analysis increases the sensitivity of the technique. The use of a computer based screening method also reduces the work involved in screening the library. In this array STM technique, each of the transposons contained two tagging sequences, which were flanked by different primers, which increased the reliability of the signal from each of the clones in the transposon deletion library.

2.4.2 Subtractive Hybridization

The technique of subtractive hybridization is used to identify sequences of DNA from one cell type (the tester), and not from another cell type (the driver). The technique can be used with either genomic DNA or with cDNA. Subtractive hybridization can be used to compare genomes of different strains if genomic DNA is used for the subtraction. Subtractive hybridization has been used to identify gross chromosomal differences implicated in some eukaryotic diseases (Kunkel, Monaco *et al.*, 1985) and to identify differences between strains of related bacteria (Emmerth, Goebel *et al.*, 1999; Zhang, Ong *et al.*, 2000; Reckseidler, DeShazer *et al.*, 2001). Differences in gene expression can be investigated when cDNA is used for the

subtraction. Subtractive hybridization using cDNA has been used to investigate cell differentiation (Hedrick, Cohen *et al.*, 1984), tissue function (Diatchenko, Lau *et al.*, 1996), and causes of disease (Hedrick, Cohen *et al.*, 1984; Duguid and Dinauer, 1989).

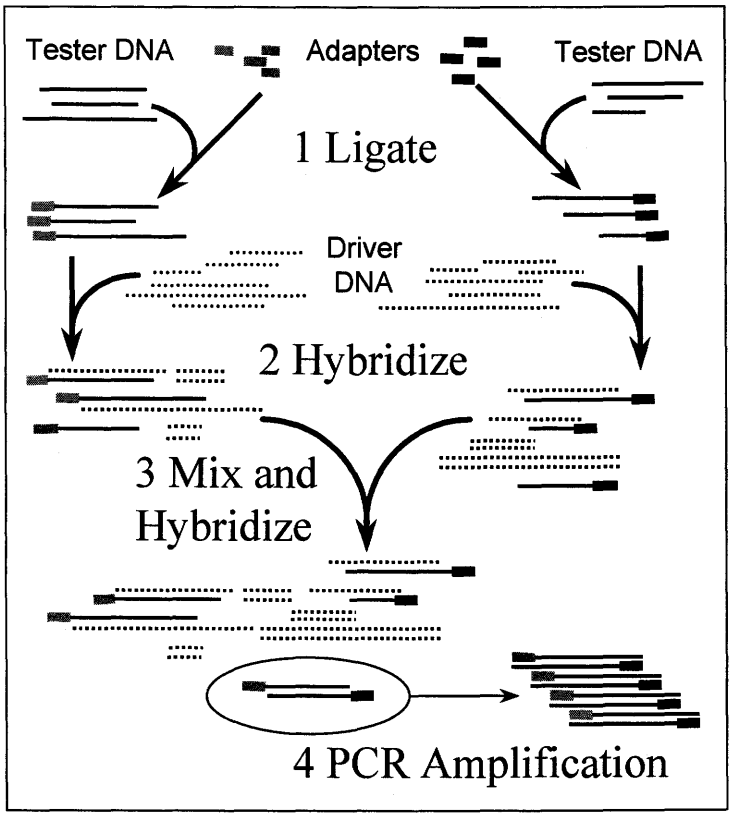
Different methods of subtractive hybridization have been used. The general procedure for subtractive hybridization has been to hybridize the tester DNA (from the strain of interest) to an excess of driver DNA (the control strain). The driver DNA is removed along with any tester DNA bound to it, thus the remaining DNA is specific to the tester strain. Initially subtractive hybridizations used single-stranded tester and driver DNA. The unique DNA was removed from the complementary DNA using hydroxyapatite columns (Hedrick, Cohen *et al.*, 1984).

Several modifications have been used with subtractive hybridization to replace the use of the hydroxyapatite columns, including biotinylation, latex beads, or PCR. One modification of subtractive hybridization used biotinylated driver DNA, which was hybridized to the tester DNA. The complementary DNA was removed by avidin binding (Duguid and Dinauer, 1989). Another modification used poly-dT oligomers linked to latex beads as the primers for the generation of single strands of cDNA. The cDNA linked to the beads was used as the driver DNA (Hara, Kato *et al.*, 1991). After hybridization, the complementary DNA was separated from the tester-specific DNA by the removal of the latex beads. These subtractive hybridization techniques that use physical separation of the common and tester-specific DNA, require multiple rounds of hybridization and subsequent removal of the common DNA before the tester-specific sequences can be isolated.

2.4.2.1 Suppression Subtractive Hybridization

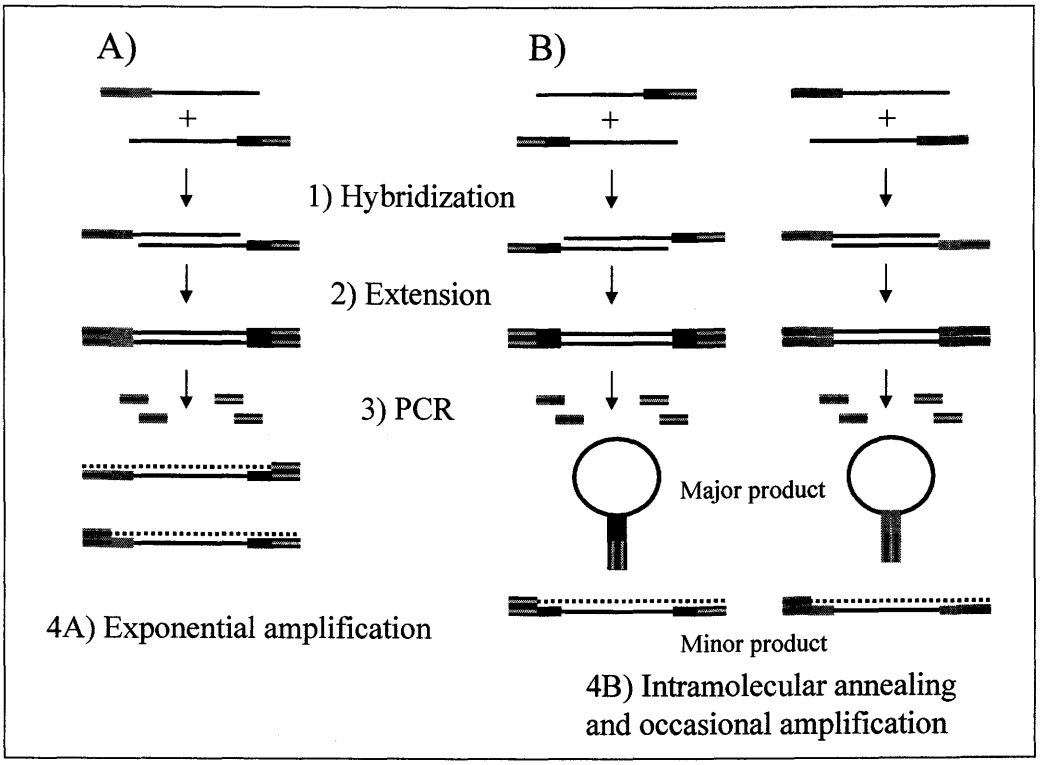
Suppression subtractive hybridization (SSH) is a type of subtractive hybridization that does not remove the common DNA sequences from the tester-specific DNA (Diatchenko, Lau *et al.*, 1996). The tester-specific DNA is amplified by PCR (Figure 2-2). In order to amplify the tester-specific DNA, the two adaptors are ligated onto the tester DNA to create two pools. Each pool has a different single-stranded adaptor ligated on to the ends of the DNA fragments. The two types of tester DNA are then hybridized to two pools containing an excess of driver DNA. This first round of hybridization is only for a short period such that the hybridization does not go to completion. The two hybridization pools are subsequently mixed together and allowed to hybridize to completion. In the second hybridization, the tester-specific DNA that did not hybridize to the driver DNA in the first hybridization is allowed to hybridize. The first hybridization allows the sequences that are common to both the tester and the driver strains to anneal, while the tester-specific DNA remains single stranded. When the two hybridization pools are combined and allowed to hybridize to completion, the tester-specific sequences from each pool can anneal to each other. The tester-specific sequences are then amplified by PCR using the adaptor sequences. This form of subtractive hybridization uses the selective suppression of PCR. With the selective suppression, only the DNA sequences with different adaptors will be amplified exponentially (Lukyanov, Gurskaya *et al.*, 1999). There are inverted repeats in the adaptor sequences that cause a panhandle structure to form in fragments of DNA that have the same adaptors on both ends (Figure 2-3). The intramolecular binding which causes the panhandle structure is much more efficient than the intermolecular binding

Figure 2-2: Suppressive subtractive hybridization.



1) DNA from the tester strain is split into two pools, and adaptors are ligated onto one end of each strand of tester DNA. 2) The two pools are then hybridized with two pools of driver DNA. An excess of driver is used to ensure the common sequences are bound leaving only the tester-specific sequences single stranded. 3) The two pools are mixed and hybridized together, which allows the tester-specific DNA from each pool to hybridize together. 4) The hybridization products that have two strands each containing a different adapter are amplified by PCR, using primers homologous to the adapters. Amplification of the sequences with one of the adaptors is linear not logarithmic. Amplification of sequences containing two identical adaptors is low due to selective suppression.

Figure 2-3: Selective suppression of PCR amplification.



Single stranded DNA with two types of adaptors anneal and are amplified by PCR. 1) The DNA fragments hybridize. 2) The DNA sequences are extended without denaturing to complete both strands. 3) PCR is used to amplify the sequences using primers to part of the adaptor sequences. 4A) The strands of DNA with different adaptors sequences are exponentially amplified. 4B) The strands of DNA that have the same adaptor sequences are inefficiently amplified. The DNA strands form a hairpin structure due to the efficiency of the intramolecular annealing, which is greater than the efficiency of the intermolecular annealing. The hairpin structure is not amplified. The binding of the primer to the adaptor sequence is a rare event that results in some amplification of the sequence containing two identical adaptor sequences.

required for primer binding, so the formation of the panhandle structure prevents primer binding and amplification of the fragment is suppressed.

The formation of the panhandle structure is dependent on the intramolecular annealing being more efficient than the intermolecular annealing, thus the efficiency of the suppression of the PCR amplification is directly dependent on the efficiency of the intramolecular annealing. There are four factors which can alter the efficient formation of the panhandles: the DNA fragment between the primer binding sites, the concentration of the PCR primers, the length of the internal repeats in the primer binding sites and the GC content of the internal repeat. The length of the sequence between the inverted repeats affects the selective suppression since longer sequences have lower probability of intramolecular binding. If the concentration of the primers is very high, the possibility of intermolecular annealing increases and the suppression is less efficient. The length and the GC content of the internal repeat affects the temperature at which the panhandle structure will form. Using this PCR technique, only the fragments that have different adaptors on the ends will be amplified efficiently.

SSH has been used with cDNA to identify genes that are differentially expressed in different tissue types (Diatchenko, Lau *et al.*, 1996), in different stages of parasite development (Bellatin, Murray *et al.*, 2002), in tissue development (Hebrok, Kim *et al.*, 1999) and in the development of cancer (Takeuchi, Nishimatsu *et al.*, 2000). SSH has also been used with genomic DNA to identify differences between strains of bacterial pathogens (Agron, Walker *et al.*, 2001; Bogush, Velikodvorskaya *et al.*, 1999; Zhang, Ong *et al.*, 2000; Reckseidler, DeShazer *et al.*, 2001).

2.4.2.2 Representational Differences Analysis

Representational differences analysis (RDA) is a type of subtractive hybridization where a subset of DNA sequences are used for the subtraction. RDA is used to compare subsets of DNA, instead of the total genomic or cDNA content of the cells involved. It still involves the subtraction of the tester DNA (strain of interest) from the driver DNA (control strain).

The original RDA technique used genomic DNA that was digested with an infrequent cutting restriction enzyme and ligated to adaptor sequences (Lisitsyn, Lisitsyn *et al.*, 1993). The adaptors were used to amplify the DNA. Due to the characteristics of PCR, the smaller fragments were amplified preferentially, and this created a subset of DNA. The subset of DNA was then hybridized to the driver DNA. The final step was to PCR amplify the tester-specific sequences. This type of RDA has been used to identify differences in the genomes of related bacteria (Tinsley and Nassif, 1996), and gene expression in different types of cancer cell lines (Lisitsyn, Lisitsyn *et al.*, 1993).

A second type of RDA uses cDNA amplified by PCR. This technique uses random-primed amplification of total DNA to amplify a fraction of the total cDNA, which is then used for the subtraction hybridization (Schatz and Hubank, 1994). Since this modification uses cDNA instead of genomic DNA, it can be used to identify differences in gene expression.

2.4.2.3 Differential Display

Differential display is a technique related to RDA in that it also uses subsets of DNA to identify differentially expressed genes. Differential display compares the cDNA from two or more sources by gel electrophoresis. The subsets of cDNA are generated by PCR using a short arbitrary sequence and a poly-dT oligomer as the primers. The cDNA sequences that are amplified represent an unbiased subset of the cDNA from the cell types. The PCR products are radiolabeled and then are visualized using gel electrophoresis, to determine which fragments of DNA represent differentially expressed genes. The DNA that is differentially expressed is extracted from the gel, amplified by PCR and sequenced to identify which gene is involved (Liang and Pardee, 1992). By this method, DNA from two or more cell types can be compared and differences in expression can be determined. This technique has been used to investigate cell differentiation (Welsh, Chada *et al.*, 1992) and infection (Sompayrac, Jane *et al.*, 1995; Tatlow, Brownlie *et al.*, 2000).

2.4.3 Genomic Analysis and Mapping by *In Vitro* Transposition (GAMBIT)

Genomic analysis and mapping by *in vitro* transposition (GAMBIT), is a method for determining if a gene is essential. The two requirements for this technique are that the strain used is naturally competent and that the sequence of the region of interest is known. A fragment of DNA containing a region of known sequence is isolated and subjected to *in vitro* transposon mutagenesis to saturating levels. The DNA is transformed into the strain of interest, where it inserts into the chromosome by

homologous recombination. If the transposon is inserted into an essential gene, it will not be able to produce a viable transposon mutant. A series of PCR reactions are done to determine where in the region the transposons are inserted using the chromosomal DNA from the transposon mutants as the template. For each of the PCR reactions, one primer anneals to the transposon sequence and the other primer anneals to part of the target DNA. A series of primers are used that bind to different parts of the target region so each segment of the region can be amplified by PCR. If a PCR product is generated, it indicates that the transposon has inserted near to the primer-binding site to produce a viable mutant. Insertion of the transposon near the primer is indicated as being lethal if no PCR amplification can be detected. Since the transposition is done to saturating levels, any stretch of DNA that does not have a transposon in it must be essential for growth of the bacteria under the conditions tested.

GAMBIT was originally used with *Haemophilus influenzae* and *Streptococcus pneumoniae* (Akerley, Rubin *et al.*, 1998). When this technique was used with transposon mutants grown on rich media, several essential genes were identified for both strains. With *S. pneumoniae*, one of the essential open reading frames that was identified was homologous to a translation initiation factor and another was homologous to a ribosomal protein. These results suggest that both the ORFs would be essential to the bacteria. These results validate the technique as one that can be used to identify essential genes. By using an infection model as the growth condition for GAMBIT, it should be possible to identify genes that are essential for virulence.

2.4.4 *In Vivo* Expression Technology (IVET)

In vivo expression technology (IVET) is a promoter capture system designed to identify genes expressed *in vivo* that are involved in virulence (Slauch, Mahan *et al.*, 1994). This technique is used with an attenuated mutant of a pathogenic strain, where the attenuating mutation is known. In a suicide vector, a promoterless operon is created which consists of a reporter gene, and a survival gene. The reporter gene such as *lacZ* or *cat* is used to monitor expression of the operon *in vitro*. The survival gene is used to compensate for the attenuating mutation in the strain. Chromosomal fragments from the pathogen are cloned upstream of the promoterless operon. The suicide vectors are moved into the attenuated strain, where they integrate into the chromosome by homologous recombination with the cloned chromosomal fragments. In the resulting library of strains, the expression of the operon *in vitro* can be monitored using the reporter gene. Expression of the reporter gene and survival gene require a chromosomal promoter to be situated upstream of the site of integration of the plasmid. The expression of the operon is monitored *in vivo* by infecting a suitable host. In an infection, only the strains that can express the survival gene will grow. After infection, the bacteria are isolated from the animal and plated on a medium selective for the expression of the reporter gene. The selection allows the identification of clones that constitutively express the survival and reporter genes. Negative clones only express the operon under *in vivo* conditions. The integrated plasmids from these clones can be rescued and sequenced to identify the genes that are only expressed *in vivo*. It is assumed that the genes expressed only *in vivo* are more likely to be involved in virulence than the genes that are constitutively expressed.

IVET was first used with *S. typhimurium* (Mahan, Slauch *et al.*, 1993) in an intraperitoneal injection model of mice. The attenuated *S. typhimurium* strain used was a *purA* (purine biosynthesis) mutant such that the survival gene was a promoterless *purA* gene. The reporter gene used was *lacZY*. Five operons were identified which were expressed *in vivo*, but not *in vitro*. Two of these operons did not have homology to known sequences and the other three operons were involved in pyrimidine synthesis, LPS synthesis, replication and recombination. Mutations in the five operons were shown to be attenuating in an oral mouse model. The results demonstrated the ability of IVET to identify genes known to be involved in virulence (Fields, Swanson *et al.*, 1986) and other genes not previously known to be involved in the ability of the bacteria to survive infection.

One of the limitations to the original IVET technique was the requirement for a known attenuating mutation of the pathogenic bacteria. An antibiotic-based system was developed to overcome this limitation. A chloramphenicol resistance gene was used as the survival gene in *S. typhimurium* (Mahan, Tobias *et al.*, 1995). To select for strains, which express the survival genes *in vivo*, chloramphenicol was administered at the infection site. Using this modified form of IVET, researchers were able to identify an operon involved in fatty acid metabolism which is proposed to function *in vivo* in the degradation of inflammatory intermediates of arachidonic acid. The antibiotic selection IVET technique is limited by the amount of the antibiotic that can reach the site of the infection. It is possible to vary the amount and the time of administration of the antibiotic thus influencing the genes that can be detected.

A third type of IVET system, called RIVET, was developed for the detection of genes expressed transiently or at low levels (Camilli, Beattie *et al.*, 1994; Camilli and

Mekalanos, 1995; Slauch and Camilli, 2000). The survival gene in this system is a site-specific recombinase that can remove a tetracycline resistance gene. The tetracycline resistance gene is flanked by recombinase recognition sites in the chromosome of the strain of interest. The IVET promoterless operon contains *lac*ZY and the recombinase gene. The expression of this operon is dependant on promoters upstream from where the operon inserts in the chromosome of the mutant strain. The *lacZY* genes are used to determine if the operon is expressed *in vitro*. This system is a negative selection system since expression of the recombinase leads to the loss of the tetracycline resistance phenotype. The expression of the recombinase at anytime will result in the removal of the tetracycline gene, there by detecting only transiently expressed genes.

RIVET is a powerful tool, but it has two limitations. The screen for tetracycline sensitive clones is a negative selection, which is more labor intensive than positive selection techniques. The second problem is the high frequency of isolation of false positive clones due to the extremely low level of expression of the recombinase required for the recombination event to occur. The rate of recombination *in vivo* and *in vitro* for each tetracycline sensitive clone needs to be compared to determine which clones are actually expressed *in vivo*. Despite these problems, the technique has been used successfully with *V. cholerae* to identify three genes that were required for full virulence in a mouse model (Camilli and Mekalanos, 1995).

2.4.5 Differential Fluorescence Induction (DFI)

Differential fluorescence induction (DFI) is a technique used to identify promoters induced under different growth conditions. Random chromosomal fragments

from the bacterial strain of interest are fused to a promoterless green fluorescence protein (Gfp) in a plasmid vector (Valdivia and Falkow, 1997). The plasmid library is transformed into the bacterial strain of interest. Any clones that are fluorescent under laboratory conditions are removed from the promoter library using fluorescence-activated cell sorting (FACS). The remaining clones are subjected to different growth conditions and sorted by FACS to identify clones that show fluorescent activation. The chromosomal fragments in the plasmids are sequenced to identify which genes are active under the conditions being tested.

This technique was first used to identify genes expressed when *S. typhimurium* associates with macrophage cells in tissue culture (Valdivia, Hromockyj *et al.*, 1996; Valdivia and Falkow, 1997). The *S. typhimurium* library was also used to identify genes expressed under acidic conditions (Valdivia and Falkow, 1996). The genes identified included known virulence factors such as type III secretion genes and genes which code for serum resistance. This technique has been used with an *E. coli* strain which can cause meningitis to determine genes that are expressed after the bacteria invades microvascular endothelial cells in the brain (Badger, Wass *et al.*, 2000).

Other bioluminescent proteins have been used to investigate gene expression. Transposon fusions using the *luxAB* reporter have been used to investigate the expression of genes in *Actinobacillus pleuropneumoniae* (Fuller, Shea *et al.*, 1999), *E. coli* (Guzzo and DuBow, 1994) and other bacterial species (Guzzo and DuBow, 1994).

DFI can be used as a technique for identifying genes that are expressed in each stage of infection or expressed under different environmental condition. The technique can be useful with a broad range of bacterial strains that have plasmid vectors and can

produce functional Gfp. As DFI does not require modification of the chromosome of the strain, it may be useful with pathogenic strains that are difficult to manipulate.

2.5 Hypothesis and Objectives

The hypothesis for this thesis was: there are DNA fragments which are highly associated with avian pathogenic *E. coli,* and virulence factors not yet identified are present in these DNA fragments.

To test the hypothesis our objectives were to identify sequences that were found in avian *E. coli* strain and to determine if these sequences encoded genes, which were involved in virulence. Identification of the virulence factors enabling avian strains to cause extraintestinal infections could lead to a better understanding of the disease process, thus facilitating the development of more effective treatments and vaccines for the prevention of diseases of poultry caused by *E. coli.*

3 MATERIALS AND METHODS

3.1 Materials and Methods for Signature Tagged Mutagenesis

3.1.1 Strains used

The strains used are listed in Appendix A1.

3.1.2 Culture conditions and Growth of Bacteria

E. coli strains were grown in LB (Luria-Bertani) broth or on LB agar (Difco Laboratories, Detroit MI USA) supplemented with the appropriate antibiotics. Cultures in tubes were grown at 37°C with shaking at 200 rpm unless otherwise stated. Bacteria in 96 well plates were grown at 37°C without shaking. Unless otherwise stated, kanamycin (Km, Sigma-Aldrich) was used at 50 μg/mL and rifampacin (Rif, Sigma-Aldrich) was used at 75 μg/mL. MacConkey agar (Difco Laboratories) plates were used for isolation of bacteria from lesions in animal experiments. *E. coli* strains were incubated on M9 agar (Sambrook, Fritsch *et al.*, 1989) without a carbon source at 30°C for conjugation. *E. coli* strains were cultured on LB agar with appropriate antibiotics at 37°C, for long-term storage. The cells were suspended in 50% brain heart infusion broth

(BHI, Difco Laboratories) with 25% w/v glycerol and stored at -70°C. Storage of cultures in 96 well plates was done by growing the cultures in the plate in LB broth with the appropriate antibiotics at 37°C for 24 hours. An equal volume of 50% w/v glycerol was added and the plates were stored at -70°C.

3.1.3 Electrophoresis of DNA

3.1.3.1 Agarose Gel Electrophoresis

Agarose gels were prepared at a concentration of 0.8% agarose (Sigma-Aldrich) in a Wide Mini-Sub Cell GT (Biorad) submerged horizontal electrophoresis apparatus. The gel sizes were 0.7 x 7 x 10 cm or 0.7 x 15 x 10 cm. The agarose gels were run in 1x TAE buffer (40 mM Tris-acetate pH 8.5, 1mM ethylenediaminetetra-acetate [EDTA]) with 0.5 mg/L ethidium bromide and between 60-120V. The DNA was visualized on an ultraviolet transiluminator or a MultiImage Light Cabinet (Alpha Innotech Corporation, San Leandro, CA, USA). The gels were photographed using Polaroid type 57 4x5 land film or using the MultiImage Light Cabinet with the AlphaImager V5.5 system.

3.1.3.2 Acrylamide Gel Electrophoresis

Acrylamide DNA gels were made and run as described by Sambrook *et al.* (1989). The gels were 10% acrylamide made up in 10 mL volumes {3.33 mL acrylamide:bis-acrylamide 30:0.8 (Biorad), 5.6 mL water, 1 mL 10 x TBE (0.45 M Tris-

borate pH 7.5, 10 mM EDTA), 88 µL 10% w/v ammonium persulfate (Biorad), 6.6 µL TEMED (N,N,N',N'-tetramethylethylenediamine; Biorad}. The gels were 0.75 mm thick and were run in a Mini-Protean gel apparatus (Biorad) in 1 x TBE buffer. They were stained in 1 µg/mL ethidium bromide in water for 10 minutes, and were destained in water for 30 minutes. DNA was visualized and photographed as described previously (Section 3.1.3.1).

3.1.4 Transfer of Plasmid DNA

3.1.4.1 Preparation of Electro-Competent Cells:

An overnight culture (10 mL) of the required strain was used to inoculate 1 L of LB, which was grown at 37°C with shaking. The culture was grown until it reached an absorbance of 0.5 to 0.8 at 600 nm as determined using an Ultraspec 3000 (Amersham Biosciences). The culture was chilled on ice for 30 minutes, and pelleted by centrifugation in a pre-chilled rotor at 4000 x g for 20 minutes. The cells were sequentially washed using 1 L of water, 0.5 L of water and 20 mL of 10% glycerol. All solutions used were at 4°C. The cells were pelleted after each wash as described. After the final wash, the cells were resuspended in 2 mL of 10% glycerol, put in 200 µL aliquots, and were flash frozen in a dry ice and ethanol bath, and stored at -70°C.

3.1.4.2 Electroporation

The electro-competent cells were thawed slowly on ice. The cells were aliquoted in 40-100 μL volumes and 0.5-2 μL of DNA was added, mixed and stored on ice for 1 minute. The cells were transferred to a chilled electroporation cuvette (Biorad). The cells were pulsed with the electroporator (Biorad, Gene Pulser) and 1 mL of SOB (20g bacto-tryptone, 5g bacto-yeast extract, 0.5g NaCl, 19 g $MgCl_2$ per liter, pH 7.0), SOC (SOB with 20 mM glucose), or BHI broth (Difco) added. For 0.1 cm electroporation cuvettes, the electroporator was set to 1.8 kV, 25 μF, and 200 Ω. For 0.2 cm cuvettes, the electroporator settings were the same except for the voltage that was set to 2.5 kV. The culture was removed from the cuvette, placed in a 12 mL snap cap tube, and incubated for 1 hour at 37°C with shaking. Aliquots of the culture were plated on the appropriate selective plates.

3.1.4.3 Mating Protocol

The STM library of transposon mutants was created by mating the strain with plasmids containing the tagged transposons (donor) to the strain to the avian pathogenic strain (*E. coli* EC317). The donor cells for mating were produced as follows. A volume of 2 μL of 1 μg/mL plasmid was electroporated into 200 μL of electro-competent S17 λpir *E. coli* cells. The electroporated culture was incubated for 1 hour at 37°C with shaking, and then the entire mixture was plated in 100 μL aliquots on LB Km agar, and incubated at 37°C overnight. The colonies were removed from the plates using a glass

hockey stick to suspend the cells in 1 mL of phosphate buffered saline pH 7.4 (PBS, 8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4, 0.24 g KH_2PO_4 to 1 liter, pH 7.4). The suspended cells from all of the plates were pooled and vortexed thoroughly. One mL of these cells was used to inoculate 25 mL of LB Km. The recipient cells were produced by inoculating 25 mL of LB Rif with 1 mL of an overnight culture of EC134. Both cultures were incubated at 37°C, with shaking, and the absorbance at 600 nm was monitored using an Ultraspec 3000. When the cultures reached an absorbance of approximately 1.0 OD 600 units, 400 μL of each culture were washed twice in 1 mL of PBS. Both the recipient (EC1347) and the donor cells (S17 λpir containing the plasmid) were resuspended in 100 μL of PBS. The two cell types were mixed and spotted in 10 μL volumes on M9 agar without a carbon source. The plates were incubated for four hours at 30°C. The cells were resuspended with 120 μL of PBS using a glass hockey stick. Each spot of cells was resuspended separately and plated on individual LB Rif/Km agar plates to select for the EC1347 cells containing the tagged transposon. The plates were incubated for 24 hours at 37°C, and the resulting colonies were streaked onto LB Rif/Km agar and stored in 96 well plates at -70°C.

3.1.5 DNA Isolation

3.1.5.1 Miniprep Isolation of Plasmid DNA

Miniprep isolation of DNA was done using Qiagen miniprep kits (Qiagen, Mississauga Ontario) according to the manufacturer's guidelines. The procedure involved a modified

alkaline lysis (Brinboim and Doly, 1979) followed by binding of the DNA to Qiagen tip-25 columns. The DNA bound to the column was washed to remove the impurities and eluted in a high salt wash (QF buffer, 1.25 M NaCl; 50 mM Tris·Cl, pH 8.5, 15% isopropanol). The DNA was desalted by isopropanol precipitation, and was resuspended in 20-50 µl of water.

3.1.5.2 Isolation of Total Bacterial DNA

The isolation of total bacterial DNA was done using the protocol of Wilson (Wilson, 1994). A 2 mL aliquot of an overnight culture (12-14 hour LB culture grown at 37°C with shaking) was pelleted, and the supernatant removed. The cells were resuspended in 576 µl of Tris and EDTA buffer (TE; 10 mM Tris-HCl, 1 mM EDTA, pH 8.0), followed by the addition of 30 µL of 10% sodium dodecyl sulfate (SDS; Biorad) and 3 µL of 20 mg/mL proteinase K (Sigma-Aldrich). The solution was mixed well and incubated at 37°C for 1 hour. Following the addition of 100 µL of 5 M NaCl, the solution was mixed thoroughly and 80 µL 10% CTAB (hexadecyltrimethyl ammonium bromide, Sigma-Aldrich) in 0.7 M NaCl was then added. The solution was mixed thoroughly and incubated at 65°C for 10 minutes. The resulting mixture was extracted with an equal volume of chloroform:isoamyl alcohol (24:1) and centrifuged at full speed (Eppendorf Centrifuge 5415 C) for 5 minutes to separate the phases. The aqueous phase was removed, and precipitated by adding 0.6 volumes of isopropanol and mixing. The DNA was pelleted by centrifugation for 20 minutes, and the supernatant removed. The DNA was washed with 1 mL of 70 % ethanol and was pelleted by centrifugation. The supernatant was removed and the pellet was dried for 5 minutes

using a roto-evaporator. The DNA pellet was resuspended in 100 µL of water and incubated at room temperature for one hour with occasional mixing. The concentration of the DNA solution was determined by measuring its absorbance at 260 nm. The DNA concentration was adjusted to 1 mg/mL, assuming that an absorbance of 1.0 at 260 nm is equivalent to 50 mg/mL of double stranded DNA (Sambrook, Fritsch *et al.*, 1989).

3.1.5.3 Maxiprep Isolation of Plasmid DNA

Maxiprep isolation of DNA was done using Qiagen maxiprep procedure (Qiagen, Mississauga, Ontario) by the manufacturer's guidelines. The procedure uses a modified alkaline lysis (Brinboim and Doly, 1979) followed by binding of the DNA to Qiagen tip-500 columns. The DNA bound to the column was washed to remove the impurities and eluted in a high salt wash (Buffer QF: 1.25 M NaCl, 50 mM Tris-HCl pH 8.5, 15% isopropanol v/v). The DNA was desalted by isopropanol precipitation and resuspended in 200 µl of TE buffer.

3.1.6 Southern Blot Analysis

3.1.6.1 Membrane Preparation

The DNA was fixed to the membranes for the colony blots as described by Holden *et al.* (1989). The membranes (Hybond-N, Amersham Biosciences) were cut to fit a 96-well plate, and were placed on the surface of a 14 cm diameter round petri plate containing LB agar. The clones were transferred from the 96-well plate, to the

membranes using a 96-prong transfer template. The agar plates were incubated membrane side up at 37°C until the colonies could be clearly seen (3-5 hours). The membranes were removed from the agar and dried at room temperature for 10 minutes. They were placed colony side up on Whatman 3MM filter paper saturated with 0.4 M NaOH, for 8 minutes, then transferred to a 0.5 M pH 7.0 Tris-HCl solution, and shaken gently for 5 minutes at room temperature. A solution of 2x SSC (20x SSC is 175.3 g NaCl, 88.2 g sodium citrate at pH 7.2 to a volume of 1 L in water) was used to wash the membranes with shaking for 5 minutes. They were placed on 3MM paper and incubated for 2 hours at 80°C to fix the DNA. The membranes were stored at room temperature in plastic bags until needed.

3.1.6.2 Hybridization and Membrane Washes

Membranes were prehybridized in the hybridization solution (0.25 M $NaHPO_4$ pH 7.2, 7% SDS, 1 mM EDTA, 50% formamide) at 42°C for at least 30 minutes but not more than two days. The membranes, in 10 mL of hybridization solution, were transferred to individual bags and 10 μL of probe (Section 3.1.6.3) was added. The bags were sealed and were incubated overnight at 42°C with gentle shaking. The membranes were transferred to a container and washed with 2x SSC for 5 minutes at room temperature. The wash solution was removed and replaced with buffer 2 (0.25 M $NaPO_4$ pH 7.2, 2% SDS, 1 mM EDTA) and the membranes were incubated with gentle shaking at 65°C for 1 hour. The wash solution was removed and replaced with fresh buffer 2 and incubated for another 20 minutes at 65°C, with shaking. The solution was

removed and replaced with buffer 3 (50 mM $NaPO_4$ pH 7.2, 1% SDS, 1 mM EDTA) and incubated for 20 minutes at 65°C. The last wash step was repeated and the membranes were removed and dried by brief blotting on 3MM paper. The membranes were put into fresh bags and exposed to Kodak XAR-5 film for 1 day to 2 weeks.

3.1.7 Polymerase Chain Reaction for Probe Generation

The probes used with the STM technique were generated by two rounds of PCR. The first reaction used chromosomal DNA as the template. The products from this reaction were used as the template for the second reaction. The second PCR reaction was done using radioactive nucleotides, and the products of the reaction were used as the probes for colony blot analysis.

The PCR for both reactions was performed with the following parameters: 95°C for 5 minutes, 20 cycles of 50°C for 2 minutes 72°C for 30 seconds 95°C for 45 seconds 50°C for 45 and held at 4°C when complete. The first PCR reactions were done in 100 µl volumes, and contained 1x Amersham Biosciences PCR reaction buffer (10x Buffer: 100 mM Tris-HCl pH 9.0, 15 mM MgCl 2, 500 mM KCl), 5 ng of chromosomal DNA, 0.5 µM of each primer (P2 5'-TACCTACAACCTCAAGCT-3', P4 5'-TACCCATTCTAACCAAGC-3'), 100 µM for each nucleotide and 1 µl Taq polymerase (Amersham Biosciences, 5-10 units/µl). To purify the template DNA, the PCR reaction was run on a 1.5% SeaPlaque agarose gel (Mandel Scientific, Guelph Ontario). The gel band containing the DNA was removed and allowed to diffuse into 200 µl of water at 4°C overnight.

The second PCR was carried out using 1 µl of eluted PCR product as the template. The reaction was done in 20 µl containing 1x Amersham Biosciences PCR reaction buffer, 0.1 µM of each primer, and 0.2 µl Taq polymerase, and 100 µM for each of dGTP (2`-deoxyguanosine 5`-triphosphate), dATP (2'-deoxyadenosine 5'-triphosphate), and dTTP (thymidine 5'-triphosphate). The dCTP (2'-deoxycytidine 5'-triphosphate) was not added with the other nucleotides, rather 10 µl of dCTP α^{32}P (3000 Ci/mmol Easytides, Mandel Scientific) was added per reaction.

3.1.8 Animal Trials

3.1.8.1 Day-Old Chick Model

The virulence of *E. coli* strains was tested by subcutaneous injection of one day-old chickens as previously described (Ngeleka, Kwaga *et al.*, 1996). Newly hatched broiler chicks were obtained from a local hatchery in Saskatchewan, Canada. Chicks were randomly allocated into groups in animal isolation rooms at the Veterinary Infectious Disease Organization, University of Saskatchewan, and were given a day to acclimatize. Water and a commercial broiler ration were provided *ad libitum*. Each room was ventilated with filtered, non-recirculated air at a rate of 10-12 changes/hour. Air pressure differentials and strict sanitation were maintained in the isolation facility.

The *E. coli* for challenge was grown overnight on LB agar plates containing antibiotics where required. The strains were grown at 37°C in 25 mL LB broth for 12-14 hours in a 100 mL Erlenmeyer flask, with shaking at 200 rpm. A 1:1000 dilution of

the culture was made in a 1 L flask containing 100 mL of LB broth. This was incubated as previously described for about 10 hours, until an absorbance at 600 nm of 1.5 was reached. This absorbance is known to produce a culture containing approximately 1.6 $x10^9$ CFU/mL of *E. coli* in the logarithmic phase of growth. The cultures were diluted in LB broth to the concentration required for the challenge. The viable counts were determined by plating serial dilutions of the diluted cultures on LB plates, incubating them at 37°C overnight, and counting the number of colonies.

Birds were inoculated subcutaneously on the neck, with 0.25 mL of LB containing 10^4 to 10^6 CFU of the *E. coli* strains. Birds were maintained for seven days post-inoculation and monitored daily for clinical signs of disease (Appendix A4), and euthanized when serious disease was found. The guidelines for the Care and Use of Experimental Animals by the Canadian Council on Animal Care were followed throughout the study. The use of animals and endpoints for euthanasia were reviewed and approved by the University Committee on Animal Care and Supply Protocol Review Committee, Animal Resources Centre, University of Saskatchewan.

Dead or euthanized birds were examined for lesions and the pericardial fluid was cultured on MacConkey agar plates. The plates were incubated at 37°C overnight and the colony forming units determined. After 7 days all of the surviving birds were euthanized and checked for lesions and samples of pericardial fluid were cultured. The birds were considered affected if the culture from the pericardial fluid contained *E. coli*, or if they had pericarditis or required euthanization.

3.1.8.3 Cellulitis Model

The ability of *E. coli* to cause cellulitis and systemic lesions was determined using the cellulitis model of infection (Gomis, Watts *et al.*, 1997). The *E. coli* for challenge was grown overnight on LB agar plates containing the appropriate antibiotics. The strains were grown aerobically at 37°C in 5 mL LB broth for 12-14 hours in a 25 mL Erlenmeyer flask, with shaking at 200 rpm. A 1:1000 dilution of the culture was made in 200 mL of BHI broth in 1 L flasks. This was incubated as previously described for 12 hours. The bacteria were diluted in LB broth, until an absorbance at 600 nm of 0.8 was reached. This absorbance is known to produce a culture containing approximately 1 $x10^9$ CFU/mL of *E. coli* in logarithmic phase of growth. The bacteria were diluted to the appropriate levels for the challenge and the actual number of bacteria was determined by plating duplicate 10-fold serial dilutions of the culture on LB agar and incubating at 37°C for 16 hours.

Newly hatched chicks were raised until 19 days of age and randomly allocated into isolation rooms at the Veterinary Infectious Disease Organization. The birds were housed as described in section 3.1.8.2. The birds were provided with a 12 hour lighting period and the room temperature was maintained at 20-22°C. The birds were individually tagged and bled just prior to challenge.

The birds were injected subcutaneously in the left caudal abdomen region with 1 mL of culture. The actual number of cfu/mL was determined by plating serial dilutions of the suspension in duplicate. The birds were monitored twice daily for clinical symptoms and euthanized when serious disease was found (Appendix A5). Blood was

taken from the birds on various days post challenge depending on the group they were in. After 7 days, all the remaining birds were euthanized and checked for lesions. At necropsy, the pericardial fluid was sampled, the cellulitis lesion and spleen were removed and processed as described in Section 3.1.8.2.1.

3.1.8.2.1 Tissue Processing

The tagged transposon library was screened by the isolation of bacteria from the infected birds. Tissue from the spleen and cellulitis lesion was removed aseptically. The tissues were weighed and approximately one half of the sample was processed in a Lab Blender 400 Stomacher in 9 mL of 0.85% saline per 1 g of tissue until the tissue was well homogenized. The other half of the sample was stored at -20°C. A 1 mL aliquot of suspension was pelleted using an Eppendorf Centrifuge 5415 C, and resuspended in 1 mL 0.5% Triton X-100 (Sigma-Aldrich). The resuspended cells were diluted and plated on LB Km plates. The pericardial fluid was collected with a sterile swab and was cultured in 1 mL LB Km for 1 hour before plating on LB Km agar. The blood was directly diluted with 0.85% saline and plated in 100 µl aliquots. The cells from plates with 1000-4000 colonies were harvested, the cells were resuspended in 5 mL LB. Total bacterial DNA was extracted as described in section 3.3.5.2.

3.2 Materials and Methods for Subtractive Hybridization and Subsequent Work.

3.2.1 Strains used

The collection of avian isolates and the ECOR collection of strains are described in the Appendix (Sections A2 and A3). The plasmids and strains used are listed in Appendix A1.

3.2.2 DNA Isolation

3.2.2.1 Miniprep Isolation of Plasmid DNA

Three types of miniprep protocol were used to isolate DNA. For the identification of plasmids containing inserts, the ten-minute miniprep protocol of Zhou, *et al.* (1990), a modified alkaline lysis protocol, was used. Plasmid DNA was isolated from 2 mL of overnight LB culture containing the subtraction hybridization clones and the DNA was resuspended in 20-50 μL of water and stored at -20°C until used.

Miniprep isolation of DNA for sequencing was done using the Qiagen Qiaprep Spin Miniprep protocol following the manufacturer's guidelines. The procedure uses an alternate modified alkaline lysis protocol (Brinboim and Doly, 1979) to lyse the bacteria and a silica gel membrane in spin columns to bind DNA. The DNA was washed with PE buffer (Qiagen) to remove impurities, then was eluted in 20-50 μL of water and

stored at -20°C until used.. Miniprep isolation of cosmid DNA done as described in Section 3.1.5.1.

3.2.2.2 Large Scale Isolation of Cosmid DNA

We used a modified alkaline lysis protocol (Sambrook, Fritsch *et al.*, 1989). The *E. coli* strain containing the cosmid was grown overnight in 500 mL of LB broth at 37°C with shaking and was lysed. Cells were centrifuged for 20 minutes at 5000 x g then resuspended in 18 mL of Solution I (50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0). A 2 mL aliquot of freshly made lysozyme solution (10 mg/mL in 10 mM Tris-HCl pH 8.0) was added followed by, 40 mL of Solution II (0.2 M NaOH, 1% SDS) was added. The sample was mixed by gentle inversion and incubated at room temperature for 10 minutes before the addition of 20 mL of ice cold Solution III (3 M potassium acetate, 5 M acetic acid). The solution was mixed by gentle inversion until there were no distinct layers, followed by incubation on ice for 10 minutes. After this incubation, the solution was centrifuged at 10,000 x g for 15 minutes, the supernatant was filtered through 2 layers of kimwipe (Kimberly Clark), and 0.6 volumes of isopropanol was added. The sample was mixed well and incubated at room temperature for 10 minutes, then was centrifuged for 15 minutes at 10,000 x g at room temperature and the supernatant fluid was removed. The pellet was rinsed with 5 mL of 70% ethanol and was dried for 20 minutes under vacuum. The DNA was resuspended in 7 mL of TE for 3 hours at room temperature. The cosmid DNA was separated from the

chromosomal DNA by cesium chloride gradient centrifugation as described in the following section.

3.2.2.3 DNA Isolation by Cesium Chloride Preparation

The protocol used for CsCl gradient purification of DNA was a modification of Sambrook *et al.* (1989). To 4 mL of DNA in TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA), 4.3 g of CsCl was added. After the CsCl was dissolved, the solution was transferred to 16x76 mm polyallomer quick seal ultracentrifuge tubes (Beckman Instruments Inc., Palo Alto, CA, USA) and 200 μL of 10 mg/mL ethidium bromide was added to each tube. The tubes were balanced with either CsCl/TE (4.3 g CsCl in 4 mL TE) or with mineral oil. The tubes were heat sealed and centrifuged in a Beckman TI-50 rotor in a Beckman L855M ultracentrifuge, at 50 K rpm for 25-30 hours at 20°C. The DNA was visualized with long wave ultraviolet light and removed from the tube using 18 gauge needles. The ethidium bromide was extracted using CsCl-saturated isopropanol, prepared by dissolving 3 g CsCl in 10 mL TE, and adding 50 mL of isopropanol. The DNA solution was dialyzed at 4°C against three 2L changes of TE (30 minutes, 3h hours and overnight).

3.2.2.4 Isolation of Total Bacterial DNA

The protocol used for isolation of total bacterial DNA was described by Ausubel (1994). The bacteria were grown overnight in 100 mL LB broth at 37°C with shaking,

then the cells were pelleted by centrifugation for 20 minutes at 5000 x g. The bacteria were resuspended in 15 mL of 10 mM NaCl, 20 mM Tris-HCl, 1 mM EDTA, 0.5% (w/v) SDS and 100 μg /mL proteinase K (Sigma-Aldrich), mixed gently, and incubated from 6 to 14 hours at 50°C. The DNA was extracted by gentle inversion with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) for 10 minutes, and then centrifuged at 4000 x g at 10-12°C for 20 minutes. The upper layer was removed without disturbing the interface and placed in a fresh tube. One-tenth volume of 3 M sodium acetate (pH 5.5) was added and the solution was mixed gently, then two volumes of absolute ethanol at room temperature were layered onto the DNA solution. The DNA at the ethanol/aqueous interface was spooled out using a glass rod. The spooled DNA was rinsed in 70% ethanol to remove the residual salt. The DNA was air-dried for at least one hour, and resuspended in 2-5 mL of TE buffer by incubation at 4°C overnight.

3.2.3 Construction of Subtractive Hybridization Libraries

Genomic DNA from *E. coli* K-12 was used as the driver DNA in each of these reactions. The genomic DNA was obtained using the large-scale CsCl prep of bacterial genomic DNA protocol (Sections 3.2.2.3 and 3.2.2.4). The two subtractive hybridizations were done using the PCR-Select bacterial genome subtraction kit (Clontech) as recommended by the manufacturer. The technique is a modification of the original suppression subtraction hybridization technique (Diatchenko, Lau *et al.*, 1996). Genomic DNA from *E. coli* strains EC317 and EC2029 were used as the tester DNA for two individual subtractive hybridization reactions. Two rounds of PCR were done, and

the secondary PCR products were cloned into pT-Adv using the AdvanTAge PCR cloning kit (Clontech). The plasmids were transformed into TOP10F' cells (Clontech) and plated on Km LB agar with 50 μg/mL isopropyl-β-D-thiogalactoside. The white colonies were picked and grown in LB Km broth overnight at 37°C. Plasmid DNA was isolated as described in Section 3.2.2.1. Restriction endonuclease digestion of plasmids with *EcoR*I, followed by the separation of the fragments on 1 % agarose gels was used to identify plasmids containing inserts.

3.2.4 Transfer of Plasmid DNA

3.2.4.1 Electroporation

See Sections 3.1.4.1 and 3.1.4.2.

3.2.4.2 Transformation

Chemically competent cells were made using the protocol described by Seidman (1994). Bacteria were grown in 4 mL of LB broth overnight at 37°C, with shaking at 200 rpm. The overnight culture was added to 400 mL of LB broth in a 2 L flask and grown at 37°C with shaking at 200 rpm, until the culture reached an absorbance at 600 nm of 0.375-0.400. The culture was centrifuged in pre-chilled 50 mL tubes for 10 minutes at 4000 x g. The supernatant fluid was decanted and the cells were resuspended gently in 10 mL of ice cold 60 mM $CaCl_2$. The cells were centrifuged for 5 minutes at

3000 x g, the supernatant decanted, and the cells resuspended in 10 mL of ice cold 60 mM $CaCl_2$. The wash was repeated and the cells were resuspended in 2 mL of ice cold 60 mM $CaCl_2$ solution. The cells were flash frozen in aliquots and stored at -70°C.

The cells were thawed by holding them in the palm of a hand. To 100 µL of cells, 10 ng of DNA in a 10-25 µL volume was added. The cells were incubated on ice for 10 minutes, followed by a further incubation at 42°C for 2 minutes, before adding 1 mL of LB broth. The culture was incubated at 37°C for 1 hour and plated in aliquots on selective media.

3.2.5 Southern Blot Analysis

3.2.5.1 Generation of Radiolabeled Probes

3.2.5.1.1 Gel Extraction of Template DNA

The DNA was run on a 1% agarose gel as per Section 3.1.3.1. The DNA fragments to be used for probe generation were cut out of the gel in a block with as little surrounding agarose as possible. The DNA was extracted from the agarose using Qiagen spin columns following the gel extraction protocol of the manufacturer (QIAquick Gel Extraction Kit, Qiagen). The agarose gel slice containing the DNA was dissolved using buffer QG (Qiagen) at 50 °C and applied to the spin column. The column uses a silica gel membrane to bind DNA, and each column can bind up to 5 µg of DNA. The DNA on the membrane was washed with PE buffer (Qiagen) to remove

impurities. The DNA was eluted with 20-50 μL of distilled water, and was either used immediately for probe generation or stored at -20°C until needed.

3.2.5.1.2 Generation of Radiolabeled Probe by Random Priming

The probes were generated by random priming of gel purified DNA using a modification of the method of Tabor (Tabor and Struhl, 1994). The DNA (30-50 μL) was boiled for 10 minutes with 5 μL of 5 mg/mL random primers {$pd(N)_6$ sodium salt, Amersham Biosciences}, and cooled on ice for 2-3 minutes. After chilling the DNA, 10 μL of 10x nick translation buffer {0.5 M Tris pH 7.2, 0.1 M $MgSO_4$, 1 mM dithiolthreitol (DTT), 500 μg/mL bovine serum albumin (BSA, Fraction V, Sigma-Aldrich)}, 3 μL nucleotides (3 mM for each of dCTP, dTTP, dGTP {Amersham Biosciences)} and enough water to bring the reaction volume up to 94 μL was added to each reaction. Four μL of dATP $\alpha^{32}P$ (3000 Ci/mmol, Easytides, Mandel), and 2 μL of Klenow fragment (5-10 u/μL, Amersham Biosciences) were added. The reaction was mixed well and was incubated at room temperature for 1-1.5 hours. To stop the reaction the unincorporated nucleotides were removed as described in Section 3.2.5.1.3. The probes were either used immediately or stored at -20°C. Before use, the probe was denatured by boiling for 10 minutes, cooled for 2-5 minutes on ice and applied to the hybridization solution.

3.2.5.1.3 Removal of Unincorporated Nucleotides by Spin Column

The unincorporated nucleotides were removed using the QIAquick PCR Purification Kit (Qiagen) following the manufacturer's guidelines. The column contained a silica gel matrix that bound to DNA. The DNA on the column was washed to remove impurities. The DNA was eluted with 20-50 μL of water or 10 mM Tris-HCl (pH 8.5), and was either used immediately for probe generation or was stored at -20°C until needed. Before use, the probe was denatured by boiling for 10 minutes, cooled for 2-5 minutes on ice and immediately applied to the hybridization solution.

3.2.5.1.4 Generation of Radiolabeled Probes by PCR

To generate radioactive probes one round of PCR was done using unlabeled nucleotides. The resulting PCR product was used as the template for the PCR with dCTPα^{32}P. The PCR reaction used for both rounds was 1 minute at 94°C with 35 rounds of 92°C for 30 seconds, 60 seconds at the annealing temperature, and then 90 seconds at 75°C. The final step was five minutes at 75°C. The annealing temperature was dependent on the primers used.

The first reaction was 20 μL in volume and contained the template DNA, 1x PCR reaction buffer (Amersham Biosciences), 0.2 μM each of dCTP, dATP, dTTP, and dGTP, 1 pMol of each primer and 0.5 μl of Taq polymerase (5-8 units/μl, Amersham Biosciences). The primers are listed in Table 3-1. This reaction was electrophoresed on a 1% agarose gel and the PCR band cut out and extracted from the gel (Section

Table 3-1: PCR primers and conditions

Primer Name	Template sequence	Sequence (5'-3')	Position on Fragment in bp	Annealing Temperature (°C)	Size of product in bp
296A	Fragment 296	AGTAGTGGTCACGCCTT	32-42	49	162
296B	Fragment 296	ATTCCGATCGCTCACAG	193-177		
358A	Fragment 358	CCTCTGGCATTACATGG	53-69	52	583 (partially plasmid)
358B	pT-Adv flanking fragment 358	TGTTCTACAGAGGCACC	Not on sequence		
361A	Fragment 361	CGCGGCATCACTAATCC	93-109	52	361
361B	Fragment 361	GAGCATCTGCGTCAGGA	454-438		
400A	Fragment 400	AGCTCTGACGCCAGTAT	31-47	52	301
400B	Fragment 400	GGTACATGCGCTGTGAA	328-315		
453A	Fragment 453	CGCCGTCGTTCCTGATA	82-98	52	818
453B	Fragment 453	TACAGCCGCTGCCTGAT	899-883		
1-63A	Fragment 1-63	GTTGCAGGTATCACCGA	474-490	53	591
1-63B	Fragment 1-63	TCTTCCTGAAGTGCGCT	1067-1050		
1-74A	Fragment 1-74	ACGGCTAAGTCCAGTGC	598-582	51	572
1-74B	Fragment 1-74	AACGGAGTGGAGCCAGA	27-43		

The PCR primers are for the amplification of subtractive hybridization fragments (Section 4.4.4)

3.2.5.1.1). The PCR products were eluted in 100 µl of water. A 1 µl aliquot was diluted into 999 µl water, and 1 µl of this solution was used for the second round of PCR. To label the product the second round of PCR used [α^{32}P]dCTP at 33 pM, instead of 0.4 pM of unlabeled dCTP. The concentration of the other components of the reactions was the same as in the standard PCR reaction, and the same PCR program was used. The unincorporated nucleotides were removed (Section 3.2.5.1.3) and the probe was stored at -20°C until used.

3.2.5.2 Generation of Membranes

3.2.5.2.1 Colony Blot Membranes

The colony blot membranes were produced as described in section 3.1.6.2. The lysis of the bacteria on the membranes was done as follows. The membranes were dried at room temperature for eight minutes, and were placed colony side up on a 3 MM filter paper (Whatman) saturated with 10% SDS, for two minutes. They were then transferred colony side up onto a 3 MM filter paper saturated with 0.5 M NaOH 1.5 M NaCl, for five minutes. The membranes were transferred onto 3 MM filter paper saturated with a solution of 0.5 M Tris-HCl pH 7.5 and 1.5 M NaCl for 5 minutes, and then placed into a solution of 2x SSC and were shaken vigorously for 5 minutes. They were placed on paper towels, air-dried for approximately 10 minutes. The membranes were fixed using a Fisher Biotech UV Crosslinker model FB-UVXL-100, on the optimal crosslink setting. The membranes were stored at room temperature in plastic bags until used.

3.2.5.2.2 Colony Lift Membranes

The colony blots were done as suggested by the manufacturer (Hybond-N, Amersham Biosciences). The clones to be screened were grown overnight on LB plates with the appropriate antibiotics. Hybond-N membranes were cut to fit the plates. The membranes were placed on fresh agar plates until they were wet. The membranes were transferred to the plates containing the colonies. The membranes and plates were marked with a syringe needle dipped in India ink. The syringe was punctured through the membranes into the agar to mark both for alignment. The membranes were left on the colonies for approximately one minute and were transferred to fresh agar plates, with the colony side up. Both the original plates and the plates with the membranes were incubated at 37°C for 4 hours. The original colony plates were then stored at 4°C until needed. The cells on the membranes were lysed and fixed as per section 3.2.5.2.1.

3.2.5.2.3 Semi-dry transfer of DNA from Agarose Gels

Genomic DNA (Sections 3.2.2.3 and 3.2.2.4) was digested with *Rsa*I (New England Biolabs) according to the manufacturer, and run on 1% agarose gels. The gels were immersed with gentle agitation in 0.25 M HCl for 10 minutes, 0.5 M NaOH with 1.5 M NaCl for 30 minutes and in 0.5 M Tris-HCl pH 7.4 with 2 M NaCl for 30 minutes. The DNA was transferred to Hybond-N nylon membranes by a modified capillary transfer technique (Sambrook, Fritsch *et al.*, 1989). After the gel was denatured, three

layers of 3MM paper cut to the size of the gel saturated with 20x SSC (3 M NaCl, 0.3 M sodium citrate) were placed on a flat surface and rolled out with a glass rod to remove any bubbles. The gel was placed on top, followed by a piece of membrane the same size as the gel. Two layers of 3MM paper were placed on the membrane followed by a stack of paper towels, and a 500g weight. The transfer was left overnight. The DNA was fixed to the membranes as described in section 3.2.5.2.1.

3.2.5.3 End Labeling of Marker DNA

Marker DNA was end labeled for use with the Southern blots. The reaction mixture contained: 12 µL of 0.5 µg/µL Drigest III (a mixture of lambda *Hind*III and ΦX174 rf *Hae*II digests, Amersham Biosciences), 12 µL of One-Phor-All Buffer PLUS (Amersham Biosciences), 80 µL of water, 4 µL Klenow fragment (Amersham Biosciences), and 4 µL of dATP α^{32}P (3000 Ci/mmol, Easytides, Mandel). The solutions were mixed well and incubated at room temperature for 20-30 minutes. The unincorporated nucleotides were removed using Qiagen spin columns (Section 3.2.5.1.3) and the DNA was eluted in 42 µL of 10 mM Tris-HCl pH 8.5, and 10 µL DNA loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF and 30% glycerol in water) was added. The labeled Drigest III DNA was stored at –20°C for up to six weeks. The marker was used in 5 to 15 µL aliquots.

3.2.5.4 Hybridization and Washes

The membranes were prehybridized in approximately 0.2 mL/cm^2 hybridization solution. The hybridization solution contained 5x SSC, 0.5 % SDS, and 5x Denhardts solution {0.5 g Ficoll (Type 400, Amersham Biosciences), 0.5 g polyvinylpyrrolidone (Sigma-Aldrich), 0.5 g Bovine Serum Albumin (Fraction V, Sigma-Aldrich) to a volume of 500 mL with water}. The membranes were individually sealed in plastic bags, and incubated at 65°C for 1-4 hours with gentle shaking (80 rpm). The probe was denatured by incubating in a boiling water bath for 5 minutes and cooling on ice for 2 minutes. The probe was added to the existing hybridization solution and was hybridized overnight at 65°C with gentle shaking. The membranes were then washed twice for five minutes in 2x SSC, 0.1 % SDS at room temperature. The membrane was washed with 0.1x SSC, 0.5% SDS at 65°C for 20 to 60 minutes depending on the strength of the signal coming from the membrane, as measured by a Geiger counter. The membranes were air-dried on paper towels for 5-10 minutes, sealed in plastic bags, and exposed to Kodak X-OMAT AR film at –70°C.

3.2.6 Sequencing and Sequence Analysis

The subtractive hybridization fragments that were found in the *E. coli* strains EC317 or EC2029 but not in *E. coli* K-12 were sequenced. Automated DNA sequencing was performed on a 373 Stretch DNA sequencer from Applied Biosystems using the Big Dye Terminator version 2.0 Kit, by the Plant Biotechnology Institute, Saskatoon,

Saskatchewan. The M13 forward and reverse primers were used to initiate the sequencing reactions. The cosmids were sequenced using primers made at VIDO or using primers obtained from Sigma-Genosys. The primers were designed using Primer Designer Version 2.0 (Scientific and Educational Software, 1994). The primers for sequencing are listed in Table 3-2.

The sequences were analyzed for DNA homology, using the National Center for Biotechnology Information (NCBI) BLASTN program (http://www.ncbi.nlm.nih.gov/BLAST). Sequences with weak DNA homology were analyzed for protein homology, using the NCBI BLASTX (Altschul, Madden *et al.*, 1997). Sequences were assembled using Clone Manager 6 Version 6.0, and Align Plus 4 Version 4.1 (Scientific and Educational Software, 1994-2000).

Cosmid sequences were analyzed using Glimmer version 2.02 (Delcher, Harmon *et al.*, 1999). Glimmer2 was used to predict which open reading frames were genes based on mathematical modeling. The Glimmer2 parameters used were minimum gene length of 90 bps, minimum overlap of 30 bp, minimum overlap percent of 10%, threshold score of 90, with the independent scores, with the first start codon and with linear DNA.

Putative genes identified by Glimmer2 and Clone Manager were translated into the predicted amino acid sequence using Clone Manager. The putative genes were examined using a BlastP homology search and a Motifs (http://www.motif.genome.ad.jp) search for protein motifs. The Motifs search screens the protein sequence through several existing protein structure libraries. These libraries are PROSITE (Hofmann, Bucher *et al.*, 1999), ProDom (Corpet, Gouzy *et al.*, 1999),

Table 3-2: Primers for sequencing cosmids.

Name of primer	Template DNA	Sequence (5' to 3')	Location on final sequence
1-63R1	Cosmid 2491	GCTCCACTAAGGCGATT	761-745
1-63R2	Cosmid 2491	CTGCAGTGACTGCTCCA	772-756, 3094-3078
1-63U1	Cosmid 2491	AACCGCCTGCTCTCACA	1545-1561
1-63U2	Cosmid 2491	TTCGGAGAGCTGGCCTT	2889-2905
296R1	Cosmid 2483	GAGAAGCCACTGTGAGC	1575-1591
296R2	Cosmid 2483	TAATCGGCAGATCGCGT	2170-2186
296U1	Cosmid 2483	GGCAACGCCTTCGATAA	1429-1413
296U2	Cosmid 2483	ATGGCGGCATGGAGTGA	842-858
YhiD	*E. coli* K-12	GTGGTATGTACGAACTGG	*yhiD* of *E. coli* K-12
453R1	Cosmid 2165	GGTATCCACACATCAGG	1600-1616
453R2	Cosmid 2165	CGCGTTCTTCGATGATG	2283-2299
453R3	Cosmid 2165	TAACCAGCATTCCGACC	3075-3091
453R4	Cosmid 2165	GACTGAAGATGGCGAAC	183-167
453U1	Cosmid 2165	GATATAGTCACAGTTCC	794-778
453U2	Cosmid 2165	ACGTGGTTGCTGTTGAG	84-69
453U2A	Cosmid 2165	TGTCTGCTGGTCTGGCT	3748-3764
CS1	SuperCos 1	AAGTGCCACCTGACGTCT	SuperCos 1, 7852-7869
USC	SuperCos 1	CCGTGGAATGAACAATG	SuperCos 1, 135-119

BLOCKS (Henikoff, Henikoff *et al.*, 1999), PRINTS (Attwood, Flower *et al.*, 1999), and Pfam (Bateman, Birney *et al.*, 1999).

Searches to find promoters for open reading frames were done using the Neural Network Promoter Prediction program (Reese and Eeckman, 1995) (http://www.fruitfly.org/seq_tools/promoter.html). The searches were done using both strands of DNA, for prokaryotes, with a minimum promoter score of 0.5. To identify possible transcription factor binding sites the Signal Scan program (http://bimas.dcrt.nih.gov/molbio/signal/) (Prestridge, 1991) was used.

3.2.7 Work with Cosmid Sequences

3.2.7.1 Heme Use and Heme Binding

The bacterial strains were grown in 5 mL of LB broth containing the appropriate antibiotics at 37°C with shaking, overnight. A 100 µL aliquot was used to inoculate 5 mL of LB that was incubated at 37°C with shaking until the absorbance at 600 nm was 1.0. The cultures were spread in 20 µL aliquots on plates. The plate contained M9 salts supplemented with 0.2% casamino acids, and either 200 µM 2,2'-dipyridyl (Sigma-Aldrich) or 200 µM EDDA (ethylenediamine-di-(o-hydroxyphenyl acetic acid)). The plates were dried for 1 hour at 37°C. The wide end of a Pasteur pipette was used to remove plugs of agar from the plates. These wells were filled with 40 µL of either 2.5mM hemin or 2.5 mM $FeCl_2$. The plates were incubated at 37°C overnight. Growth around the $FeCl_2$ well on 2,2'-dipyridyl or EDDA supplemented plates indicated the

ability of the bacteria to grow on those plates. Growth around the hemin well indicates the ability of the strain to use heme as an iron source.

EDDA was prepared by dissolving 10 g EDDA in 190 mL of boiling HCl and cooled to 20°C (Rogers, 1973). The EDDA was filtered and added to 1.5 L of 4°C acetone. The solution was adjusted to pH 6 using 1 M NaOH, and stirred overnight. The suspension was vacuum-filtered (Whatman #1 paper) and washed with ice-cold acetone. The precipitated EDDA was air-dried for 48 hours. The EDDA was dissolved in 0.1 M NaOH to a concentration of 0.1 M and filter sterilized. The 2,2'-dipyridyl was prepared at 0.1M in 50% ethanol (vol/vol). $FeCl_3$ was prepared at 1M in 0.1 M sodium citrate with 0.1 M sodium carbonate and filter sterilized. Hemin was added to 20 mM NaOH to a concentration of 10 mg/mL, and then heated to 65°C to dissolve. The hemin solution was filter sterilized. All of the solutions were stored at 4°C.

The heme binding assay was done using strains grown on M9 agar plates supplemented with 0.4% glucose, 0.2% casamino acids and 1 mg/mL hemin. Some of the plates were supplemented with either 100 μg/mL ampicillin or 50 μg/mL km. The bacterial strains were streaked onto the plates with the appropriate antibiotics and incubated at 37°C until 2-4 mm colonies were visible (24-48 hours). *E. coli* K-12 (MG1655) was used as a negative control and *E. coli* strain pRAP117 in JM105 was used as a positive control (Pontarollo, Rioux *et al.*, 1997). Strains that were gray on the plates were considered positive for heme binding. Strains that were white on the plates were considered negative for heme binding.

3.2.8 Statistical Analysis

Statistical analysis was done using Prism Graphpad (version 2.01, 32-bit for Windows 95). Comparison of the occurrence of the fragments in the ECOR strains and in the cellulitis isolates was done by contingency table analysis (Kleinbaum, Kupper *et al.*, 1988; Altman, 1991). The comparison of the GC content of the fragments was done using a unpaired t test with Welch's correction (Altman, 1991).

3.3 Materials and Methods for the Disruption of Sequences

3.3.1 Strains and Plasmids

The strains and plasmids use are listed in Appendix A1.

3.3.2 Ligations

DNA ligations were done using a vector to insert molar ratio of about 1:3 for blunt end ligations and 1:1 for cohesive end ligations. The concentration of the DNA for vector and insert was estimated after the DNA was run on a 1% agarose gel. The ligation reactions were set up in 10 μL volumes. For blunt-end ligations, the reaction contained 2 μL T4 DNA ligase (Amersham Biosciences, 4-10 unit/μL), 1 μL buffer 3 (New England Biolabs, 50 mM Tris-HCl pH 7.9, 10 mM MgCl2, 100 mM NaCl, 1 mM DTT), 0.25 μL Bovine Serum Albumin/DTT supplement (2 mg/mL BSA, 0.36 M DTT),and 1 μL 10 mM ATP (Sigma-Aldrich). The blunt end ligations were incubated at

14°C overnight. For cohesive-end ligations, the amount of ligase was reduced to 0.5 μL per reaction, and the incubation was at room temperature for 1-4 hours.

3.3.3 Removal of DNA Overhangs

The DNA was blunt ended using T4 polymerase. To 42.25 μL of DNA, 5 μL New England Biolabs buffer 2 (10 mM Tris-HCl pH 7.9, 10 mM MgCl2, 50 mM NaCl, 1 mM DTT), 0.25 μL of 10 mg/mL BSA, 0.5 μL of 10 mM dNTP's (10 mM for each type of nucleotide), and 2 μL of T4 polymerase (USB, 5-10 unit/μL) were added. The solution was incubated at 12°C for 20 minutes. The DNA was purified using a Qiagen Spin column PCR purification protocol following the manufacturer's guidelines.

3.3.4 Dephosphorylation of Cloning Vectors.

The DNA was dephosphorylated by addition of DNA in 43 μL of water to 5 μL of One-Phor-All Buffer PLUS (Amersham Biosciences) and 2 μL of diluted calf intestinal alkaline phosphatase (CIP). To dilute the CIP, 1 μL (20 u/μl, Amersham Biosciences) was added to 2 μL of One-Phor-All Buffer PLUS and 17 μL water. The DNA and CIP solution were mixed and incubated at 37°C for 30 minutes and then at 50°C for 30 minutes. The DNA was purified using a Qiagen Spin column PCR purification kit following the manufacturer's guidelines.

3.3.5 Allele Replacement.

The allele replacement constructs were electroporated into *E. coli* EC317. Transformants were streaked on an LB plate containing 25 μg/mL chloramphenicol (Cm) and grown at 30°C overnight. Five colonies were picked, used to inoculate 1 mL of LB broth at 42 °C, and serial dilutions were plated on LB plates containing 25 μg/mL Cm, which had been pre-warmed to 42°C. The plates were incubated at 42°C overnight. Five colonies from the plates were used to inoculate 5 mL of LB broth, which was shaken at 30°C overnight. A 100 μL aliquot of overnight culture was used to inoculate 5 mL of LB broth, which was incubated at 30°C overnight. This procedure was repeated 3-7 times. After 4 or more passages at 30°C, 50 μL of the overnight culture was added to 5 mL of LB broth, and the culture was incubated with shaking at 30°C for 1 hour. Chloramphenicol was added to the culture to a final concentration of 50 μg/mL. The culture was incubated with shaking at 30°C for another hour. Ampicillin was added to a final concentration of 150 μg/mL and the culture was incubated for 4 hours at 30°C with shaking. Serial dilutions were made of the culture and the dilutions of 1/100 to 1/10000 plated on LB Cm plates, which were incubated at 30°C overnight. Colonies were picked and inoculated into 100 μL of 0.85% NaCl in a 96 well plate. The colonies were replica plated on LB with 25 μg/mL chloramphenicol, LB with the antibiotic in the allele replacement cassette (either Tc or Km), LB with both Cm and the allele replacement antibiotic, and on MacConkey agar. The allele replacement mutants were expected to be lactose positive on MacConkey agar and negative for growth on LB Cm but able to grow on LB with the allele replacement antibiotic. Mutants were tested to confirm the allele replacement by Southern blot analysis and PCR.

3.3.6 Virulence Testing

3.3.6.1 Day-Old Chick Model

Virulence testing for the day-old chick model was done as described in Section 3.1.8.1.

3.3.6.2 Cellulitis Scratch Model

Challenge strains of *E. coli* were grown on LB agar with appropriate antibiotics for 12-14 hours at 37°C. The *E. coli* strains were used to inoculate 10 mL BHI broth in a 50 mL Erlenmeyer flask, and were grown at 37°C for 12-14 hours with shaking at 200 rpm. A 1:3000 dilution of the culture was made into 200 mL of BHI broth in a 1 L flask. This was incubated at 37°C with shaking for 6-8 hours, until an absorbance at 600 nm of 0.8 was reached. This absorbance is known to produce a culture containing approximately 1 x10^9 CFU/mL of *E. coli* in the logarithmic phase of growth. The challenge inoculum was washed once in normal saline and resuspended to an absorbance at 600nm of 0.8.

The model used for production of cellulitis and colibacillosis in broiler chickens was a modification of a model previously described (Gomis, Watts *et al.*, 1997). The birds were housed and monitored as described in section 3.1.8.3. At 25 days of age, two parallel, three cm long full thickness scratches were made through the skin on the left caudal abdominal region with a diabetic lancet (Lifescan Lancets, Lifescan Canada Ltd. Burnby, B.C.). The challenge inoculum was applied with two cotton applicators

(Hardwood Products Company LLC, Guilford, Maine, USA) to the scratches. Application of *E. coli* on the scratches was repeated 15 minutes later. After 7 days, all the remaining birds were euthanized and examined for lesions. The air sacs, pericardial fluid and cellulitis lesion from each bird were cultured on MacConkey agar.

3.3.7 Statistical Analysis

The comparison of the survival curves was done using the log rank test, Prism Graph Pad, version 3.00 for Windows. Comparison of mortality and lesions in the animal experiments was done by contingency table analysis (Altman, 1991).

3.3.8 Reverse Transcription PCR

3.3.8.1 RNA isolation

Chicken liver and spleen samples were taken from broiler chickens two days after administration of *E. coli* EC317 in the cellulitis scratch model (Section 3.3.6.2). Liver and spleen samples were also taken from uninfected broiler chickens. The tissues were aseptically removed, and stored on ice. One gram of each organ was homogenized in 1 mL of 0.85% saline. The viable bacteria counts were determined by plating serial dilutions on MacConkey plates. The plates were incubated at 37°C for 12-16 hours before the colonies were counted.

The RNA was stabilized using the RNAProtect bacteria reagent (Qiagen) as recommended by the manufacturer. Two volumes of RNAProtect bacteria reagent were added to the remaining sample and incubated at room temperature for 5 min. After incubation, the sample was centrifuged and the pellet was stored at -20°C until the RNA was isolated. RNA isolation was done using the Qiagen RNeasy kit (Qiagen) as recommended by the manufacturer. The RNA on the RNeasy spin column was treated with DNase I (RNase-Free DNase Set, Qiagen) as described by the manufacturer. The RNA was eluted from the column, and was further treated with DNase I (GibcoBRL Life Technologies) as described by the manufacturer. The concentration of the total RNA isolated was determined by measuring the absorbance of the sample at 260nm. The RNA was stored at -20°C until used.

3.3.8.2 Production of cDNA

First strand cDNA was synthesized using Superscript II RNase H^- reverse transcriptase (Invitrogen) as described by the manufacturer. A random hexamer primer (Amersham Biosciences) was used for the first strand cDNA synthesis of the bacterial RNA. Oligo (dT) primer was used for chicken cDNA synthesis. The single stranded cDNA was stored at -20°C, until used.

3.3.8.3 PCR

The three sets of primers used are listed (Table 3-3). The *gapA* primers were the positive control for RNA from *E. coli* strains. The 354 primers were primers to the 3-54 subtractive hybridization fragment. The reactions were in a total volume of 50 μL, and contained 0.2 μM of each primer, 0.2 mM of each dNTP, 2 μL of cDNA (Section 3.3.8.2), 2 units Taq DNA polymerase (Amersham Biosciences), and 1x PCR buffer (10 mM Tris-HCl, pH 9.0, 1.5 mM $MgCl_2$ and 50 mM KCl, Amersham Biosciences). The PCR reactions were incubated at 94°C for 1 minute. The following three steps were repeated 33 times: 92°C for 30 seconds, 52°C for 60 seconds, and 75°C for 90 seconds. The reaction was then incubated at 75°C for 5 minutes. The PCR products were analyzed by electrophoresis on a 1% agarose gel.

Table 3-3: RT-PCR primers

Name	Gene	Sequence (5'-3')	Position	Product Size (bp)
GapA1	GADPH*	TCACGCTACTACCGCTACTC	728	416
GapA2	GADPH*	CGTTGTCGTACCAGGATACC	1143	416
354-1	ORF1 of 3-54	CCGGCATCGTTATGGTGATT	8	350
354-2	ORF1 of 3-54	TGCCGGATTACTGCATCTGA	357	350

* GADPH is the *gadA* gene that codes for glyceraldehyde 3-phosphate dehydrogenase-A subunit, which is constitutively expressed.

4 RESULTS

4.1 Standard Signature Tagged Mutagenesis

The original STM technique (Hensel, Shea *et al.*, 1995) and a modification were used. The standard STM protocol was employed initially to identify virulence factors in avian *E. coli* strains; transposons containing random tagging sequences were transferred into the avian strain creating a tagged insertion library. The members of the tagged insertion library were screened by colony blot analysis after the library was created to ensure the members had a strong signal when probed with the tagging sequence. Pools of bacteria were screened by animal infection, and isolates recovered from animals were used to generate probes for a colony blot of the original members of the pool.

4.1.1 Generation of a Rifampicin Mutant

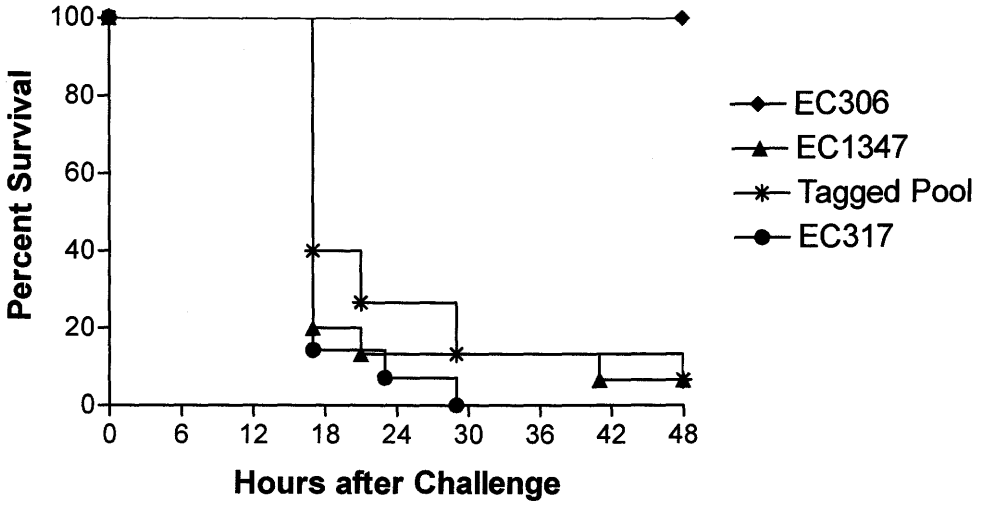
A spontaneous rifampicin (Rif) mutant was isolated from *E. coli* EC317 to produce a strain with a selectable marker necessary to create the transposon insertion library. A rifampicinresistant EC317 derivative was isolated by plating an overnight broth culture on LB plates containing Rif at concentrations of 25, 50, 75, and 100 μg/mL. After overnight culture, each plate had approximately 10 colonies. Five

colonies were chosen from the 100 μg/mL Rif plates and streaked onto plates containing 75 μg/mL rifampicin to confirm resistance. Of the five strains tested, four grew and were considered rifampicin resistant. One of the four resistant strains was confirmed to be serogroup O2 by slide agglutination. This strain was designated *E. coli* EC1347.

4.1.1.1 Virulence of the Rifampicin Resistant Mutant

Rifampicin resistance mutations are known to attenuate *Salmonella* and *E. coli* strains (Bjorkman, Hughes *et al.*, 1998; Linde, Fthenakis *et al.*, 1998), therefore the virulence of the *E. coli* EC1347 strain had to be tested before it could be used for STM. This test was done using the day-old chick model of infection. Groups of 15 or 16 day-old chicks were challenged by subcutaneous injection in the neck. *E. coli* strains EC317, EC1347, EC306 (HB101), and a pool of 96 tagged transposon mutants derived from *E. coli* EC1347 were used to challenge the birds. *E. coli* EC306 is known to be avirulent in day-old chicks and was the negative control for the trial. *E. coli* EC1347 was the rifampicin resistant strain derived from *E. coli* EC317. *E. coli* EC317 was the positive control and was known to be highly virulent. A pool of 92 STM-knockout strains derived from *E. coli* EC1347 was also used to determine if a pool was fully virulent, even though individual members of the pool might be attenuated. The doses per bird were: 1.7×10^5 CFU for the *E. coli* EC317 and *E. coli* EC1347 groups, 1.4×10^5 CFU for the pooled plate of transposon insertion mutants, and 4.5×10^4 CFU for *E. coli* EC306 group. The results of the experiment are shown in Figure 4-1. All of the birds challenged with *E. coli* strains EC317, EC1347 or the pool of tagged mutants were dead

Figure 4-1: The survival curves for day-old chicks after subcutaneous challenge with *E. coli* strains.



The survival curves of the groups challenged with the *E. coli* EC306 was significantly different from the survival curves of the groups challenged with the other strains ($P<0.0001$). The survival curves of the groups challenged with *E. coli* strains EC317, EC1347 or with the pool of tagged insertion mutants were not significantly different ($P=0.6846$) as determined by the log rank test. Each group contained 15 or 16 birds.

or required euthanasia 48 hours post challenge. The last bird from each of these groups had to be euthanized because solitary birds are unduly stressed.

4.1.2 Generation of the Tagged Transposon Insertion Library

The transposon insertion library was made by electroporation of the tagged pUT plasmid into *E. coli* S17 λpir (EC1266), which is a permissive strain for the replication of the pUT plasmids. The resulting transformants were mated with *E. coli* EC1347, which is the rifampicin resistant mutant of *E. coli* EC317. In total, 1749 transposon insertion mutants were made from 95 independent matings. Each transposon insertion mutant was streaked for isolated colonies on Rif/Km LB plates. The clones were stored at -70°C in duplicate 96 well plates. The pUT plasmid contains the *bla* gene that codes for ampicillin resistant phenotype. To determine how many of the clones lost the pUT plasmid, the clones were plated on LB amp plates. All the clones were tested and 192 (9.1%) were ampicillin resistant, indicating that the pUT plasmid was not lost. The ampicillin resistant clones were not used for subsequent work.

4.1.3 Removal of False Positives

Colony blot analysis was used to identify false negative clones, before the library was screened through the animal model. The membranes were probed using PCR generated probes. The total DNA isolated from 92 tagged mutants was used as the template for the radioactive PCR. The clones from the 96 well plate were used to

produce colony blot membranes. The controls were *E. coli* EC1347, *E. coli* EC1266 (which is permissive for the replication of pUT), plus a well containing all of the 92 mutant strains. The screen to identify negative colonies was done with 920 of the transposon insertion mutants and 419 were positive. Initially, 276 of the positive colonies were arranged into three new 96 well plates for further testing.

4.1.4 Pilot Animal Trial

A pilot animal experiment was done to determine the optimal times for isolation of the tagged transposon disruption library from birds challenged using the cellulitis model of infection. Three-week old broiler chickens were challenged with a subcutaneous injection of *E. coli* on the lower abdomen. Three types of *E. coli* were used to challenge the birds. The negative control group contained five birds challenged with 1.1×10^5 cfu/mL of *E. coli* HB101 (EC306). The positive control group contained six birds challenged with 1.2×10^6 cfu/mL of *E. coli* EC317. Fifteen birds were challenged with 1.2×10^6 cfu/mL of a pool of 92 tagged mutant strains.

All of the chickens challenged with *E. coli* EC306 survived until seven days after challenge; following euthanasia, no *E. coli* was isolated from any of the pericardial fluid samples. All the birds challenged with *E. coli* EC317 had pericarditis, but no *E. coli* were isolated from two birds that were euthanized. Cellulitis was observed in five of the six birds. The birds in the control groups had the expected lesions and mortality.

Additional sampling of the birds challenged with the pool of transposon insertion mutants was done. The plan was to divide the birds into three groups of five birds. One group was to be bled one day post-challenge and euthanized after three days. Another

group was to be bled two days post challenge and euthanized after four, while the final group was to be bled at three days and euthanized after seven. The blood samples were plated on Km LB agar.

Many of the birds died or were euthanized before the day they were scheduled for examination. Four of the five birds to be bled after three days died in the first two days after the challenge. Swabs of the pericardial fluid were taken from each bird, diluted in LB broth, and plated. The spleens and cellulitis lesions were aseptically removed from the birds, homogenized and plated. Table 4-1 and Figure 4-2 show the concentration of *E. coli* in the tissues. The cellulitis lesions contained the most bacteria initially and remained high throughout the examination period in most of the birds. There were large numbers of *E. coli* present in the pericardial fluid, but they were cleared by day seven. The number of bacteria isolated from the spleen and blood was much lower than in the pericardium. The plates of bacteria isolated from the post mortem examination of the chickens were used to generate PCR probes for colony blots. To ensure that there was sufficient diversity to represent all 92 of the transposon insertion mutants used for the challenge, the entire plates with 1,000-10,000 colonies were used for probe generation. The bacteria from the plates with 10^3-10^4 colonies were used to generate probes for Southern blots. The membranes for the blots were generated from the input plate of mutant strains. None of the blood samples, and only one of the spleen samples had enough bacteria for probe generation.

Colony blots were performed using membranes generated from the colonies of plate 1F, containing mutants used to challenge the chickens. The probes were generated from the bacteria isolated from the infected chickens. A positive control blot was done using probe generated from the DNA from the pooled cells in the challenge. It was

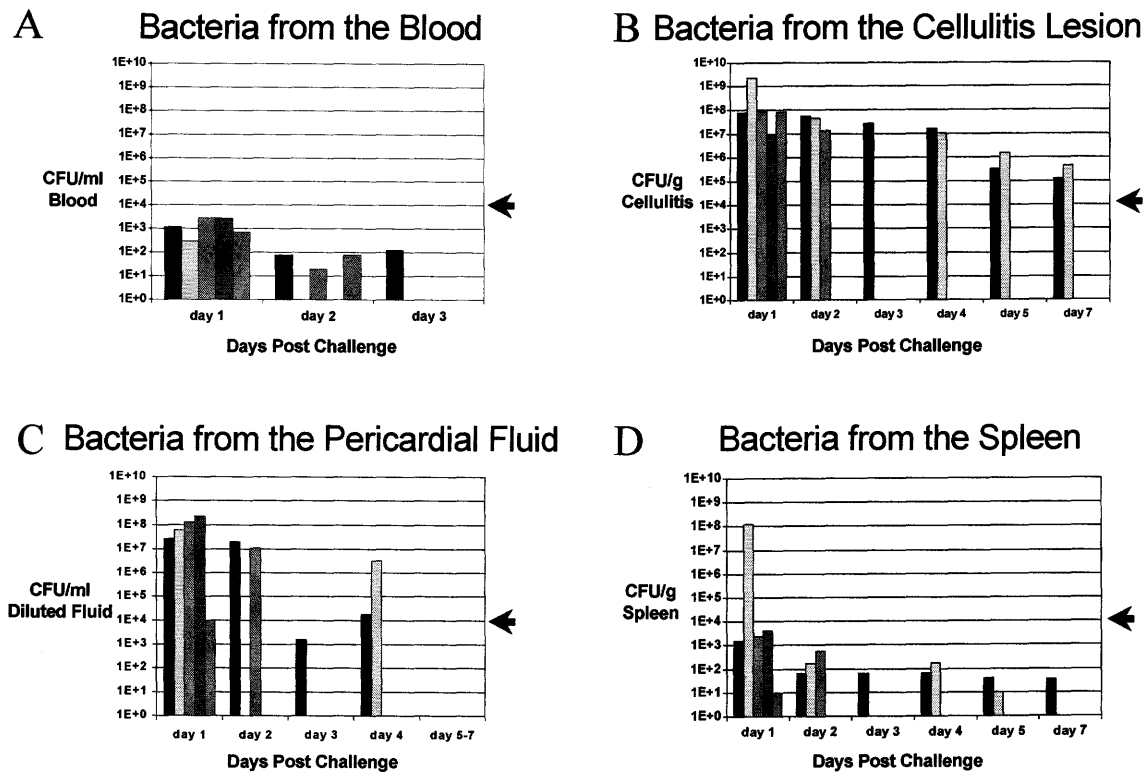
Table 4-1: Bacteria isolated from lesions of birds challenged with a pool of 92 tagged transposon mutants.

Bird	Days Post Challenge	Spleen CFU/100 mg	Pericardial Fluid[a] CFU/mL	Cellulitis Lesion CFU/100 mg	Blood CFU/mL
3	1	1.60×10^3	2.65×10^7	7.8×10^7	2.82×10^3
11	1	1.24×10^8	6.07×10^7	2.17×10^9	Not done
15	1	2.37×10^3	1.32×10^8	8.8×10^7	Not done
1	1	1.03×10^1	1×10^4	8.3×10^7	1.18×10^3
4	1	4.17×10^3	2.25×10^8	$\sim 1 \times 10^7$	2.75×10^3
12	2	6.7×10^2	1.95×10^7	5.5×10^7	Not done
5	2	1×10^2	0	4.52×10^7	7.2×10^2
13	2	1.51×10^3	1.08×10^7	1.36×10^7	Not done
2	3*	6.5×10^1	1.6×10^3	2.67×10^7	3.0×10^2
9	4*	1.7×10^2	1.89×10^4	1.67×10^7	8×10^1
14	4*	5.48×10^2	3.29×10^6	1.04×10^7	1.3×10^2
7	5*	4×10^1	0	3.25×10^5	0
10	5*	1×10^1	0	1.50×10^6	0
6	7*	3.5×10^1	0	1.31×10^5	8×10^1
8	7*	0	0	4.35×10^5	2×10^1

[a] The pericardial fluid was removed with a swab which was placed in 1ml of LB, so the cfu/mL was per mL of eluted pericardial fluid.

* The bird was killed for a time point and did not require euthanization.

Figure 4-2: Isolation of bacteria from infected chickens.



The amount of *E. coli* isolated from A) blood, B) cellulitis lesion, C) pericardial fluid, and D) spleen. The birds were challenged with a pool of 92 tagged insertion mutants. The arrows indicate the minimum level for use with STM, which was 10^4 CFU/mL. The limit was determined by the number of bacteria required to form 1000 colonies when 100 µl of processed tissue was plated.

expected that all of the colonies would have strong signals since each colony was strongly positive on previous blots and was present in the pool used to make the probe (Figure 4-3). The control blot did not show strong signals from all of the colonies, nor did the blots using probes derived from the bacteria isolated from the infected tissues.

4.2 Modifications of Standard Signature Tagged Mutagenesis

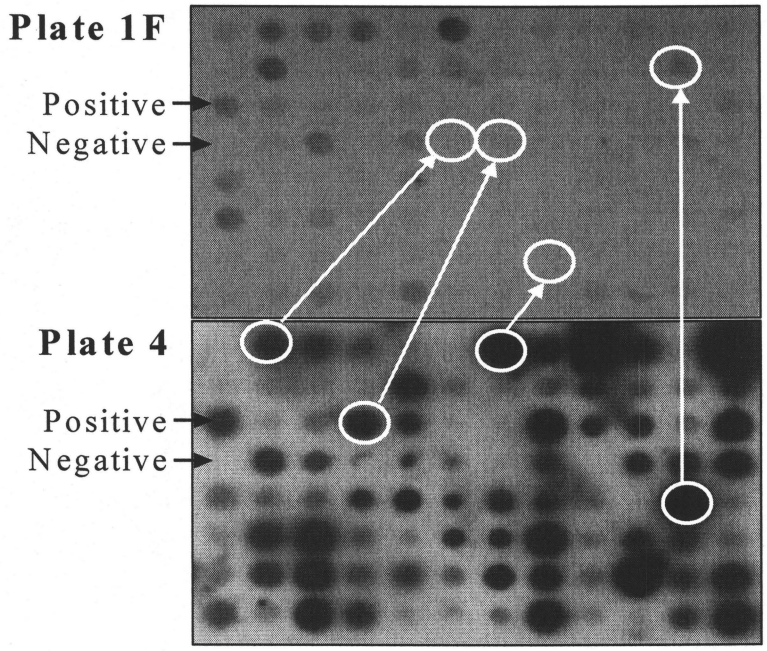
Since the previous experiment using STM was unsuccessful, we modified the STM protocol in an attempt to make the technique work. The two major areas that were altered were the PCR reaction used to generate the probes and the colony blot protocol.

4.2.1 Polymerase Chain Reaction Modifications

PCR was used to generate the probes for the blots. In an attempt to obtain reproducible results in the colony blots used with the STM technique, the PCR conditions were altered. Changes were made to the primer and nucleotide concentration, to the reaction buffer, to the type of polymerase used and the annealing temperature of the reaction.

The primer and nucleotide concentrations used were altered to those developed by laboratories that have successfully used STM (Hensel, Shea *et al.*, 1995). The primer concentrations were changed from 0.5 μM to 1.2 μM, and the nucleotide concentration was raised from 100 μM to 200 μM in the reaction. The higher concentrations of primers and nucleotides

Figure 4-3: Comparison of positive colonies from the first and the second round of colony blot analysis.



Colony blot analysis of clones to be used for STM. Plate 1F was assembled from clones that were shown to be positive by colony blot analysis. Plate 4 was one of the initial plates used for the colony blot analysis. The circles on plate 4 indicate some of the clones that were used in plate 1F. The white arrows to the clones of plate 1F indicate the new positions of the clones on plate 1F. The controls for the colony blots are indicated by the block arrows. The positive controls were a pool of all of the colonies in the plate and the negative control was strain *E. coli* EC317 that is the parent strain for the transposons mutants.

increased the amount of the 80 bp PCR product (Figure 4-4), but did not improve reproducibility of the colony blots.

The buffers used with the PCR reaction were tested to determine the optimal buffer for this reaction. The buffer recommended by Hensel *et al.* (1995) was compared to a number of commercially available buffers (Table 4-2). The PCR products generated using the different PCR buffers are shown in Figure 4-5. The product generated using Buffer F (Invitrogen) had the highest intensity band with the least smear. Use of Buffer F to generate the probe did not affect the reproducibility of the colony blots.

Taq polymerases from Amersham Biosciences and Qiagen were compared and no significant difference was observed when the PCR products were visualized by gel electrophoresis.

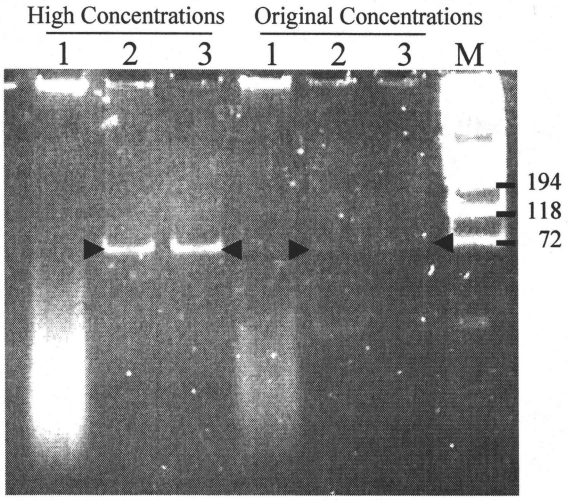
4.2.1.4 Annealing Temperatures

The annealing temperature used for the PCR reactions was originally 50°C. Comparisons using annealing temperatures of 45°C and 52°C did not affect the PCR products as seen by gel electrophoresis.

4.2.2 Blotting Conditions

The blotting conditions for the colony blots were altered in order to obtain more consistent results. The original protocol used two short washes to remove the excess

Figure 4-4: Comparison of different primer and nucleotide concentrations.



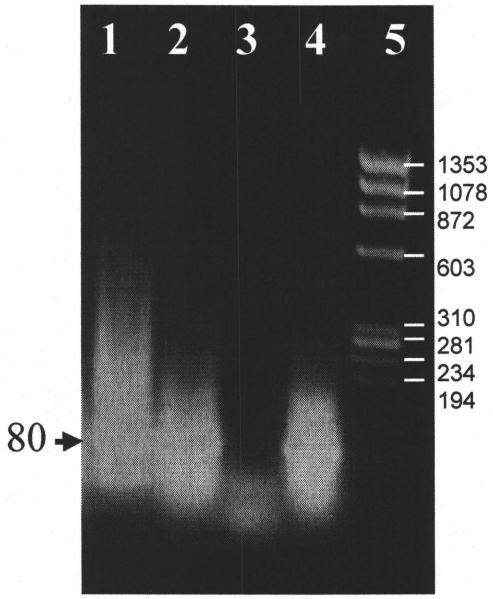
A 10 % acrylamide gel of the PCR products using two concentrations of primers and nucleotides. Two sets of reactions were used: the original concentrations (0.5 μM primers and 100 μM nucleotides), and the higher concentrations (1.2 μM primers and 200 μM nucleotides). For each set of reactions the same templates were used: 1) chromosomal DNA without tagged transposons, 2) chromosomal DNA containing tagged transposons, and 3) tagged transposon on a plasmid. Lane M was the markers and the sizes of the bands in bp are indicated. The expected PCR product were indicated by the black arrows. The bands for the original concentration of primers and nucleotides were very faint, where as the bands for the higher concentration were much more intense. The result indicated that PCR using the higher concentration of primers produced more product.

Table 4-2: The composition of the PCR buffers tested.

Buffer	Composition
STM PCR Buffer	10x buffer: 200 mM Tris-HCl pH 8.3, 20 mM $MgCl_2$, 500 mM KCl 0.1% Tween 20
Amersham	10x buffer: 100 mM Tris-HCl pH 9.0, 15 mM $MgCl_2$,500 mM KCl.
Qiagen*	10x buffer: Tris-HCl pH 8.7, KCl, $(NH_4)_2SO_4$, 15 mM $MgCl_2$
Invitrogen	
B	5x buffer: 10 mM $MgCl_2$, 300 mM Tris-HCl pH 8.5, 75 mM ammonium sulfate.
F	5x buffer: 10 mM $MgCl_2$, 300 mM Tris-HCl pH 9.0, 75 mM ammonium sulfate.
J	5x buffer: 10 mM $MgCl_2$, 300 mM Tris-HCl pH 9.5, 75 mM ammonium sulfate.
N	5x buffer: 10 mM $MgCl_2$, 300 mM Tris-HCl pH 10.0, 75 mM ammonium sulfate.

* Qiagen does not publish the exact composition of their PCR buffer.

Figure 4-5: Comparison of PCR buffers.



A 1.5% SeaPlaque agarose gel containing the PCR products from the reaction, using bacterial genomic DNA from the mutants of plate 1Fa as the template with different buffers: 1) Invitrogen buffer B, 2) Invitrogen buffer F, 3) STM PCR buffer, and 4) Pharmacia buffer. Lane 5 contained the markers. The sizes in bp are listed to the right of the gel and are indicated by the white lines. The expected product for the PCR reactions is a 80 bp fragment that is indicated by the black arrow. The Invitrogen buffer F reaction buffer had the most 80 bp product with the least amount of contaminating smear.

probe followed by two lower stringency washes and two higher stringency washes. When the high stringency washes were omitted, the signal of the colonies on the blot increased, as did the intensity of the negative control. Increasing and decreasing the quantities of probe used was tried as well. The lower concentrations of probe gave reduced signal without increasing the demarcation between the positives and negatives, and higher concentrations increased the background of the blot.

4.2.3 Pool Size

There is a limit to the number of tagged transposon mutants that can be detected in a pool as the signal from each member of the pool is reduced as the size of the pool increases. The pool size originally used for STM with *S. typhimurium* was 96 colonies (Hensel, Shea *et al.*, 1995). Colony blot analysis was done using different numbers of clones (8,32,54, and 96) to generate the probes. Membranes prepared from the entire pools were used for the blots. The results showed a uniform increase in the signal from the clones that were used to generate the probe. The signal intensity between colonies used to generate the probe remained variable. The reduction in the pool size did not eliminate the problem of variable signal strength. The colonies that had been used to generate the probes had an increase in the signal intensity, but the differences in the signal intensity between colonies remained. The colonies that were not used to generate the probe had a significant background on the colony blot. The reduced pool size did not correspond to an increase in the intensity of the signal from the weak signal members. It was expected that the members of the smallest pool would exhibit less colony-to-colony variation than seen among the members of larger pools. If this effect

was observed, it would have indicated the pools were too large and the individual signals were washed out. This effect was not seen, so other factors must be responsible for the weak signals from the blots.

4.3 Modified Signature Tagged Mutagenesis

In the modified STM technique, tagging sequences identified as strongly positive were used to make the transposon insertion library. The tagged transposons were screened for signal strength by colony blot analysis and tags that had a strong signal were used to generate the insertion mutant library. The size of the screened pools was limited by the number of strong signal tags since each tag can only be in a pool once. Pools of 96 were used previously with the modified technique (Mei, Nourbakhsh *et al.*, 1997).

4.3.1 Identification of Strong Signal Tagged Transposons

Colony blot analysis of the clones containing the tagged transposon on a plasmid vector identified thirty-five clones with strong signal tags. Results were confirmed by Southern blot analysis using the plasmids containing the positive tagged transposons. Twenty-two of the 35 clones were positive by this method. The colony blot analysis identified some plasmids as positive when the Southern blot analysis showed they were negative, so the tags were checked by Southern blot analysis. The plasmids from 152 clones were screened by Southern blot analysis and 36 tags were found to have a strong signal when probed.

At this point we decided to use subtractive hybridization to investigate the genomic differences between the avian *E. coli* strains EC317 and EC2029 and the *E. coli* K-12, thus the STM approach was abandoned.

4.4 Subtractive Hybridization of Two Avian *E. coli* Strains Against *E. coli* K-12 MG1655

4.4.1 Subtractive Hybridization

To identify sequences unique to avian *E. coli* strains, subtractive hybridization was performed individually with two avian *E. coli* strains (EC317 and EC2029) and *E. coli* K-12. The technique is a modification of the original suppression subtraction hybridization technique (Diatchenko, Lau *et al.*, 1996). The common DNA and the *E. coli* K-12 specific DNA were not removed after the hybridization, but PCR was used to amplify the avian-specific hybridization products.

4.4.2 Cloning of the DNA Fragments

The PCR products from the two suppressive subtractive hybridizations were cloned creating two libraries of fragments, one derived from *E. coli* EC317 and the other from *E. coli* EC2029. From the EC317 library, 78 of the 457 plasmids tested had an insert. From the EC2029 library, 31 of the 300 plasmids tested contained an insert.

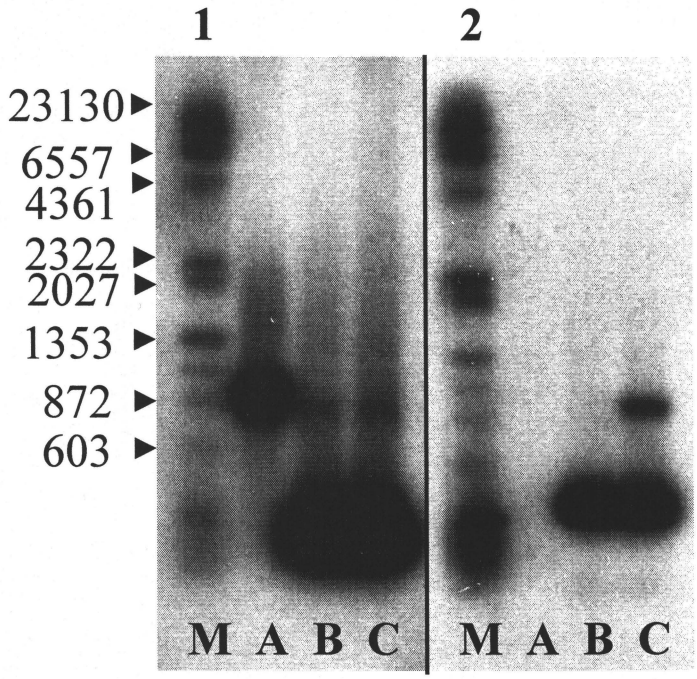
4.4.3 Verification of the Absence of Each DNA Fragment in *E. coli* K-12 M1655

The suppression subtractive hybridization technique used did not remove the common or driver specific DNA fragments so Southern blot analysis was necessary to confirm that the plasmid inserts were unique to the tester strain. In this case, digested chromosomal DNA from *E. coli* strains EC317, EC2029 and MG1655 was probed with the plasmid inserts (see Figure 4-6). Fragments present in either EC317 or EC2029, but not in MG1655, were used for subsequent work. Of a total 104 Southern blots, 65 fragments were found that were present in a pathogenic strain (either *E. coli* EC317 or *E. coli* EC2029) but absent in *E. coli* K-12 strain MG1655. From the EC317 library, 48 of 73 fragments were found to be in EC317, but not in *E. coli* K-12. Similarly, 15 of 31 fragments from the EC2029 library were not found in *E. coli* K-12. The size of the fragments from both libraries varied from 203 to 1882 bp in length.

4.4.4 Sequence Analysis of Pathogen Specific Fragments

The subtractive hybridization fragments absent in *E. coli* K-12 were sequenced on one strand (Table 4-3). The GC content of the fragments varied from 32.2% to 57.8%. The EC317 derived fragments had a GC range of 42.6% to 57.8% with a mean of 50.3, while the EC2029 derived fragments had a GC range of 32.2% to 55.2% with a mean of 42.7%. The mean values of GC content for the two groups of fragments was significantly different ($P > 0.0001$). The sequences were analyzed by BLAST homology search (Altschul, Madden *et al.*, 1997) (Table 4-3). Three of the fragments were repeated (3-13 and 1-64, 338 and 364, 119 and 348), so there were only 62 unique

Figure 4-6: Southern blot analysis of subtractive hybridization fragments.



Southern blot 1 was probed with fragment 336. Southern blot 2 was probed with fragment 364. Lane M: labeled lambda Hind III digest and phiX-174 *Hinc*II digest. The sizes of the molecular weight markers are indicated in bp. Lane A: *E. coli* K-12 strain MG1655 genomic DNA digested with *Rsa*I. Lane B: *E. coli* EC 2029 genomic DNA digested with *Rsa*I. Lane C: *E. coli* EC 317 genomic DNA digested with *Rsa*I. The strong bands in lanes A, B, and C in the blot 1 indicate the presence of the fragment sequence in all three strains, so fragment 336 was not specific to the avian *E. coli* strains. The absence of a band in lane A of blot 2, shows that fragment 364 is not found in *E. coli* K-12, and so is specific to the avian *E. coli* strains.

Table 4-3: Summary of BLAST search of clones that were generated by subtractive hybridization of DNA from avian *E. coli* strains and DNA from *E. coli* K-12.

Clone	Size (bp)	Sequences with similarity	Score	Probability (e^x)	GenBank accession no.	% GC
Phage sequences						
17	409	Bacteriophage 21 lysis genes	266	-69	M65239	46.5
28*	465	*E. coli* clone SauE4.D1 neonatal meningitis strain-specific DNA sequence	434	-119	AF222136.1	51.2
50	307	DNA replication primase protein - phage N15 (P)	157	-38	NP 046932.1	50.2
53	603	Coliphage phiX174,	1195	0	NC 001422.1	42.6
101	320	Bacteriophage VT2-Sa provirus	60	-7	NC 000902.1	44.7
103	289	Bacteriophage lambda	446	-123	J02459.1	57.8
152	987	*E. coli* O157:H7 DNA	1671	0	AP002556.1	51.0
199	348	Bacteriophage lambda	642	0	NC 001416.1	55.7
202	995	Bacteriophage N15	345	-92	AF064539	43.6
214	1547	gp29, Bacteriophage N15	436	-121	NP 046924.1	47.2
235*	317	Prophage P-EibA J protein-like protein	504	-140	AF151091.1	48.9
238	249	Bacteriophage lambda	377	-102	J02459.1	56.2
240*	444	ORF Y.-1, Bacteriophage T4 (P)	70.9	-12	NP 049798.1	45.8
243	322	Bacteriophage N15 virion	127	-27	NC 001901.1	48.2
254	365	Bacteriophage lambda	573	-161	J02459.1	57.8
266	455	*E. coli* O157:H7 genome, Bacteriophage lambda	839	0	AE005331.1	54.3
268	428	Bacteriophage lambda	730	0	NC 001416.1	57.2
289	308	*E. coli* O157:H7 genomic DNA, prophage (Sakai-VT1) inserted region	502	-140	AP000400.1	47.7
305*	1119	*E. coli* clone SauE4.D1 neonatal meningitis strain-specific DNA sequence	226	-56	AF222136	53.0
314	1285	*E. coli* O157:H7 DNA	2040	0	AP002556	50.1
317	1140	Bacteriophage N15 virion	180	-42	NC 001901.1	49.6
338	302	Bacteriophage lambda	527	-147	NC 001416.1	54.6
347*	913	unknown protein encoded by cryptic prophage CP-933M *E. coli* O157:H7 (P)	117	-25	AAG55470.1	50.0
350	1714	side tail fiber protein from lambdoid prophage RAC (P)	66.2	-10	STFR ECOLI	49.1
356	562	Bacteriophage lambda	1090	0	NC001416.1	54.6
373	995	Bacteriophage N15 virion	345	-92	NC 001901.1	43.6
375	246	*E. coli* O157:H7 genome, lambda	436	-120	AE005330.1	52.4
401	936	Bacteriophage N15 virion	204	-50	NC 001901.1	48.9
3-13	639	*E. coli* O157:H7 genome	377	-102	AE005330.1	39.5
Plasmid sequences						
230	402	*E. coli* plasmid pKL1	557	-156	ECU81610	44.3
233	1552	*E. coli* plasmid pKL1	813	0	NC 002145.1	52.0
303	1202	*E. coli* plasmid pAPEC-1 IS100 putative transposases, IS30 putative transposase, IS911 putative transposase, Tsh (*tsh*)	444	-122	AF218073.1	48.3

Table 4-3: Continued

Clone	Size (bp)	Sequences with similarity	Score	Probability (e^x)	GenBank accession no.	% GC
Plasmid sequences (continued)						
307	1003	*E. coli* plasmid pKL1	537	-150	NC 002145.1	52.2
341	980	*Enterobacter aerogenes* plasmid R751	1166	0	EAU67194	48.5
366	252	*E. coli* DNA for plasmid pColV-K30	500	-139	AJ223631.1	52.0
1-63*	1112	*Shigella flexneri* virulence plasmid pWR100	1782	0	AL391753.1	55.2
1-74*	676	*E. coli* plasmid EAF *bfp* operon	351	-94	AF119170.1	45.4
2-76*	1146	*E. coli* plasmid pO157	309	-81	NC 002128.1	39.3
3-8	353	*E. coli* clone TspE15.D1 neonatal meningitis strain-specific DNA sequence	547	-153	AF222156.1	46.2
Sequences with known function						
219*	458	acetyl-CoA:acetoacetyl-CoA transferase A subunit, *Mesorhizobium loti* (P)	225	-58	BAB50030.1	52.4
296*	203	*E. coli* outer membrane hemin receptor ChuA	379	-103	AF280396.1	49.1
410	841	*E. coli* clone SauE4.B1 neonatal meningitis strain-specific DNA sequence	297	-78	AF222130.1	45.2
1-44*	941	phosphoglycerate transporter protein *Salmonella typhimurium* (P)	301	-81	NP_461340.1	42.6
1-47*	888	*E. coli* O157:H7 genome	1461	0	AE005354.1	43.8
2-69	451	*E. coli* clone TspE15.A3 neonatal meningitis strain-specific DNA sequence	402	-110	AF222152.1	47.3
3-12	903	*E. coli* transposon Tn5396 transposase gene, inverted repeat region	61.9	-7	U04362.1	45.9
3-19*	1726	dTDP-6-deoxy-D-glucose-3,5 epimerase *Shigella boydii* (P)	168	-40	AF402312 5	32.2
Sequences without known function						
201*	1280	hypothetical protein, *Legionella pneumophila* (P)	58.2	-7	CAC33469.1	50.8
282*	537	conserved hypothetical integral membrane protein *Borrelia burgdorferi* (P)	102	-21	AAC67189.1	52.1
308*	542	*Shigella flexneri* 2a *she* pathogenicity island	289	-76	AF200692.2	52.2
315*	372	hypothetical protein T10P11.16.1 *Arabidopsis thaliana* (P)	32.7	1.9	T01095	51.9
340	1038	*E. coli* clone TspE15.G6 neonatal meningitis strain-specific DNA sequence	537	-150	AF222171.1	43.8
358*	307	putative aldolase *E. coli* O157:H7 (P)	51.6	-6	AAG59284.1	54.0
361*	510	conserved hypothetical protein *Pseudomonas aeruginosa* (P)	165	-40	AAG05047.1	48.6
381*	451	putative structural protein (partial) *E. coli* O157:H7	97.8	-20	AAG54720.1	49.7

Table 4-3: Continued

Clone	Size (bp)	Sequences with similarity	Score	Probability (e^x)	GenBank accession no.	% GC
Sequences without known function (continued)						
390	499	probable integrase, *Salmonella typhimurium*	135	-31	T03001	44.8
400*	328	*E. coli* clone TspE15.F7 neonatal meningitis strain-specific DNA sequence	418	-115	AF222167.1	55.8
453*	936	putative kinesin-like protein, *Arabidopsis thaliana* (P)	33.9	1.9	AAB70034.1	44.0
1-4*	1306	*E. coli* clone TspE15.E10 neonatal meningitis	438	-120	AF222165.1	40.0
		E. coli clone SauE15.E9 neonatal meningitis	212	-52	AF22089.1	
2-42		transketolase 1 VCA0624, *Vibrio cholerae* (P)	85.5	-16	AAF96525.1	51.4
		putative transcriptional regulator, *Neisseria meningitidis*	70.9	-11	CAB84834.1	
3-54*	1038	similar to PTS cellobiose-specific enzyme IIC *Listeria innocua* (P)	253	-66	NP 472233.1	39.2
3-72*	1175	*E. coli* O157:H7 genome	309	-81	AE005336.1	40

* Fragments screened for in the ECOR and APEC strains. The sequences with the highest homology to the fragment are listed. The homology is not usually to the entire length of the fragment. The (P) indicates that the homology listed is an amino acid sequence (BLASTX). All other homologies listed are based on (BLASTN) (Altschul, Madden *et al.*, 1997). The probability and score values are an indication of length and level of homology. As the homology of the sequences increases, the probability approaches zero. The score increases as the level and length of homology increase.

fragments. The sequences were grouped into four general types: phage sequences, plasmid sequences, sequences with known functions, and sequences without known function.

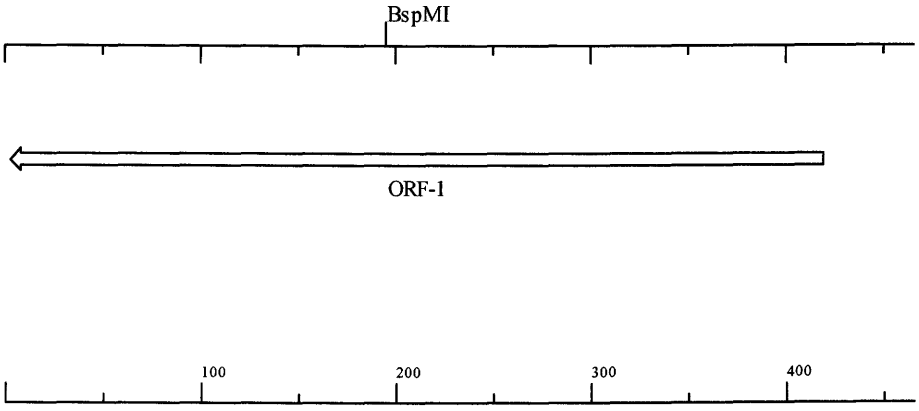
There were 29 DNA fragments with homology to phage sequences. Of the phage sequences, 28 were derived from the EC317 fragments, and one was from the EC2029 fragments. There were nine fragments with homology to lambda phage and seven with homology to phage N15. One of the fragments (101) had homology to an open reading frame (ORF) of the Vt2-Sa prophage (Miyamoto, Nakai *et al.*, 1999), but not to the phage genes for the production of Shiga toxin type 2. Another fragment (289) had homology to part of the Vt1 prophage, but not to the section that encodes Shiga toxin type 1. There were four fragments with homology to prophages of *E. coli* O157:H7 strains.

Fragment 28 had homology to phage-like sequences. It was 465 bps long and had one reading frame that did not contain termination codons (Figure 4-7), as determined from the single stranded sequence. In this reading frame, one ORF was identified which had the potential to encode a protein 154 amino acids long. The ORF shared identity with the *S. enterica* serovar Typhi hypothetical protein STY4603 (54% over 147 amino acids). The homologous region corresponded to residues 290 to 436 of the 926 residue hypothetical protein. The fragment had a reading frame that did not contain any stop codons, so it was possible that fragment 28 was a portion of a larger ORF or gene.

There were 10 fragments with homology to plasmid sequences. Six of the fragments had homology to parts of virulence plasmids. Fragment 3-8 had a 69 bp region with 98% homology to a putative reverse transcriptase from the pWR501 plasmid

Figure 4-7: Fragment 28 homology.

A)



B)

```
Frag 28: 464 QLKKTPEAAAEFAAKMQDATGTASEDMMGLFDTIQKAFYLGVDDTNMLSFFTKTSSVLKM 285
             QL+  P AAAEFAAK+QDAT T+ +DMM L D IQK FY GVD  NML  F+K SS + +
STY4603: 290 QLQMAPTAAAEFAAKLQDATQTSEKDMMNLMDVIQKGFYAGVDSGNMLQGFSKISSAMNI 349

Frag 28: 284 VNKDGLQAAQSLAPISVMMDQMGMNGESAGNALRKVIQSGLSVKKIRDVNKVMARQKLGV 105
             +NK GL+A ++ AP+ VM DQ  M GESAGNA RK+ Q+ L    I+ VN  +  +  G+
STY4603: 350 INKKGLEAVKTFAPLLVMADQGSMAGESAGNAYRKIFQAALDADNIKAVNDDLKEKGAGI 409

Frag 28: 104 QLDFTDGKGSFGGLDNMFRQLAKLRKL 24
             + +F+DGKG FGGL+NM+ QL KL+KL
STY4603: 410 KFNFSDGKGGFGGLENMYAQLEKLKKL 436
```

A) Fragment 28 restriction map and open reading frames as determined using the Clone Manager program. The open reading frames were determined using an ATG start codon and a minimum length of 100 amino acids. The frame that contains the ORF did not contain any stop codons, so the true start codon may be located on sequences flanking the fragment 28 sequence. B) The alignment of the translated fragment 28 sequence with the *S. typhi* hypothetical gene STY4603 as determined using BLASTX. The sequences shared 54% identity over 146 residues.

of *Shigella flexneri,* which is a 210 kb plasmid containing a large number of virulence factors and insertion sequences (IS) (Venkatesan, Goldberg *et al.*, 2001). The fragment had very high homology (100%, over 276 bp) to a fragment of DNA isolated from an *E. coli* strain that causes neonatal meningitis, but the function of the neonatal meningitis-specific DNA fragment is not known.

Fragment 2-76 had 96% homology to 181 bp of the pO157 plasmid. This plasmid contains many putative genes, some of which are believed to be virulence factors (Burland, Shoa *et al.*, 1998), but the region of homology was to an IS element. One of the fragments (1-74) had 100% homology to an IS element on the EAF plasmid (Bortolini, Trabulsi *et al.*, 1999), but also has a region with no significant homology to any sequence in GenBank. Fragment 1-63 was 96% homologous over 1042 bp to IS elements found on the pWR100 plasmid, which is another large *S. flexneri* virulence plasmid (Buchrieser, Glaser *et al.*, 2000). The homology of fragments 1-63, 1-74, and 2-76 to IS elements suggests that the two avian strains that were used to derive the DNA fragments were exposed to transposons. Transposons have been shown to facilitate the transfer of genes between strains, so it is possible that these fragments were adjacent to genes transferred into the avian *E. coli* strains by transposition, but it is unlikely that these IS elements have a direct effect on the virulence of the avian *E. coli* strains.

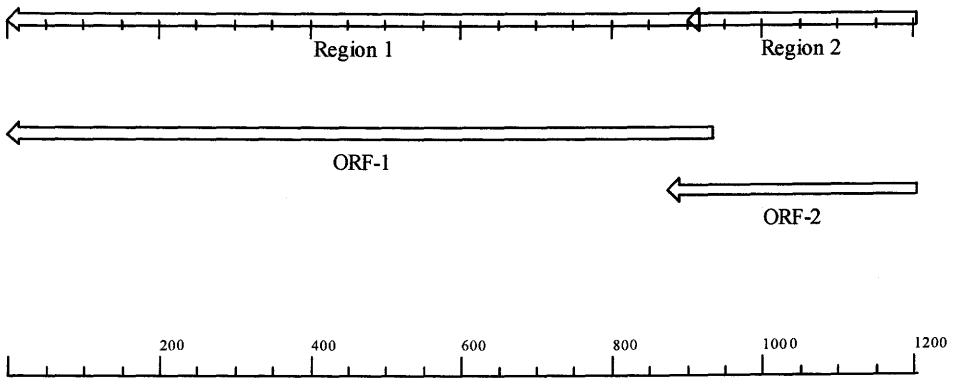
Two fragments (303 and 366) had high homology to two virulence plasmids, pColV and pAPEC-1. These fragments were homologous to two genes, *tsh* and *hbp,* which code for proteins varying by two residues. The *tsh* gene has been associated with virulent avian *E. coli* strains (Dozois, Dho-Moulin *et al.*, 2000), while the *hbp* gene codes for a hemoglobin protease (Otto, van Dooren *et al.*, 1998). The entire length of fragment 366 was 100% homologous to part of the *hbp* gene, and varied from the

sequence of the *tsh* gene by one bp. The fragment 366 sequence does not contain the entire *tsh* or *hbp* gene, so it is not known if the complete gene is present in *E. coli* EC317 or *E. coli* EC2029. Fragment 303 had two regions which were translated and shared high identity to the both Hbp and Tsh (Figure 4-8). Fragment 303 may contain part of a gene with function similar to *tsh* and *hbp*.

There were eight fragments with homology to genes with known functions, which were not found on plasmids or in phage. One of the fragments (219) shared 68% amino acid identity to the acetyl-CoA:acetoacetyl-CoA transferase A subunit (Mlr3047) from *Mesorhizobium loti*, as determined by BLASTX (Figure 4-9). This enzyme functions in the degradation of fatty acids (Sramek and Frerman, 1975; Karp, Riley *et al.*, 2000). The region of homology covered 452 of the 458 bp of fragment 219. It is possible that this fragment is part of a gene with similar function to the acetyl-CoA:acetoacetyl-CoA transferase A subunit.

Fragment 1-44 had 99% homology to 105 bp of the *pap* genes, but not to the part that codes for pili proteins. The fragment 1-44 sequence contained an ORF that coded for a protein 179 residues long, and the predicted product had 84% identity to the *S. typhimurium* protein PgtP (Figure 4-10), a phosphoglycerate transporter (Jiang, Yu *et al.*, 1988). The ORF was not large enough to contain the entire PgtP gene. It is possible that the fragment 1-44 sequences codes for a protein with similar function to PgtP. There was very high identify (98%) shared between fragment 1-47 and part of a putative fimbrial usher protein from *E. coli* O:157:H7, but this homology was in two sections (Figure 4-11). Two ORFs were identified for this fragment in frames +1 and +2. The homology at the DNA level indicated that base 317 was not in the O157:H7 sequence. The DNA homology of the fragment with the O157:H7 sequence was very high

Figure 4-8: Fragment 303 homology.
A)



B) Region 1 homology to Hbp:

```
Frag 303:  915  GLPGGVLDVNGNDLTFHKLNAADYGATLGNSSDKTANITLDYQTRPADVKVNEWSSSNRG 736
               G  GG LDVNGN LTFH+L AADYGA L N+ DK A ITLDY  R   V +N WS S +G
     Hbp:  491  GYRGGTLDVNGNSLTFHQLKAADYGAVLANNVDKRATITLDYALRADKVALNGWSESGKG 550

Frag 303:  735  TVGSLYIYNNPYTHTVDYFILKTSSYGWFPTGQVSNEHWEYVGHDQXSAQALLAXRINNK 556
               T G+LY YNNPYT+T DYFILK S+YG+FPT Q SN  WE+VGH Q  AQ L+A R N
     Hbp:  551  TAGNLYKYNNPYTNTTDYFILKQSTYGYFPTDQSSNATWEFVGHSQGDAQKLVADRFNTA 610

Frag 303:  555  GYLYHGKLLGNINFSNKATPGTTGALVMDGSANMSGTFTQENGRLTIQGHPVIHASTSQS 376
               GYL+HG+L GN+N  N+   G TGALVMDG+A++SGTFTQENGRLT+QGHPVIHA  +QS
     Hbp:  611  GYLFHGQLKGNLNVDNRLPEGVTGALVMDGAADISGTFTQENGRLTLQGHPVIHAYNTQS 670

Frag 303:  375  IANTVSSLGDNSVLTQPTSFTQDDWENRTFSFGSLVLKDTDFGLGRNATLNTTIQADNSS 196
               +A+ +++ GD+SVLTQPTSF+Q+DWENR+F+F  L LK+TDFGLGRNATLNTTIQADNSS
     Hbp:  671  VADKLAASGDHSVLTQPTSFSQEDWENRSFTFDRLSLKNTDFGLGRNATLNTTIQADNSS 730

Frag 303:  195  VTLGDSRVFIDKKDGQGTAFTLEEGTSVATKDADKSVFNGTVNLDNQSVLNINEIFNGGI 16
               VTLGDSRVFIDK DGQGTAFTLEEGTSVATKDADKSVFNGTVNLDNQSVLNIN+IFNGGI
     Hbp:  731  VTLGDSRVFIDKNDGQGTAFTLEEGTSVATKDADKSVFNGTVNLDNQSVLNINDIFNGGI 790

Frag 303:  15   QANNS 1
               QANNS
     Hbp:  791  QANNS 795
```

C) Region 2 homology to Hbp:

```
Frag 303:1202 TWTGAGIIVDKDASVNWQVNGVKGDNLHKIGEGTLVVQGTGVNEGGLKVGDGTVVLNQQA 1023
              TWTGAGI+VD   SVNWQVNGVKGDNLHKIGEGTL VQGTG+NEGGLKVGDG VVLNQQA
     Hbp: 395 TWTGAGIVVDNGVSVNWQVNGVKGDNLHKIGEGTLTVQGTGINEGGLKVGDGKVVLNQQA 454

Frag 303:1022 DSSGHVQAFSSVNIASGRPTVVLADNQQVNPDNISWGYRGG 900
              D+ G VQAFSSVNIASGRPTVVL D +QVNPD +SWGYRGG
     Hbp: 455 DNKGQVQAFSSVNIASGRPTVVLTDERQVNPDTVSWGYRGG 495
```

A) A map of fragment 303. Two ORFs identified using the Clone Manager program are indicated. The DNA sequence of fragment 303 was used to search the GenBank protein database (BLASTX) and two regions were identified as homologous to the Tsh and Hbp proteins. B) Region 1 shares 73% identity with a portion of Hbp and Tsh. C) Region 2 shares 85% identity with Hbp and Tsh.

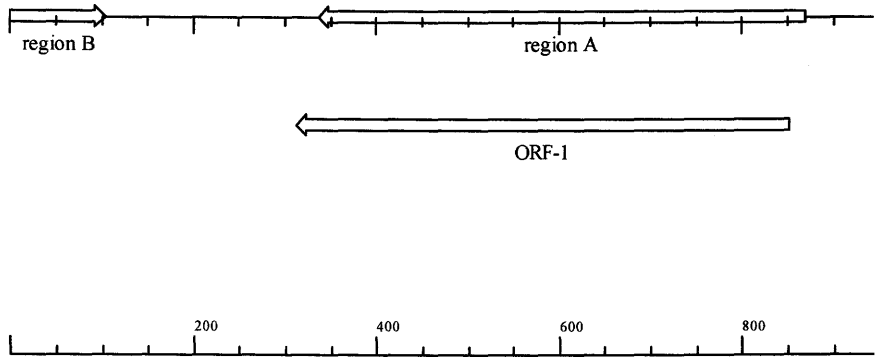
Figure 4-9: Fragment 219 homology.

```
Frag 219: 457 PRILLEQGRHGDVTWLLEQGAIGGVPLLEFQFGCASNAEAFLPSPQQFTYFQGGGFDLTL 278
              PRILLE+G+HG VTW++EQGA+GGVPLL+F+FGCASNAEA +PSP QF YFQ GGFD +L
   Sbjct: 310 PRILLEEGQHGKVTWVIEQGAVGGVPLLDFKFGCASNAEAIMPSPHQFIYFQAGGFDASL 369

Frag 219: 277 MSFLQIGADGSVNVSHLPARPHVTAGCGGFIDITSHAKRIIFSGFFNAGAQLQLEEGQLR 98
              +SFLQI   GSVNVS L ARPHVTAG GGF+DIT+ AK+I+FSGFFNAGA+L L +G +R
   Sbjct: 370 LSFLQIDRHGSVNVSKLSARPHVTAGAGGFVDITARAKKIVFSGFFNAGAKLSLADGGIR 429

Frag 219: 97  IIKEGKAKKLVQDVAHVTFSGKRAIRLGQQV 5
              I +EGK KK+V +V H++FSGKRA+  GQ +
   Sbjct: 430 IDQEGKVKKVVNEVEHISFSGKRAVAQGQDI 460
```

The sequence of fragment 219 was used to search for homology with the GenBank protein database (BLASTX). The sequence shared 68% identity with the Acetyl-CoA:acetoacetyl-CoA transferase A subunit of *Mesorhizobium loti.*

Figure 4-10: Homology of fragment 1-44

A)



B)

```
1-44: 869 YLQRFGMISWLPIYLLTVKHFSKEQMSVAFLFFEWAAIPSTLLAGWLSDKLFKGRRMPLA 690
          Y+ RFGMISWLPIYLLTVKHFSKEQMSVAFLFFEWAAIPSTLLAGWLSDKLFKGRRMPLA
PgtP: 270 YMVRFGMISWLPIYLLTVKHFSKEQMSVAFLFFEWAAIPSTLLAGWLSDKLFKGRRMPLA 329

1-44: 689 MICMALIFICLIGYWKSESLFMVTIFAAIVGCLIYVPQFLASVQTMEIVPSFAVGSAVGL 510
          MICMALIF+CLIGYWKSESL MVTIFAAIVGCLIYVPQFLASVQTMEIVPSFAVGSAVGL
PgtP: 330 MICMALIFVCLIGYWKSESLLMVTIFAAIVGCLIYVPQFLASVQTMEIVPSFAVGSAVGL 389

1-44: 509 RGFMSYIFGASLGTSLFGIMVDHIGWHXXXXXXXXXXXXXXXXCWLSHRGAIELERHR 336
          RGFMSYIFGASLGTSLFG+MVD +GW+                C+LSHRGA+ELER R
PgtP: 390 RGFMSYIFGASLGTSLFGVMVDKLGWYGGFYLLMGGIVCCILFCYLSHRGALELERQR 447
```

A) The map of fragment 1-44 showing the two regions with homology to sequences in the GenBank database and one ORF. The translation of region A shared 84% identity to the *S. typhimurium* protein PgtP that is a phosphoglycerate transporter protein. Region B had 99% homology to a non-coding region of the pap gene cluster of *S. typhimurium.* The ORF codes for a putative protein with homology to the PgtP protein.
B) The alignment of the translated fragment 1-44 sequence to the sequence of PgtP as determined by BLASTX analysis. Area of the alignment indicated as region A on the map.

Figure 4-11: Fragment 1-47 alignment to a putative fimbrial usher.

A)

```
Frame = +2
1-47: 317 YAYLNLQSGINIGSGRLRDNSTWSYNSGSSNSSDSNKWQHINTSAERDIIPLRSRLTVGD 496
          YAYLNLQSGINIGS RLRDNSTWSYNSGSSNSSDSNKWQHINTSAERDIIPLRSRLTVGD
O157: 211 YAYLNLQSGINIGSWRLRDNSTWSYNSGSSNSSDSNKWQHINTSAERDIIPLRSRLTVGD 270

1-47: 497 SYTDGDIFDSVNFRGLKINSTEAMLPDSQHGFAPVIHGIARGTAQVSVKQNGYDVYQTTV 676
          SYTDGDIFDSVNFRGLKINSTEAMLPDSQHGFAPVIHGIARGTAQVSVKQNGYDVYQTTV
O157: 271 SYTDGDIFDSVNFRGLKINSTEAMLPDSQHGFAPVIHGIARGTAQVSVKQNGYDVYQTTV 330

1-47: 677 PPGPFTIDDINSAANGGNLQVTIKEADGSIQTLYVPYSSVPVLQRAGYTRYALAMGEYRS 856
          PPGPFTIDDINSAANGG+LQVTIKEADGSIQTLYVPYSSVPVLQRAGYTRYALAMGEYRS
O157: 331 PPGPFTIDDINSAANGGDLQVTIKEADGSIQTLYVPYSSVPVLQRAGYTRYALAMGEYRS 390

1-47: 857 GNNLQSTPKF 886
          GNNLQS+P+F
O157: 391 GNNLQSSPRF 400

Frame = +1
1-47: 1   TDLLVSLGIKKSALLDNKEHSADKHVPDNSACTPLQDRLADASSEFDVGQQHLSLSVPQI 180
          TDLLVSLGIKKSALLDNKEHSA+KHVPDNSACTPLQDRL DAS+EFDVGQQHLSLSVPQI
O157: 106 TDLLVSLGIKKSALLDNKEHSAEKHVPDNSACTPLQDRLVDASTEFDVGQQHLSLSVPQI 165

1-47: 181 YVGRMARGYVSPDLWEEGINAGLLNYSFNGNSINNRSNHNAGKSN 315
          YVGRMARGYVSPDLWEEGINAGLLNYSFNGNSINNRSNHNAGKSN
O157: 166 YVGRMARGYVSPDLWEEGINAGLLNYSFNGNSINNRSNHNAGKSN 210
```

B)

```
Frame = +1
1-47: 1   TDLLVSLGIKKSALLDNKEHSADKHVPDNSACTPLQDRLADASSEFDVGQQHLSLSVPQI 180
          TDLLVSLGIKKSALLDNKEHSA+KHVPDNSACTPLQDRL DAS+EFDVGQQHLSLSVPQI
O157: 106 TDLLVSLGIKKSALLDNKEHSAEKHVPDNSACTPLQDRLVDASTEFDVGQQHLSLSVPQI 165

1-47: 181 YVGRMARGYVSPDLWEEGINAGLLNYSFNGNSINNRSNHNAGKSNYAYLNLQSGINIGSG 360
          YVGRMARGYVSPDLWEEGINAGLLNYSFNGNSINNRSNHNAGKSNYAYLNLQSGINIGS
O157: 166 YVGRMARGYVSPDLWEEGINAGLLNYSFNGNSINNRSNHNAGKSNYAYLNLQSGINIGSW 225

1-47: 361 RLRDNSTWSYNSGSSNSSDSNKWQHINTSAERDIIPLRSRLTVGDSYTDGDIFDSVNFRG 540
          RLRDNSTWSYNSGSSNSSDSNKWQHINTSAERDIIPLRSRLTVGDSYTDGDIFDSVNFRG
O157: 226 RLRDNSTWSYNSGSSNSSDSNKWQHINTSAERDIIPLRSRLTVGDSYTDGDIFDSVNFRG 285

1-47: 541 LKINSTEAMLPDSQHGFAPVIHGIARGTAQVSVKQNGYDVYQTTVPPGPFTIDDINSAAN 720
          LKINSTEAMLPDSQHGFAPVIHGIARGTAQVSVKQNGYDVYQTTVPPGPFTIDDINSAAN
O157: 286 LKINSTEAMLPDSQHGFAPVIHGIARGTAQVSVKQNGYDVYQTTVPPGPFTIDDINSAAN 345

1-47: 721 GGNLQVTIKEADGSIQTLYVPYSSVPVLQRAGYTRYALAMGEYRSGNNLQSTPKF 885
          GG+LQVTIKEADGSIQTLYVPYSSVPVLQRAGYTRYALAMGEYRSGNNLQS+P+F
O157: 346 GGDLQVTIKEADGSIQTLYVPYSSVPVLQRAGYTRYALAMGEYRSGNNLQSSPRF 400
```

The homology of the fragment 1-47 to protein sequences was determined by BLASTX search. The sequence with the highest homology to fragment 1-47 was a putative fimbrial usher gene from *E. coli* O157:H7. A) Alignment of the translation of fragment 1-47 with the *E. coli* O157:H7 gene. Parts of two reading frames (+1 and +2) of the 1-47 sequence were homologous to the gene. The overall identity shared between the sequences was 98%. B) Alignment of the edited fragment 1-47 sequence with the O157:H7 gene. The 1-47 sequence was edited to remove one base. The entire +1 frame of translated sequence was homologous to the O157:H7 gene. The identity shared between the sequences was 97%.

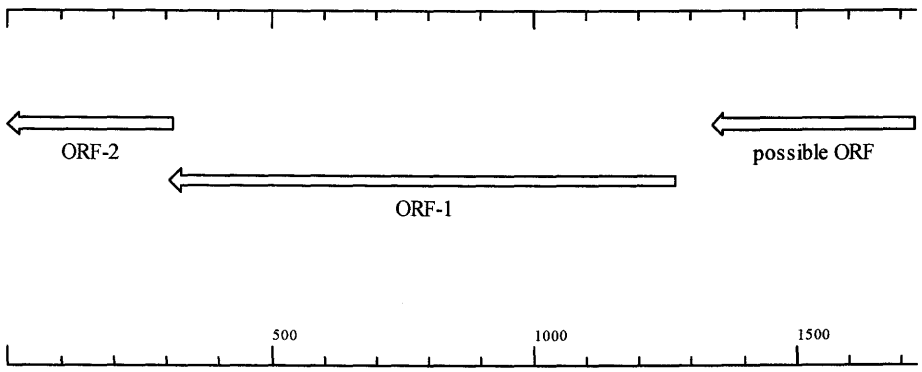
(98% over the entire fragment sequence), suggesting that the base is an error due to sequencing only the single strand, or an error introduced during PCR. When the sequence was edited to remove the base, the fragment did not have any stop codons in reading frame +1 and the homology to the putative usher protein was contiguous (Figure 4-11). Fimbrial function in adherence, which is necessary for colonization. Usher proteins are not structural proteins, but aid in the folding and assembly of other proteins (Sussman, 1997). The strong homology of the fragment 1-47 sequence to part of a putative fimbrial usher gene suggests that the fragment may be part of a gene that functions as a fimbrial usher and may be involved in the virulence of avian *E. coli* strains.

The entire sequence of fragment 296 had 98% homology to the *chu*A gene, which belongs to a heme uptake pathway (Torres and Payne, 1997). There was also a fragment (3-12) with a 231 bp region that shares 82% identify with a transposase by BLASTX analysis, but the region of homology was not to the entire gene or to the entire sequence of the fragment. Fragment 3-12 did not have any ORFs so it is likely that this sequence does not code for a protein, though it might have been a transposon.

Fragment 3-19 was a large fragment (1726 bp) with homology to three different proteins involved in lipopolysaccaride (LPS) synthesis (Li and Reeves, 2000). Some serotypes of LPS have been shown to have a role in virulence due to their function in serum resistance and protection from phagocytosis (Burns and Hull, 1998; Burns and Hull, 1999). Fragment 3-19 had two ORFs and an area without stop codons (Figure 4-12). The third region might be an ORF since the start codon might be located off the three prime end of the fragment.

Figure 4-12: Fragment 3-19 homology

A)



B)

```
ORF1: 74  YETIIFIILTCSFFIKARFLANYSRNFSGLTLFFIYYASVALWVLDYTQFRNGLCISILM 133
          ++TII  I   +  +    + N++++F   +   +    + LW +     R  L +SI++
 Rfc: 83  FQTIIIFIAAVNVIL----ILNFAKHFENGSFVIVAIMCMFLWSVYVEAIRQALALSIVI 138

ORF1: 134 FSVYYLFINKPTYFYFSVLCAIATHWSALPFLLLYPFVYSTKIRRL 179
          F ++ LF+ +   F   VL A   H +AL   LL   ++S K+ ++
 Rfc: 139 FGIHSLFLGRKRKFITLVLFASTFHITALICFLLMTPLFSKKLSKI 184
```

C)

```
ORF2: 1   MCKAKVLAIIVTYNPEIIRLTECINSLAPQVERIILVDNGSNN--SDXXXXXXXXXXXXX 58
          M K KV AII+TYNP++  L E   SL  QV++IIL+DN S N
RfbQ: 1   MIKKKVAAIIITYNPDLTILRESYTSLYKQVDKIILIDNNSTNYQELKKLFEKKEKIKIV 60

ORF2: 59  XXXXXXGIAFAQNHGVKKGLEAKEFDYLFFSDQDT 93
                G+A AQN G+   ++   + Y    DQD+
RfbQ: 61  PLSDNIGLAAAQNLGLNLAIK-NNYTYAILFDQDS 94
```

D)

```
 ORF: 1   LRGLHYQLAPYAQAKLVRCVVGQVFDVAVDLRKNSPTFKKWFGITLSAENKRQLWIPEGF 60
          LRGLHYQL PYAQ KLVRC VG+VFDVAVD+RK+SPTF KW G+ LSAENKRQLWIPEG
RmlC: 58  LRGLHYQLEPYAQGKLVRCAVGEVFDVAVDIRKSSPTFGKWVGVNLSAENKRQLWIPEGL 117

 ORF: 61  AHGFLVTSDEAEFIYKTTNYYAPGHQQAIIYNDPILNIDWPFCSSALSLSQKDQEAKLFS 120
          AHGFLV S+ AEF+YKTTNYY P   + II++DP +++ WP       LS KD++ K+F
RmlC: 118 AHGFLVLSETAEFLYKTTNYYHPESDRGIIWDDPDIDVKWPLSIHKPILSIKDEKQKMFK 177

 ORF: 121 ELLDSE 126
          E++  E
RmlC: 178 EMIALE 183
```

A) Fragment 3-19 indicating the position of the two ORFs and the possible ORF. B) The alignment of the translation of ORF-1 with the O antigen polymerase of *E. coli.* The sequences share 24% identity and 48% positives. C) The alignment of the translation of the ORF-2 sequence with rhamnosyl transferase I (RfbQ) of *S. dysenteriae.* The sequences share 35% identity and 47% positives. D) The alignment of the translation of the possible ORF with dTDP-6-deoxy-D-glucose-3,5 epimerase (RmlC) of *S. boydii.* The sequences share 64% identity and 77% positives.

The ORF1 was translated and the 320 amino acid sequence was analyzed for homology to known protein sequences. It shared 22% identity (44% positives) over 101 residues with the *E. coli* O antigen polymerase (Rfc, Figure 4-12-B) (Lukomski, Hull *et al.*, 1996). The Rfc protein was 395 residues long, so the area of homology did not cover the entire length of the protein.

ORF2 shared 35% identity with the rhamnosyl transferase I (RfbQ) of *S. dysenteriae* (Sturm, Jann *et al.*, 1986) (Figure 4-12-C). The RfbQ gene was 303 residues long, and the homology shared with ORF2 was to the first 95 residues. ORF2 ended at the 5' end of the fragment 3-19 sequence so it was possible that the ORF extends into the region flanking the fragment. If this is the case, the second ORF may be a homologue of the *rfbQ* gene.

The translated sequence of the third region shared 66% identity to 126 residues of the dTDP-6-deoxy-D-glucose-3,5 epimerase (RmlC) of *Shigella boydii* (Figure 4-12-D). The gene was part of the O-antigen gene cluster responsible for different serotypes of LPS (Wang, Qu *et al.*, 2001). The RmlC protein was 189 residues long and the area of homology was 126 residues at the C-terminal end of the sequence. There were 57 residues missing, which could have been coded on the sequence flanking fragment 3-19. As well, there are six residues missing from the end of the translated 3-19 sequence.

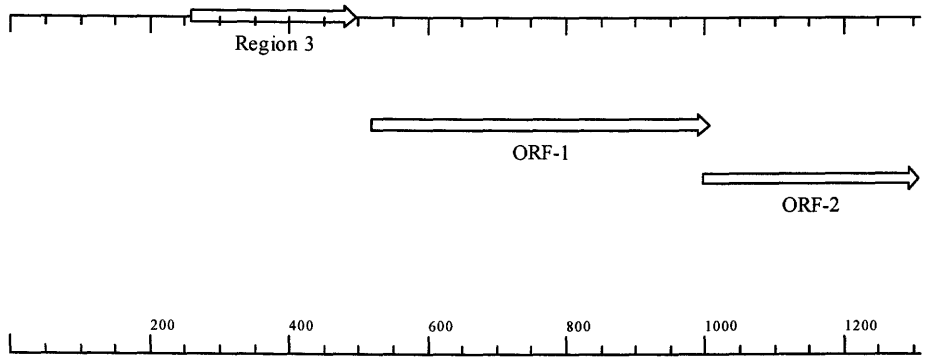
If these three regions from fragment 3-19 were expressed, they could be part of the LPS synthesis operon for the *E. coli* strain EC2029. This theory seems likely since the ORFs are in the same orientation, and because they are homologues to genes that are involved in LPS synthesis in other strains. There are many different types of LPS and each requires different genes for their synthesis (Reeves, Hobbs *et al.*, 1996), so it is possible that these genes make up part of an LPS operon.

There were 15 fragments grouped as "unknown" sequences. These sequences had either small regions of homology or low levels of homology to known sequences or had homology to putative genes or ORFs. Fragment 1-4 had high homology (96% over 245 bp, and 98% over 123 bp) to two fragments of DNA with unknown function which had been isolated from neonatal meningitis *E. coli* strains. The fragment also contained two ORFs with homology to two putative *S. typhimurium* genes (Figure 4-13). ORF2 had homology at the protein level (77% identity to 94 residues) to the putative arylsulfatase (gene STM0035) which was 497 residues long. Arylsulfatases are involved in sulfatide metabolism (Murooka, Higashiura *et al.*, 1978; Karp, Riley *et al.*, 2000). An ORF from *S. typhimurium* (STM0034) shared identity with two parts of the fragment 1-4 sequence (53% identity to 191 residues, and 53% to 79 residues), including the ORF1 sequence. STM0034 codes for a protein 253 amino acids long and is adjacent to the STM0035 ORF in the *S. typhimurium* LT2 chromosome (McClelland, Sanderson *et al.*, 2001). The fact that there are three regions with homology to LPS synthesis genes suggests that fragment 3-19 was part of an LPS synthesis operon. The level of homology was quite low, so it is possible that these ORFs have other function if they are expressed.

Fragment 3-54 was a 1040 bp fragment with three ORFs (Figure 4-14), as determined using Clone Manager on the single stranded fragment sequence. Putative promoters were predicted for both ORFs 1 and 3 using the Neural Network Promoter Prediction program (http://www.fruitfly.org/seq_tools/promoter.html) (Reese and Eeckman, 1995). ORFs 1 and 3 were homologous to different genes involved in cellobiose metabolism. ORF 1 had 67% homology to the first 178 of the 450 amino acid PTS cellobiose-specific enzyme IIC. The promoter for ORF 1 was shifted three bases from the ORF start codon predicted using Clone Manager. ORF 3 had 58% homology

Figure 4-13: Fragment 1-4 sequence analysis.

A)



B)

```
    1-4: 1024 MQKTLMASLIGLAVCTGNAFNPVVAAETKQPNLVIIMADDLGYGDLATYGHQIVKTPNID 1203
              M++T++AS+IGLA+C G   +   AA  K+PNLVII+ADDLGYGDLATYGH+IVKTPNID
STM0035: 1    MKRTVVASMIGLALCAGCVLSTAQAATAKRPNLVIILADDLGYGDLATYGHRIVKTPNID 60

    1-4: 1204 RLAQEGVKFTDYYAPAPLSSPSRAGLLTGRMPFR 1305
              +LAQEGVKFTDYYAPAPL SPSRAGLLTGRMPFR
STM0035: 61   KLAQEGVKFTDYYAPAPLCSPSRAGLLTGRMPFR 94
```

C)

```
    1-4: 432  ILKKPSFTHFMKAMCLTMP*PGRNHTVSAMTHVRLFD-SNMTFFGKIYGQWDNSWGDDLD 608
           +  +  F  F +   L  P  GRNH    + H R+ + SN T   K+Y   +N +GD+L+
STM0034: 63   VFDRGEFASFYEMNVLNHPVEGRNHVTQFLGHYRVVEGSNFTAMMKLYMSMENKFGDELN 122

    1-4: 609  MFYGLGYLGWNGSWGFFKPYIGLHNQSGDYVSAKYGQTNGWNGYVVGWTAVLPFTLFDEK 788
              M YG+GYLG     GF KPY  +HN S DY S KYGQ  G+NGYV+GW A   F +F+EK
STM0034: 123  MMYGVGYLGLTSPSGFIKPYFAVHNLSNDYTSKKYGQATGFNGYVLGWVAAYNFDMFNEK 182

    1-4: 789  FVLSNWNEIELDRNDAYAEQQFGRNGLNGGLTVAWKFYPRWKASVTWRYFDNKLGYDGFG 968
              FV+SNWNE+E+DRNDAYAEQQ G  GLNG +T  WKF PRW ASV++RYF+NKLGYDG+G
STM0034: 183  FVISNWNEVEMDRNDAYAEQQGGTTGLNGAVTFTWKFMPRWTASVSYRYFNNKLGYDGYG 242

    1-4: 969  DQMIYMLGYDF 1001
              D+M Y++G++F
STM0034: 243  DRMNYLIGFEF 253
```

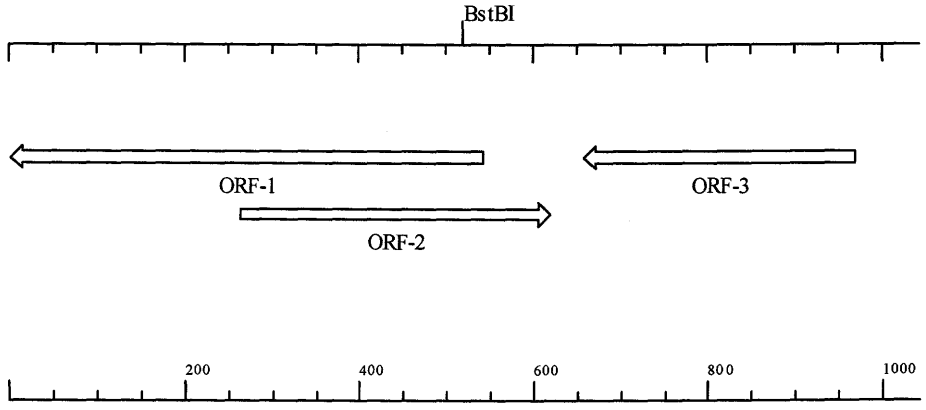
D)

```
    1-4: 260 LLIACALTTIFYHLPAYSSLEYKGSFGSINAGYADWNSGFINTHRGEVWKVTADFGVNFK 439
           LL AC L + F    A    E K  +G ++ GY DWN GF+N  RGEVWK  ADFG  F
STM0034: 10  LLAACVLFSCFNGQAA----ELKRVYGKLSFGYGDWNKGFVNVDRGEVWKAVADFGAVFD 65

    1-4: 440 EAEFYSFYESNVLNHAVAG 496
             EF SFYE NVLNH V G
STM0034: 66  RGEFASFYEMNVLNHPVEG 84
```

A) The map of the fragment 1-4 sequence indicating the two ORFs that were identified using Clone Manager. B) The alignment of the translation of fragment 1-4 to a putative arylsulfatase of *S. typhimurium*. The two sequences share 77% identity. The fragment 1-4 sequence in this alignment corresponds to the ORF-2 sequence. C) The alignment of the putative *S. typhimurium* protein STM0034 with one part of the translation of fragment 1-4. The sequences share 53% identity. The fragment 1-4 sequence in this alignment corresponds to the ORF-1 sequence. D) The alignment of the putative *S. typhimurium* protein STM0034 with the second part of the translation of fragment 1-4. The sequences share 53% identity. The fragment 1-4 sequence in this alignment corresponds to the region 3 sequence.

Figure 4-14: Fragment 3-54 homology.

A)



B)

```
  ORF-1: 1   MGLMASFERGMERFLVPVAIKLNSQKHVAAVRDGFVFTFPIIMASSLIILINFAILSPDG 60
             M +M+ FE GMER LVPVA KLNSQ+H+AA+RD F+  FP+IMA S+I LINFA+LSPDG
lin2905: 1   MSIMSKFEHGMERVLVPVANKLNSQRHIAAIRDAFILVFPLIMAGSIITLINFAVLSPDG 60

  ORF-1: 61  FIAGLLHLNSIFPNLEKAQAIFTPVMNGSVNIMSIMIAFLVARNVAISYEQDDLLCGLTA 120
             FIA +L L  IFPNL  AQA+F+PVM GS NIM+I+I FLVARN+AI ++QDDLLCGLT+
lin2905: 61  FIAKILFLGKIFPNLADAQAVFSPVMQGSTNIMAILIVFLVARNLAIFFKQDDLLCGLTS 120

  ORF-1: 121 IGAFFIVYTPYQMIDGQAFLTTKYLGAQGLFVAVIVALITSEIFCRLARNPKITITMP 178
             IGAFFIVYTPY ++D  +++T K+LGAQGLFVA+IVA+IT E+F RLAR+P++ I MP
lin2905: 121 IGAFFIVYTPYTVVDNASYMTIKFLGAQGLFVAIIVAIITGEVFSRLARSPRLMIKMP 178
```

C)

```
  ORF-3: 1   MKKILLVCAAGMSTSMLVKRMIDHATAISLEVNISALAIAEAKGKIKNNEVDVVLLGPQV 60
             MKKILLVCAAGMSTS+LV +M  HAT+I  E+ I AL ++EA   +  +++D+V+LGPQV
lin2906: 1   MKKILLVCAAGMSTSLLVTKMKAHATSIGEEIEIEALPVSEASNVV--DKMDIVMLGPQV 58

  ORF-3: 61  RFQKPEIEAVAQGKMPVAVIEMKDYGTMNGQAVLEFA 97
             R+QKP+++ + QG++PV VI+MKDYG +NG+AVLE A
lin2906: 59  RYQKPQVDELVQGRIPVIVIDMKDYGMLNGKAVLEKA 95
```

A) Fragment 3-54 restriction map and open reading frames as determined using the Clone Manager program. The open reading frames were determined using an ATG start codon and a minimum length of 100 amino acids. B) The alignment of the translation of ORF-1 with the *L. innocua* gene (lin2906) similar to PTS cellobiose-specific enzyme IIC. The sequences share 67% identity. C) The alignment of the translation of ORF-3 with the *L. innocua* gene (lin2905) similar to PTS cellobiose-specific enzyme IIB. The sequences share 58% identity.

to 95 of 100 of the amino acids in the PTS cellobiose-specific enzyme IIB. The promoter for ORF 3 used the same ATG start codon as the ORF. The fact that these two ORFs had homology to proteins in the same metabolic pathway, and were in the same orientation suggests that they may be part of an operon. These putative genes may have been transcribed together, possibly with other genes contained on the sequences flanking fragment 3-54.

ORF 2 from fragment 3-54 coded for a 132 amino acid peptide with low homology to a steroid sulfatase from *Mus musculus* (30% homology over 72 amino acids). ORF 2 did not have a predicted promoter, so it does not seem likely that ORF 2 is a gene, but the ORF has not been tested to determine if it is expressed.

Fragment 308, which was one of the fragments without known function, had homology to sequences from two pathogenicity islands. There was homology to two sections of the *S. flexneri* 2a *she* pathogenicity island: one section was 41 bp with 100% homology, and the other section was 236 bp with 91% homology. Both sections were part of the ORF8 of the *she* island which did not have a proposed function (Al-Hasani, Rajakumar *et al.*, 2001). The fragment was also homologous to the LEE II island. The homology was to three parts of the ST55 ORF (91% over 216 bp, 93% over 145 bp, and 89% over 139 bp). ST55 did not have a proposed function. The fragment 308 sequence did not have an ORF with a start codon, but there was a 390 bp region without stop codons. This entire region was translated and a BLASTP search was done. The amino acid sequence had some homology to the ST55 sequence (86% identity over 52 residues and 88% identity over 36 residues), but it also had homology to some transposases (28% identity over 107 residues). Fragment 308 did not have an ORF with a start codon, and

the sequence was not homologous to any with known function, so it is not likely to code for a protein.

Some of the subtractive hybridization fragments had homology to sequences from other pathogenic *E. coli* strains. Eleven fragments from all four groups had high homology to DNA from *E. coli* O157:H7 strains (Hayashi, Makino *et al.*, 2001; Perna, Plunkett *et al.*, 2001). Of these fragments, seven had homology with phage sequences, one had homology to putative aldolases, one had homology to a flagellin protein, and another had homology to a putative fimbrial usher protein. One of the fragments with homology to *E. coli* O157:H7 DNA was also homologous to an IS element. There were eight fragments with high homology to sequences from neonatal meningitis *E. coli* strains (Bonacorsi, Clermont *et al.*, 2000). Fragment 1-4 contained two of the neonatal meningitis sequences. Fragment 410 had homology to a phage protein. Fragment 3-8 exhibited a high degree of homology to a meningitis specific sequence and to the pWR501 virulence plasmid of *S. flexneri*. The function for six of the neonatal fragments was unknown.

In summary, the fragments could be divided into four groups based on sequence homology: phage sequences, plasmid sequences, other sequences with known function, and sequences with unknown function. In the phage sequences there were two fragments that had homology to parts of two Shiga toxin converting phage, though the homology was not to the toxin genes. The fragments with homology to plasmid sequences included six with homology to virulence plasmids. Three of these fragments had homology to IS elements which have been implicated in the horizontal transfer of genes between species. Two of the fragments had homology to the *tsh* gene previously associated with avian *E. coli* strains (Dozois, Dho-Moulin *et al.*, 2000). Of the

fragments with known function, three had homology to genes that may have a function in virulence. These functions were heme use, fimbrial assembly, and LPS synthesis. The final group of fragments had homology to sequences with unknown function, or did not have homology to sequences in GenBank. It is difficult to predict if the fragments in this group have any function, so the role of these fragments in the genome of the avian *E. coli* strains is difficult to predict.

4.4.5 Screen for the Unique DNA Fragments in Avian *E. coli* Isolates

A collection of 41 avian *E. coli* isolates were screened for the presence of selected fragments. From the collection of 62 fragments, 25 were chosen for further use based on the homology of the sequences (Table 4-3). The fragments chosen included 12 without known function, five with homology to phage, three with homology to virulence plasmids, and five with homology to sequences with known functions. These functions were LPS synthesis, phosphoglycerate transport, flagellar usher, fatty acid degradation, and heme uptake.

The avian *E. coli* strains used were primarily from a collection isolated from cellulitis lesions in broilers from ten flocks located in Saskatchewan, Canada (Ngeleka, Kwaga *et al.*, 1996). Two isolates from turkeys with systemic lesions were also used. Both turkey isolates are capable of causing cellulitis in a three-week-old broiler model (Allan, van den Hurk *et al.*, 1993). Colony blot using strains EC317, EC2029, and *E. coli* K-12 as controls determined the presence of the fragments in the avian isolates (Table 4-4). Nine of the 25 fragments used were from the EC2029 library and 16 were from the EC317 library. The EC2029 fragments were found in 12 to 45% of the avian

Table 4-4: The presence of selected DNA fragments in a collection of *E. coli* strains isolated from poultry.
The + indicates the presence of DNA fragments as determined by colony blot analysis. The - indicates a that the fragment is absent from the strain. The cellulitis isolates were previously analyzed by multilocus enzyme electrophoresis (MLEE) (Ngeleka, Kwaga *et al.*, 1996) and grouped by electrophoretic type (ET) except for the last group of strains that have not been typed (nd). The *E. coli* strains EC317 and EC2029 served as the positive controls for the dot blots. *E. coli* K-12 strain MG 1655 was used as the negative control.

ET	Strains	*E. coli* EC2029 fragments									*E. coli* EC317 fragments															
		1 4	1 4	1 4	1 6	1 7	2 7	3 1	3 5	3 7	2 0	2 0	2 1	2 3	2 4	2 8	2 9	3 0	3 0	3 1	3 4	3 5	3 6	3 8	4 0	4 5
		1 4	1 4	1 7	1 3	1 4	1 6	1 9	1 4	1 2	2 8	2 1	2 9	3 5	4 0	2 2	2 6	5	8	5	7	8	1	1	0	3
1	421	-	-	-	+	-	+	-	+	-	+	-	-	-	-	-	-	+	+	-	+	-	-	-	-	-
	422	-	-	-	+	-	+	-	+	-	+	-	-	-	-	-	-	+	+	-	+	-	-	-	-	-
	423	-	-	-	+	-	+	-	+	-	+	-	-	-	-	-	-	+	+	-	+	-	-	-	-	-
	424	-	-	-	+	-	+	-	+	-	+	-	-	-	-	-	-	+	+	-	+	-	-	-	-	-
	471	-	-	-	+	-	+	-	+	-	+	-	-	-	-	-	-	+	+	-	+	-	-	-	-	-
	475	-	-	-	+	-	-	-	+	-	+	-	-	-	-	-	-	+	+	-	+	-	-	-	-	-
	476	-	-	-	+	-	-	-	+	-	+	-	-	-	-	-	-	+	+	-	+	-	-	-	-	-
2	420	-	-	-	-	-	-	-	+	-	+	-	-	+	-	-	-	+	+	-	-	-	-	-	-	-
	435	-	-	-	-	-	-	-	+	-	+	-	-	+	-	+	-	+	+	-	-	-	-	-	-	-
	436	-	-	-	-	-	-	-	+	-	+	-	-	+	-	-	-	+	+	-	-	-	-	-	-	-
	437	-	-	-	-	-	-	-	+	-	+	-	-	+	-	-	-	+	+	-	-	-	-	-	-	-
3	447	-	+	+	-	+	-	-	-	+	+	-	+	+	-	+	+	+	-	-	+	+	+	+	+	+
4	430	-	+	+	-	+	+	-	-	+	+	-	+	+	-	+	+	+	+	+	+	+	+	+	-	+
5	477	+	+	+	-	+	-	+	+	+	+	-	+	+	-	+	+	+	+	-	-	+	+	+	+	-
	479	+	+	+	-	+	-	+	+	+	+	-	+	+	-	+	+	+	+	-	-	+	+	+	+	-
6	443	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
7	480	-	-	+	-	+	-	+	+	-	-	-	-	-	-	+	+	-	-	-	+	+	+	+	+	-
8	425	-	-	+	-	+	+	-	-	+	-	-	+	-	-	+	+	-	+	-	-	+	+	+	-	-
	426	-	-	+	-	+	+	-	-	+	-	-	+	-	-	+	+	-	+	-	-	+	+	+	-	-
	427	-	-	+	-	+	+	-	-	+	-	-	+	-	-	+	+	-	+	-	-	+	+	+	-	-
	428	-	-	+	-	+	+	-	-	+	-	-	+	-	-	+	+	-	+	-	-	+	+	+	-	-
	448	-	-	+	-	+	+	-	-	+	-	-	+	-	-	+	+	-	+	-	-	+	+	+	-	-
	450	-	-	+	-	+	+	-	-	+	-	-	+	-	-	+	+	-	+	-	-	+	+	+	-	-
	470	-	-	+	-	+	+	-	-	+	-	-	+	-	-	+	+	+	+	-	-	+	+	+	-	-
9	478	-	-	+	-	-	+	-	-	-	-	-	-	+	-	-	+	-	+	-	+	+	+	+	-	-
10	441	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	+	-	-	+	+	+	-	-
	444	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	+	-	-	+	+	+	-	-
	445	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	+	-	-	+	+	+	-	-
11	473	-	+	-	+	+	+	-	-	-	-	-	-	+	-	+	+	-	+	-	-	+	+	+	-	-
12	429	-	+	-	-	+	+	-	-	-	-	-	-	+	-	+	+	-	+	-	+	+	+	+	+	-
13	449	-	+	-	-	+	+	-	-	-	-	-	+	+	-	+	+	-	+	-	+	+	+	+	-	-
	469	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	+	+	+	-	+	+	+	+	-	-
	472	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	+	-	+	+	+	+	-	-
14	438	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	+	+	+	-	-	+	-	+	-	-
15	439	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	+	-	-	+	-	+	-	-
	440	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	+	-	-	+	-	+	-	-
	442	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	+	-	-	+	-	+	-	-
16	474	-	+	+	+	-	+	-	-	-	-	-	-	+	-	-	+	-	+	-	-	+	-	+	-	-
17	419	-	+	-	+	-	-	+	-	-	-	-	-	+	-	-	+	-	+	-	-	+	-	-	-	+
nd	222	+	-	+	-	+	-	-	+	+	-	-	+	+	-	-	+	-	+	-	-	+	+	+	+	-
nd	234	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
nd	317	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
nd	2029	+	+	+	+	+	+	+	+	+	+	-	+	+	-	+	+	+	+	+	-	+	+	+	+	+
nd	K-12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

isolates tested. The EC317 fragments were found in 2 to 93% of the strains. The overall distribution of the EC317 fragments and EC2029 fragments in the avian isolates was significantly different (P=0.0005), with the fragments from the EC317 library found more frequently than from the EC2029 library.

4.4.6 Screen for the Unique DNA Fragments in the ECOR Strain Collection

The ECOR strains were screened for the presence of the selected fragments used above. The ECOR collection consists of 72 strains isolated from mammals from diverse geographic locations. The strains were originally assembled to represent the genetic diversity of the *E. coli* species, as determined by multilocus enzyme electrophoresis (Ochman and Selander, 1984). All the strains were isolated from feces of mammals, except for 10 strains isolated from women with urinary tract infections (UTI). The UTI isolates were chosen to represent pathogenic strains from each major cluster of *E. coli* strains. Each of the urinary tract isolates has the same MLEE pattern as one of the fecal isolates. The presence of the 25 fragments in the ECOR strains was determined by colony blot (Table 4-5). The fragments were found in 10 to 97% of the strains. The overall distribution of the fragments in the ECOR strains was significantly greater for the EC317 fragments than for the EC2029 fragments ($P< 0.0001$). The distribution of the fragments in the UTI strains did not differ significantly from the other strains in the ECOR collection ($P= 0.51$).

The ECOR reference collection of strains contains only mammalian isolates, and only 10 of the 72 strains were associated with disease. The avian strain collection used was composed entirely of isolates from poultry lesions. The occurrence of the fragments

Table 4-5: The presence of selected DNA fragments in the ECOR collection of *E. coli* strains.

	E. coli EC2029 fragments									*E. coli* EC317 fragments														
	1	1	1	1	1	2	3	3	3	2	2	2	2	2	2	3	3	3	3	3	3	4	4
	1	4	4	6	7	7	1	5	7	0	1	3	4	8	9	0	0	1	4	5	6	8	0	5
Strains	4	4	7	3	4	6	9	4	2	8	1	9	5	0	2	6	5	8	5	7	8	1	1	0	3
1*	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-
5	-	-	-	-	+	-	-	-	-	-	-	-	+	+	-	-	-	+	+	+	-	-	-	+	-
8	-	-	-	-	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
10	-	-	-	+	+	+	+	+	+	-	-	-	+	-	-	-	-	+	-	+	-	-	-	-	-
11*	-	-	-	-	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	+
25	-	-	-	-	+	+	-	+	+	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-
2	-	+	-	+	+	+	-	-	+	-	-	-	+	-	-	-	+	+	-	+	-	-	-	-	-
3	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	+	+	-	+	-	-	-	-	-
9	-	-	-	-	+	+	-	-	+	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-
12	-	-	-	-	+	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	+	-
2	-	-	-	-	+	+	-	+	+	-	-	-	+	+	-	-	-	-	-	+	-	-	-	-	-
6	-	-	+	-	+	-	-	-	-	-	-	-	+	-	-	-	-	+	-	+	-	-	-	-	-
16	-	-	-	+	+	-	-	+	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	+
22	-	-	+	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-
7	-	+	-	-	+	+	-	+	+	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	-
14	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	+	+	-	+	-	-	-	-	-
13	-	-	-	-	+	+	-	-	+	-	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-
18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-	-
19	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
20	-	-	-	+	+	+	-	-	+	-	+	-	+	+	-	-	-	-	-	-	-	-	-	+	-
21	-	-	-	+	+	+	-	-	+	-	+	-	+	+	-	-	-	-	-	+	-	-	-	-	-
17	-	-	+	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
24	-	+	-	-	+	+	-	-	+	-	-	-	+	-	-	-	-	+	+	+	+	-	-	+	+
15*	-	-	-	-	+	+	-	-	+	+	-	-	+	-	-	-	+	-	-	+	-	-	-	-	-
23	+	+	+	+	+	-	-	+	-	-	-	+	+	-	+	+	-	+	-	+	+	+	+	-	+
58	-	-	+	-	+	-	-	+	-	-	-	-	+	-	-	-	-	+	-	+	-	-	+	-	-
67	-	-	+	-	+	-	-	+	-	-	-	-	-	+	-	-	+	+	-	-	-	-	-	-	-
26	-	-	+	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	-	+	-	-	-	+	-
27	-	-	+	-	+	-	-	+	-	-	-	-	+	+	-	-	-	-	-	+	-	-	-	+	-
69	-	-	+	-	+	-	-	+	-	+	-	-	-	-	-	-	+	-	+	+	-	-	-	-	-
28	-	+	+	+	+	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	+	-	+	-
45	-	-	+	-	+	-	-	+	-	-	-	-	+	-	-	-	+	-	-	+	-	-	-	+	-
29	-	-	+	-	+	-	-	-	-	+	-	-	+	-	-	-	+	-	-	+	-	-	-	-	-
32	+	-	+	-	+	-	-	+	-	-	-	+	+	-	+	+	-	+	-	+	+	+	+	-	+
33	-	-	+	-	+	-	-	+	-	-	-	-	+	-	-	-	-	+	-	+	-	-	+	+	-
34	-	-	+	-	+	+	-	+	+	+	+	-	+	+	-	-	+	+	-	+	-	-	+	-	-
30	-	-	+	-	+	+	-	+	-	-	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-
68	-	-	+	-	+	-	-	+	-	+	-	-	+	-	-	-	+	+	+	+	-	-	+	-	-
70	-	-	+	-	+	+	-	+	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
71*	-	-	+	-	+	+	-	+	+	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-
72	-	+	+	+	+	-	-	+	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-
51	-	-	+	-	+	-	-	-	-	-	-	+	+	-	+	+	-	+	-	-	-	-	-	-	-
52	+	+	+	+	+	-	-	+	-	-	-	+	+	-	+	+	-	+	-	+	-	+	+	-	+
54	+	-	+	-	+	-	-	+	-	-	-	+	+	-	+	+	+	+	-	+	-	+	+	-	+
56*	-	-	+	+	+	+	-	+	-	-	-	+	+	-	+	+	-	-	-	-	+	-	+	-	+
57	+	+	+	-	+	-	+	-	+	-	+	+	+	+	+	+	-	+	-	-	-	+	+	-	+
55	-	-	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-
65	+	-	+	-	+	-	-	+	-	-	-	-	-	-	+	+	-	-	-	-	+	-	-	-	-
61	-	-	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-
62*	+	-	+	-	+	+	+	+	+	+	-	+	+	-	-	+	-	-	-	-	+	-	+	+	+
63	+	-	+	-	+	-	-	-	-	-	+	-	+	+	+	+	-	+	-	+	+	-	+	-	+
64*	-	-	+	-	+	+	-	+	+	-	-	-	-	-	+	+	-	+	-	-	-	-	-	-	-
53	+	+	+	-	+	-	-	+	-	-	-	+	+	-	-	+	-	+	-	+	-	+	+	+	+
59	+	+	+	-	-	+	+	+	+	+	+	+	+	+	-	+	+	+	-	-	+	+	+	+	+
60*	-	-	+	-	+	-	-	-	-	-	-	+	-	-	+	+	-	+	-	-	-	-	-	-	-
66	+	+	+	-	+	-	-	-	-	-	-	+	-	-	-	+	-	+	-	+	+	-	+	-	-
35	+	-	+	-	+	-	-	+	-	-	-	-	+	-	-	+	-	-	-	+	+	-	-	-	+
36	+	-	-	+	+	-	-	+	-	-	-	-	+	-	-	-	-	-	-	+	+	-	-	-	+
38	+	+	+	-	+	+	-	+	+	-	-	-	+	-	+	+	-	-	-	+	+	-	-	-	+
39	+	+	+	+	+	+	-	+	+	-	-	-	+	-	+	+	-	-	-	+	+	-	-	-	+

Table 4-5: Continued

	E. coli EC2029 fragments									*E. coli* EC317 fragments															
	1	1	1	1	2	3	3	3		2	2	2	2	2	2	3	3	3	3	3	3	4	4
	1	4	4	6	7	7	1	5	7	2	0	1	3	4	8	9	0	0	1	4	5	6	8	0	5
Strains	4	4	7	3	4	6	9	4	2	8	1	9	5	0	2	6	5	8	5	7	8	1	1	0	3
40*	-	-	+	-	+	+	-	+	+	-	-	-	+	-	+	+	-	-	-	+	+	-	-	-	+
41	-	-	+	-	+	-	-	-	-	-	-	-	+	+	-	+	-	+	-	-	+	-	+	+	-
46	-	+	-	-	+	-	-	+	-	-	-	-	+	-	-	+	-	+	-	-	-	-	-	-	+
49	+	-	+	-	+	+	+	+	+	-	-	-	+	+	-	+	-	+	-	+	+	-	-	+	+
50	-	+	+	-	+	+	+	+	+	-	+	-	+	+	-	+	-	+	+	-	+	-	-	-	+
44	-	-	-	+	+	-	+	-	-	+	-	-	+	-	-	+	-	-	-	+	-	-	-	+	+
47	-	-	+	-	+	-	-	-	-	-	-	-	+	-	-	+	+	-	-	+	-	-	-	+	-
48*	-	-	+	-	+	-	-	-	-	+	+	-	+	-	-	+	-	+	-	-	-	-	-	-	+
31	+	+	+	+	+	-	+	+	-	+	+	-	+	-	+	+	+	+	+	+	-	+	+	+	-
43	-	-	-	-	+	-	-	-	-	-	-	-	+	-	-	+	-	-	-	+	+	-	+	-	-
37	-	-	+	-	+	+	-	-	+	-	-	-	+	-	-	+	-	-	+	-	-	-	-	-	-
42	-	-	+	-	+	-	-	-	-	-	-	-	+	+	-	+	-	+	-	+	+	-	+	+	-
317	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
2029	+	+	+	+	+	+	+	+	+	+	-	+	+	-	+	+	+	+	+	-	+	+	+	+	+
K-12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

The + indicates the presence of DNA fragments determined by colony blot analysis. The - indicates that the fragment is absent from the strain. The *E. coli* strains EC317, EC2029, and *E. coli* K-12 do not belong to the ECOR collection. *E. coli* K-12 was the negative control. *E. coli* strains EC317 and EC2029 were the positive controls. The * indicates strains that were isolated from urinary tract infections. The ECOR strains were listed in order of relation as determined by MLEE analysis of the strains (Herzer, Inouye *et al.*, 1990).

in the two collections was compared and was significantly different (P= 0.0029). The presence of individual fragments in the two collections was not significantly different for 14 of the fragments (Table 4-6). Fragments 28, 235, 282, 296, 305, 308, 347, 358, 361, 381, and 1-44 were present in significantly more of the avian isolates than the ECOR strains. Fragments 240, 1-47 and 1-74 were present in significantly more ECOR strains than avian strains.

4.4.7 Screen for EHEC genes in Avian *E. coli* Isolates

The presence of EHEC related sequences in the unique fragments generated from the subtractive hybridization suggested that *E. coli* strains pathogenic for poultry may contain virulence factors from EHEC strains. There were 11 fragments with high homology to *E. coli* O157:H7 genomic sequences. This result suggested that EHEC virulence factors might be present in the avian *E. coli* strains. To test this theory, *E. coli* EC317, *E. coli* EC2029, and the collection of 41 avian isolates were screened for the presence of *eae*, *espA*, *espB*, *espD,* and *tir* by colony blot analysis. Restriction fragments containing these genes were used to generate the probes by random priming. Dr. Bret Finlay generously provided the plasmids containing these genes (Appendix A1). None of the *E. coli* strains isolated from the avian sources, including *E. coli* EC317 and *E. coli* EC2029, contained the EHEC virulence genes. The positive control for the colony blots was *E. coli* EDL933, which is an *E. coli* O157:H7 strain. *E. coli* EDL933 was positive in all cases. Other *E. coli* O157:H7 strains were not used for the colony blot analysis. The *E. coli* K-12 colony was negative in all of the blots, which confirms the test worked properly.

Table 4-6: Percentage of strains containing subtractive hybridization fragments in the avian strain and ECOR strain collections as determined by colony blot analysis.

Strain of origin for the fragment	Fragment	Avian strains (%)[a]	ECOR strains (%)	P value[b]
EC317	28	39	13	0.0019
	201	2	11	0.1522
	219	34	17	0.1123
	235	60	72	0.2937
	240	2	19	0.0095
	282	42	18	0.0141
	296	71	42	0.0034
	305	46	15	0.0007
	308	93	53	<0.0001
	315	5	7	1.0000
	347	39	63	0.0193
	358	71	24	<0.0001
	361	57	11	<0.0001
	381	68	25	<0.0001
	400	17	24	0.4798
	453	10	31	0.0115
EC2029	1-4	10	24	0.0818
	1-44	46	24	0.0202
	1-47	39	63	0.0193
	1-63	27	23	0.4924
	1-74	41	97	<0.0001
	2-76	44	39	0.6914
	3-19	12	11	1.0000
	3-54	41	56	0.1738
	3-72	31	36	0.6848

[a] A collection of 41 avian *E. coli* isolates.
[b] As determined by contingency table analysis.

4.4.8 The Identification of Cosmids Containing Unique Fragments.

The suppression subtractive hybridization (SSH) technique used digested chromosomal DNA as the substrate. The size of the fragments obtained by SSH was dependent on the location of restriction sites and not on the actual size of the unique region of the chromosome. The fragment sequences identified by subtractive hybridization may have been parts of larger unique areas of the chromosome. It was also possible that the genes important to virulence or other phenotypes were not on the fragments themselves, but were coded on flanking DNA. To identify the sequences flanking the subtractive hybridization fragments, a cosmid library was screened for the presence of fragments containing subtractive hybridization sequences.

4.4.8.1 Use of Pooled Probes

Cosmids containing fragments had to be identified so the flanking regions could be sequenced. An *E. coli* EC317 cosmid library, which had previously been made in our laboratory, was used (Elizabeth Umelo, unpublished). Seventeen of the fragments previously used to screen the avian and ECOR strain collections were used to screen the cosmids. Only two fragments from *E. coli* EC2029 were used because it was thought that while the fragments had sufficient homology to *E. coli* EC317 to be positive by colony blot analysis, there might be enough difference to make sequencing difficult. The two fragments were 1-63 and 1-74, and were chosen because they had homology to virulence plasmids from EPEC and EHEC strains. Ten of the fragments used were from the group with unknown function, employed because we were interested in identifying

novel virulence factors. To maximize the number of fragment-containing cosmids found, four pools of probes were used. The first pool contained fragments 235, 240, 315, 347, and 358 from the EC317 library. The second pool contained fragments 282, 305, 361, 400, and 453 from the EC317 library. The third pool contained fragments 28, 201, 219, 296, and 308. The final pool contained fragments 1-63 and 1-74 from the EC2029 library of fragments.

Cosmids were identified using three rounds of colony blot analysis. The first round used colony lifts of the cosmid library spread on plates. The membranes were probed with the pooled probes. Colonies with strong signals were removed as plugs from the plates, eluted in LB broth, and plated on LB ampicillin plates. Up to 16 colonies from each plug were inoculated into wells of a 96 well plate containing LB broth with ampicillin. The cultures in the 96 well plate were used to inoculate membranes for colony blots. These colony blot membranes were tested with the pooled probes to confirm the cosmids contained fragment sequences. The positive colonies were reassembled in new 96 well plates, and duplicate colony blot membranes were made. These membranes were screened using probes to individual fragment sequences. The cosmids positive in this final screen were used for sequencing.

The three rounds of blots were done with the four pools of probes, and 70 cosmids were identified which contained fragments. Due to the use of the pooled probes, only eight of the subtractive hybridization fragments were found on the cosmids. The fragments found on the cosmids were 235, 240, 308, 315, 400, 453, 1-63, and 1-74. Every cosmid tested for fragment 235 appeared to contain it. Six of the cosmids appeared to contain more than one of the fragments. Cosmids 2158, 2161, and 2162 appeared to contain fragments 240 and 315. Cosmid 2160 appeared to contain

fragments 240 and 453. Cosmid 2169 appeared to contain fragment 240 and 400. Cosmid 2165 appeared to contain fragments 240, 315 and 453.

Cosmid vectors can contain inserts from 32 to 47 kb in size. A cosmid insert containing a subtraction hybridization fragment could be composed entirely of novel sequences. The quantity of DNA that would require sequencing to determine the limits of the unique area was uncertain. Due to the cost of sequencing a potentially large area, it was decided that only a few cosmids would be sequenced. Cosmid 2165 was chosen for sequencing because it contained three fragments (240, 315 and 453), and could have a large area of unique sequence. The other cosmid chosen for sequencing was cosmid 2181, since it contained fragment 1-74 with homology to a virulence plasmid of EPEC strains.

Attempts to sequence the two cosmids did not yield useful data. Cosmid DNA for sequencing was prepared using a cesium chloride purification that yields high quality DNA. The primers for the sequencing were designed using the fragment sequences. In the case of cosmid 2165, the primers were designed from fragment 240. A sequencing run using cosmid 2181 and a primer to the cosmid vector was successful, so the problem was not due to the template DNA. The cosmids believed to contain the subtractive hybridization fragments were tested for the presence of the fragments by Southern blots. The cosmid library was generated using *E. coli* EC317 genomic DNA, so if the cosmid DNA and *E. coli* EC317 genomic DNA were digested with the same restriction enzyme, the band containing the fragment should be the same size. The genomic DNA was digested with *Rsa*I before it was used for subtractive hybridization, so *Rsa*I was used to digest the cosmid and genomic DNA for the Southern blots. The subtractive hybridization fragments were cloned into pT-Adv, adding about 42 base pairs to the

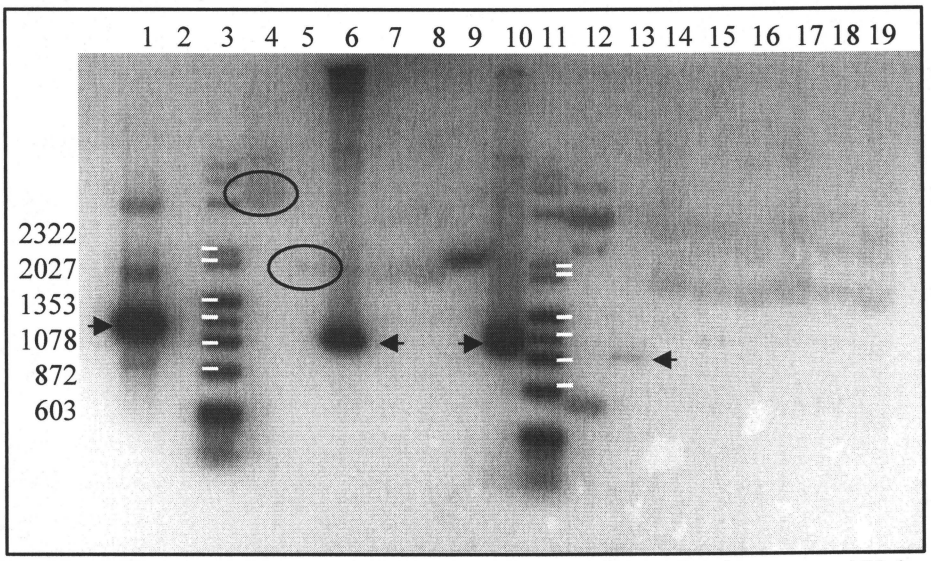
sequences when they are cut from the vector using an *EcoR*I digest. The Southern blots were done using *Rsa*I digests of the cosmids, which were probed with subtractive hybridization fragments. Figure 4-15 shows a Southern blot of cosmids thought to contain fragment 453. Of the seventy cosmids probed, only two contained fragments (cosmids 2158, and 2165). The Southern blot analysis demonstrated that the probes bound to the cosmid vector. The level of cross reactivity with the cosmid vector varied for the different probes, which was probably due to the quality of the DNA used to generate the probes.

4.4.8.2 The Use of Plasmid DNA Blocker and PCR Generated Probes.

The Southern blot analysis of the false positive clones showed cross reactivity between the cosmid vector and the probes (Figure 4-15). The labeled probes were made from a template extracted from an agarose gel, so there may have been short plasmid sequences labeled along with the fragment. A 1828 bp region of the pT-Adv plasmid vector and the SuperCos 1 cosmid are 98% homologous, so the plasmid and cosmid vectors would cross react strongly.

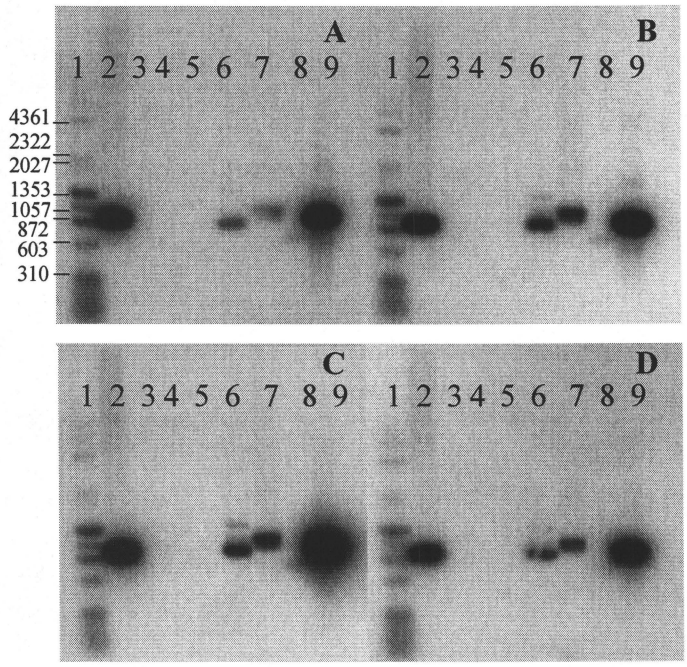
Two modifications to the Southern blots were done to prevent false positive clones. The first modification was to use PCR to generate the probes. Secondly, 50 μg/mL of denatured pT-Adv plasmid was added to the hybridization solution to block the membranes. Figure 4-16 shows the effect of the modifications. The comparison was done with probes to 1-74 and 453. The use of PCR probes considerably reduced the amount of cross-reactivity between the probe and the cosmid vector (Figure 4-16A).

Figure 4-15: Southern blot analysis of cosmids determined to contain fragment 453 by colony blot analysis.



Southern Blot analysis of cosmids that were positive for fragment 453 by colony blot analysis. The probe was made by random priming using the fragment 453 *EcoR*I fragment as a template. The markers were in lanes 3 and 11 and the sizes of the markers in bp were listed on the left and were indicated by the white lines. The positive controls were lane 1 (pT-Adv containing fragment 453 digested with *EcoR*I) and lane 13 (*Rsa*I digest of *E. coli* EC317). The black arrows indicate the bands containing fragment 453. The fragment was 936 bp in the *Rsa*I digest of *E. coli* EC317 (lane 13) and 978 bp in the plasmid (lane 1). Lane 4 and lane 5 were the negative controls. Lane 4 was undigested SuperCos 1 vector. Lane 5 was *Rsa*I digested cosmid 2186 that did not contain fragment 453 as determined by colony blot analysis. The sizes from top to bottom are 2322, 2027, 1353, 1078, 872, and 603. Lanes 6 to 10 and 14 to 17 are *Rsa*I digests of cosmids believed to contain fragment 453. Lanes 6 and 10 were cosmids 2165 and 2158 respectively that contained the fragment indicated by the black arrows. The faint bands indicated by the circles in lanes 4 and 5 indicated that the probe bound to the cosmid vector.

Figure 4-16: Comparison of Southern blot modifications for screening cosmids.

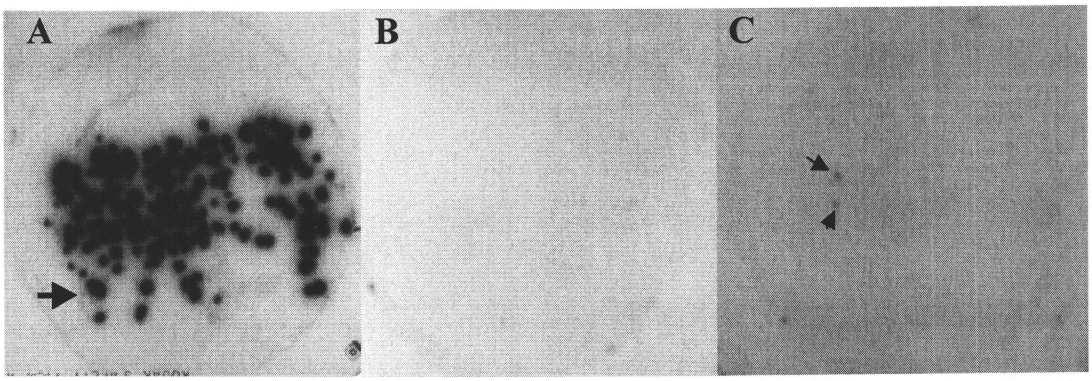


Four identical membranes were probed using the PCR generated probe to fragment 453 with other modifications to the protocol. Southern blot A used the original hybridization solution and PCR product. Southern blot B used the original hybridization solution and the gel purified PCR product as the probe. Southern blot C used the hybridization solution containing 50 μg/mL of denatured pT-Adv plasmid and PCR product. Southern blot D used hybridization solution containing 50 μg/mL of denatured pT-Adv and used the gel purified PCR product as the probe. Lane 1 contains the labeled Drigest III marker, the sizes of the markers are indicated in bp for blot A. Lane 2: *Rsa*I digested cosmid 2165 that contains fragment 453. Lane 3: *Rsa*I digested cosmid 2181 that does not contain fragment 453. Lanes 4 and 5 are negative controls and contain *Rsa*I digested SuperCos 1 and pT-Adv respectively. Lane 6: *Rsa*I digested genomic DNA from *E. coli* EC317. Lane 7: *Rsa*I digested genomic DNA from *E. coli* EC2029. Lane 8: pT-Adv containing fragment 1-74 that was digested with *Eco*RI. Lane 9: *Rsa*I digested pT-Adv containing fragment 453 that was digested with *Eco*RI. Lanes 3, 4, 5, and 8 are expected to be negative and lanes 2,6, 7 and 9 are expected to contain fragment 453. The size of the band in lanes 2 and 6 should be 936 bp and the band in lane 9 should be 978 bp.

The additional steps of gel purification of the PCR product and use of the pT-Adv plasmid to block the membranes did not have as dramatic an effect as the use of PCR generated probes. There was a slight improvement using the purified PCR product and the blocker, so these modifications were used in subsequent work.

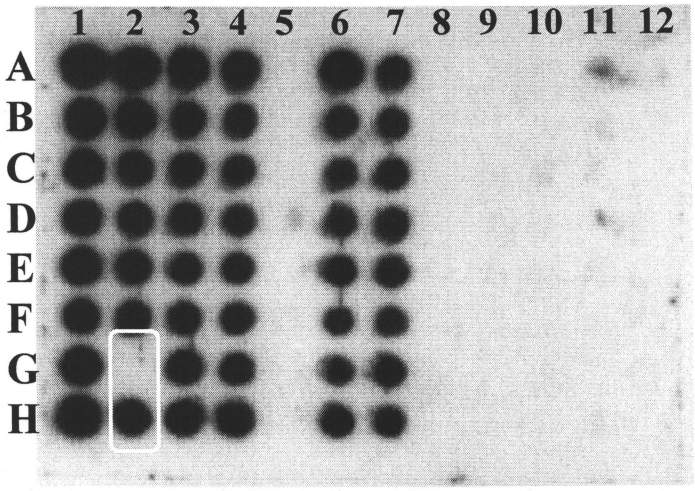
The modified blot protocol was used over the entire cosmid screen protocol to reduce the number of false positives found at each step. The cosmid screen protocol was tested using *E. coli* strains EC2165, and EC2196 as the controls for blots probed with fragment 453. *E. coli* strain EC2165 was used as the positive control since it contained a cosmid with fragment 453. *E. coli* strain EC2196 was the negative control since it contained the SuperCos vector without an insert. Figure 4-17 shows the result of the initial step used to screen the cosmid library. The modified blot protocol did not show any positive colonies for the negative control plate, and the positive control plates had all positive colonies. To test the next step of the screen, positive, weakly positive and negative colonies from the cosmid library were removed as plugs. The plugs were eluted and used to generate new colony blot membranes. An example of a membrane was shown in Figure 4-18. The strongly positive colonies from the initial colony blot were all positive. The negative colonies remained negative. The weakly positive colonies had some positive and some negative colonies. Some of the colonies were checked by Southern blot analysis as the final stage of the screen (Figure 4-19) All of the clones in row D of the colony blot (Figure 4-18) were used for Southern blot (Figure 4-19), and every clone that was positive by the colony blot remained positive on the Southern Blot. These results confirm that the modified blot protocol reduced the number of false positive cosmids identified. This modified protocol was used to identify more cosmids containing the subtractive hybridization fragments.

Figure 4-17: Screen of the cosmid library for fragment 453.



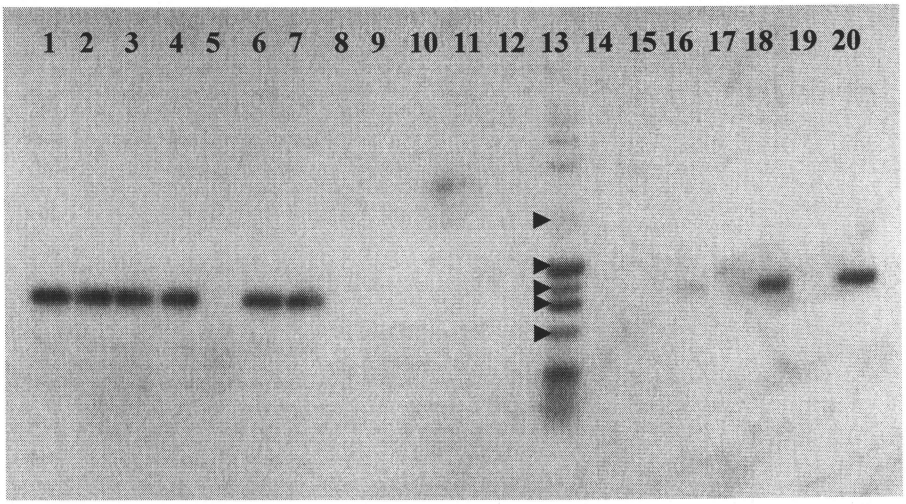
These membranes were blocked with denatured pT-Adv, and were probed with the gel purified radiolabeled PCR product to the fragment 453. Plate A: colonies of *E. coli* EC2165, which contains fragment 453 in a cosmid. This plate was used as the positive control. Plate B: colonies of *E. coli* EC2196, which contains only the SuperCos 1 vector. This plate is used as the negative control. Plate C: a portion of the *E. coli* EC317 cosmid library. The arrows indicate the weakly positive colonies.

Figure 4-18: Colony blot of cosmid library probed with fragment 453.



The probe was generated by PCR amplification of fragment 453 using radio-labeled nucleotides. The blot is made up of members of a cosmid library. The controls are indicated by the box. The negative control was at 2G and was *E. coli* EC2196, which contains the cosmid vector without an insert. The positive control was at 2H and was *E. coli* EC2165, which contains fragment 453 in a cosmid. Each column was derived from one colony from the cosmid library, which had been picked as a plug, eluted and streaked onto LB agar with kanamycin. From each plug, 8 colonies were picked, (except for column 2 which has the 2 control colonies) to insure the plug contained a pure culture. Columns 1 to 4 were of strongly positive colonies from the previous blot (Figure 4-17), columns 5 to 8 were from the weakly reactive colonies, and columns 9 to 12 were from colonies the negative colonies.

Figure 4-19: Cosmid library Southern blot probed with fragment 453.



Cosmid DNA was digested with *Rsa*I and used in a Southern Blot. Lanes 1-12: well D from columns 1-12 on the colony blot (Figure 4-18). Lane 13: labeled Drigest III marker. Starting from the top arrow, the bands indicated correspond to 2322, 1353, 1078, 872, and 603 bps. Lane 14: pT-Adv, the plasmid vector. Lane 15: SuperCos 1, the cosmid backbone present in all of the cosmids shown. Lane 16: chromosomal DNA from *E. coli* EC317 (the strain the fragment and the cosmids were derived from) digested with *Rsa*I. Lane 17: *E. coli* EC2029 chromosomal DNA digested with *Rsa*I. Lane 18: the cosmid pEC2165 (known to contain fragment 453) digested with *Rsa*I. Lane 19: no sample. Lane 20: pEC2110, pT-Adv containing fragment 453, digested with *Eco*RI. The band in lane 20 indicates that the blot worked. The faint band in lane 17 and the lack of a band in lane 16 were due to insufficient DNA.

The use of PCR to generate probes for the Southern blot analysis required specific primers for each subtractive hybridization fragment used. The number of fragments used was therefore reduced, and the primers were not pooled. Fragments 296, 358, 361, and 400 were chosen because they were in all of the cellulitis isolates known to be highly virulent. The *E. coli* strains EC429, EC477, EC479, and EC480 had been tested for virulence in a day-old chick model of infection and were found to affect 100 % of the birds (Ngeleka, Kwaga *et al.*, 1996). Fragments 1-63 and 1-74 were used because they have homology with virulence plasmids.

Using the three rounds of blots previously described, cosmids were found which contained fragments 296, 358, 361, and 1-63.

4.4.8.3 Sequencing Identified Cosmids

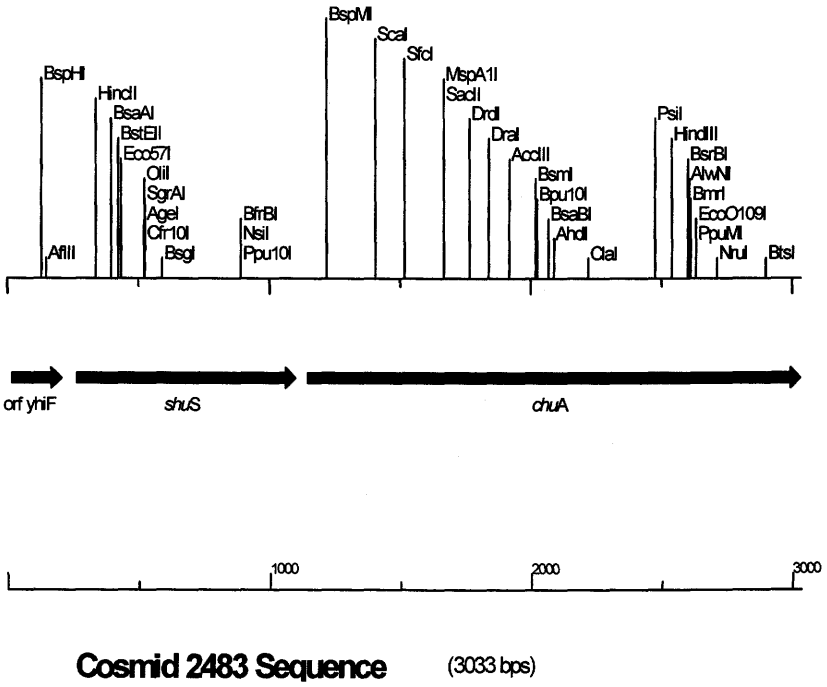
Cosmids were identified that contained fragments 296, 358, 361, 453 and 1-63. Two cosmids containing fragment 453 were found in the initial cosmid screens using the pooled, randomly primed probes (Section 4.4.8.1). The cosmids containing fragments 296, 358, 361, and 1-63 were identified using PCR generated primers in the modified protocol. Cosmids containing fragments 296, 453, and 1-63 were sequenced. The cosmid DNA was isolated using large-scale cesium chloride purification. Primers that anneal to the ends of the subtractive hybridization fragment sequence were used to sequence the flanking DNA on the cosmid. Subsequent sequencing used the primers to the ends of the previous round of sequencing. Both strands of DNA were sequenced for all three of the cosmids.

4.4.8.3.1 Cosmid Containing Fragment 296

Ten cosmids were identified that contained fragment 296. The cosmid in *E. coli* strain EC2483 (cosmid 2483) was used for sequencing. A total of 3033 bps of the cosmid was sequenced (see Figure 4-20). There was 96% homology between 3018 of 3033 bp of the cosmid sequence and the *shu* gene cluster of *Shigella dysenteriae*. The cosmid sequence did not cover the entire *shu* gene cluster, but the genes that were present were in the same orientation as the gene cluster in *S. dysenteriae*. The ORFs indicated on Figure 4-20 showed two *shu* cluster genes and a part of ORF *yhiF* flanking the shu gene cluster (Wyckoff, Duncan *et al.*, 1998). The function of *yhiF* is not known. The ShuS protein was known to bind both heme and double stranded DNA (Wilks, 2001). ShuS is thought to function in storage of heme and in the protection of DNA from oxidation induced by high iron levels. The ShuA protein was an outer membrane heme receptor (Wyckoff, Duncan *et al.*, 1998). Both ShuA and ShuS function in the use of heme as an iron source for bacterial strains.

Additional sequencing was done to determine if the other junction of the *shu* gene cluster matched the reported sequences for *S. dysenteriae*. The other junction was sequenced from the *E. coli* K-12 gene *f215* into the *shuV* gene. The single strand DNA sequence for 45bp *f215/shuV* junction was identical to the *S. dysenteriae* junction except for two point mutations. The sequence for this junction was completed for only one strand, so it may have contained some errors. The *shuS/yhiF* junction was part of the cosmid sequence, and had a 5 bp deletion and four point mutations in 54bp (Wyckoff, Duncan *et al.*, 1998). It was not determined if the other *shu* genes were present on the cosmid.

Figure 4-20: The cosmid containing fragment 296



The open reading frames indicated by the arrows were derived from the homology of the sequence to genes in *E. coli* and *Shigella dysenteriae*. ORF *yhiF* is an open reading frame originally identified in *E. coli* K-12 strain MG1655, which is known to flank an insertion in *E. coli* O157:H7 strain EDL933, and some *Shigella* species. The cosmid sequence of ORF *YhiF* is only the amino terminal end of the ORF. The sequence indicated by *shuS* on the figure is 96% homologous to the entire *shuS* gene of the *S. dysenteriae*. The ShuS protein is known to bind both heme and double-stranded DNA (Wilks, 2001). The sequence indicated by chuA on the figure is 96% homologous to the *chuA* gene of *E. coli* O157:H7 strain EDL933 and to the *shuA* gene of *S. dysenteriae*. The area of the cosmid, which was sequenced lacks the last 299 bp of the *chuA/shuA* gene.

Most *E. coli* strains produce enterobactin (*ent*), which is a very strong chelator of iron and is able to strip iron from transferrin (Ratledge and Dover, 2000). Cosmid 2483 was transformed into *E. coli* EC2503, which is a strain of *E. coli* HB101 that cannot produce enterobactin (Wyckoff, Duncan *et al.*, 1998), and the resulting strain (EC2506) was tested for its ability to use heme as a sole iron source. The *ent* strain containing the cosmid was unable to grow on minimal media agar with heme as the sole iron source, but strains EC317 and EC2029 could. This result indicates that the *shu* gene cluster on the cosmid was unable to confer the heme-use phenotype. The expression levels of the *shu* genes were not determined, so it is possible that they are not expressed from the cosmid sequence.

There was no indication of heme binding in the *E. coli* strains EC317, EC2029, or in *E. coli* K-12 MG1655. Heme binding was determined by the color of the colony on M9CA plates supplemented with hemin. For unknown reasons, the *E. coli* EC2506 strain did not grow on the M9CA heme plates, so it could not be determined if the cosmid could confer a heme binding phenotype.

4.4.8.3.2 Cosmid Containing Fragment 453

Two cosmids containing fragment 453, one of the sequences with unknown function, were identified: cosmids 2165 and 2158. A total of 3905 bp including the fragment, were sequenced from cosmid 2165. This area had some homology to phage proteins, but the homology was scattered and did not include the entire sequence of the individual proteins.

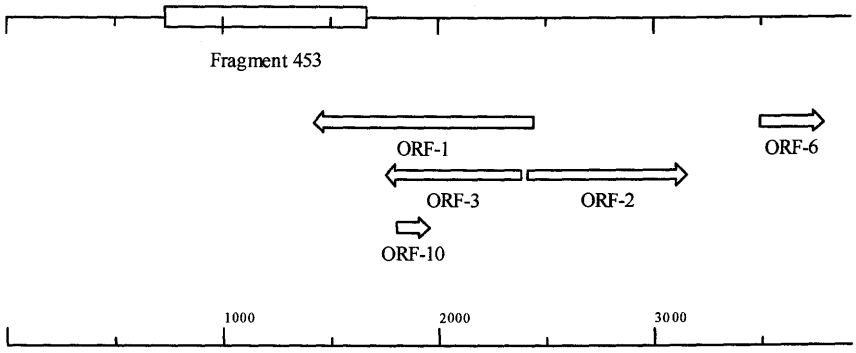
The sequence was analyzed further by screening for ORFs using Clone Manager, and by screening for genes using the Glimmer2 program. Using Clone Manager, 11 ORFs were identified that were greater than 50 amino acids and contain a start codon. Using Glimmer2, 10 potential genes greater than 30 amino acids were identified. The ORFs identified were compared to the putative genes generated by Glimmer2. Five of the Clone Manager ORFs were identified as genes by Glimmer2 (Figure 4-21). The five putative genes were examined using the BlastP (http://www.ncbi.nlm.nih.gov/BLAST) homology search and Motifs (http://www.motif.genome.ad.jp) searches for protein motifs (Table 4-7).

The BlastP search using ORF1 indicated that most of the sequence had identity (49%) with the L-shaped tail fiber protein of bacterophage T5. The homologous area covered 200 of the 340 residues. The Motifs search found over 20 proteins with similar motifs to the ORF. The best match was to a domain of the enzyme phosphoenolpyruvate carboxykinase (PCK). The portion of the PCK that was homologous (amino acids 275-316 of PCK) included the connecting region between two globular domains of the protein. The motif was 41 amino acids long and contained 16 amino acids found in both PCK and ORF1. One residue that was not conserved was an active site amino acid (K288). Due to the low homology with this motif, and the fact that an active site amino acid was not conserved, it is not likely that this ORF has a similar function to PCK. A variety of other proteins had motifs that covered areas of the ORF.

ORF2 shared 30% identity with a probable tail fiber protein of phage P2. The Motifs search identified a number of motifs but none had high score. This ORF had low

Figure 4-21: Partial sequence of cosmid 2165 including fragment 453.

A)



B)

```
ORF-1:  72  TASENSATAAKKSE-----TNAKNSETAAKTSET--NAKSSQTAAKTSETNAKASETAAK 124
            T  +NS TA+K  +      +N+ +S TAA  +    ++K+S  AAK SE NAK SE  AK
  LTR:  13  TMDQNSITASKYPKYTVVLSNSISSITAADVTSAIESSKASGPAAKQSEINAKQSELNAK 72

ORF-1: 125  SSQDAAAESESAAAGSATSAAGSATAAANSQKAAKTSETNAKSSQTAAKTSETN------ 178
             S++ A  S +++  SAT +A SATA+ANS KAAKTSETNA +S+ AAKTSETN
  LTR:  73  DSENEAEISATSSQQSATQSASSATASANSAKAAKTSETNANNSKNAAKTSETNAASSAS 132

ORF-1: 179  -AKASETAAKSSQDAAAESESAAAGSASAAAASATASANSQKAAKTSETNAKVSETAAAN 237
             A +  TAA++S  AA  SE+ A  SA AA AS TA+ANS  AAKTSETNAK SETAA
  LTR: 133  SASSFATAAENSARAAKTSETNAGNSAQAADASKTAAANSATAAKTSETNAKKSETAA-- 190

ORF-1: 238  SAKASAASQTAAKASEDAAREYASQAAEPYKYVLQ 272
                   S+T AK SE+ A+EY   A+E    V Q
  LTR: 191  -----KTSETNAKTSENKAKEYLDMASELVSPVTQ 220
```

C)

```
Orf2: 110  TNLYGGTMIFPG--DSGYLKMGNCLMSYSKRGSNALIKFDYTDTLQIKYAN----HGSTM 163
           TN  GG+ I  G  D+G+ + G+ ++      +N+   F + + + I + N
 gpH: 486  TNALGGSSIVLGDNDTGFKQNGDGILDVY---ANSQRVFRFQNGVAIAFKNIQAGDSKKF 542

Orf2: 164  TLNTQGTAYAGVTAQLWGNSRRPVVYEVGVDGGAYMFYAQKNTDNTYMLSVNGACHATAF 223
           +L++  T+   +T  LWG S RPVV E+G + G + FY+Q+NTDN+ + +VNG    + +
 gpH: 543  SLSSSNTSTKNITFNLWGASTRPVVAELGDEAG-WHFYSQRNTDNSVIFAVNGQMQPSNW 601

Orf2: 224  NQHSDRDLKD 233
                R +KD
 gpH: 602  GNFDSRYVKD 611
```

A) The ORFs are indicated by the arrows and were determined using both Clone Manager, with ATG as the start codon and a minimum length of 150bp, and the Glimmer2 program. The sequence of the SSH fragment 453 is indicated by the box. B) The alignment of the ORF1 sequence with the amino acid sequence of the L-shaped tail protein of bacterophage T5. The sequences shared 49% identity over the 200 residues. C) The alignment of the ORF2 sequence with the GpH protein of phage P2. The sequences share 30% identity over 130 amino acids.

Table 4-7: The homology of open reading frames found on cosmids containing fragments 453 and 1-63.

ORF	Size (aa)	Sequences with similarity[a]	Score [b]	Probability (e^x)[c]	GenBank accession no.
Fragment 453					
1	340	L-shaped tail fiber protein Bacteriophage T5	74.7	-12	CAA04591.1
2	246	gpH Enterobacteria phage P2	61.2	-9	NP_046775.1
3	210	No significant similarity found			
6	99	No significant similarity found			
10	50	No significant similarity found			
Fragment 1-63					
1	223	orf29 [*E. coli* B171]	209	-53	NP_053091.1
2	223	orf29 [*E. coli* B171]	209	-53	NP_053091.1
3	196	orf31 [*E. coli* B171]	312	-84	NP_053093.1
5	131	conjugative transfer: assembly [*S. typhimurium* LT2]	40	0.004	NP_490573.1
6	101	No significant similarity found			
9	79	TraC [*E. coli* plasmid F]	113	-26	AAB61935.1
11	59	Orf50 [*E. coli* B171]	116	-26	NP 053112.1

[a] Sequences with the highest homology to the fragment are listed. The homology is not usually to the entire length of the fragment.
[b] The score increases as the level and length of homology increase.
[c] The homology of the sequences increases, as the probability approaches 0.

homology with known proteins, so the function of this ORF could not be predicted, and it was not determined that the ORF was actually expressed.

ORF3 was identified as having homology with a large number of RNA and DNA binding proteins using the BlastP program. The homology spanned 160 of the 210 residues in the ORF, but the best match had only 26% homology. This level of homology was not considered significant. The Motifs search also identified a number of DNA or RNA binding motifs. There were motifs from 10 different ribosomal proteins, as well as a sigma factor and basic leucine zipper transcription factor.

One protein with homology to ORF6 was identified using the BlastP search. The homology was to a sodium/proline symporter of *Neisseria meningitides* and was 30% of 95 residues, which was not considered significant. The Motifs search identified 10 proteins with motifs similar to the ORF. The best match was with the Tub family of proteins, which are related to mammalian obesity (Kleyn, Fan *et al.*, 1996). The next best matches were to the glycosyl hydrolases family 10, and to legume lectin alpha domain, which may indicate carbohydrate binding motifs.

Neither the BlastP nor the Motifs search with ORF10 found any homologous sequences.

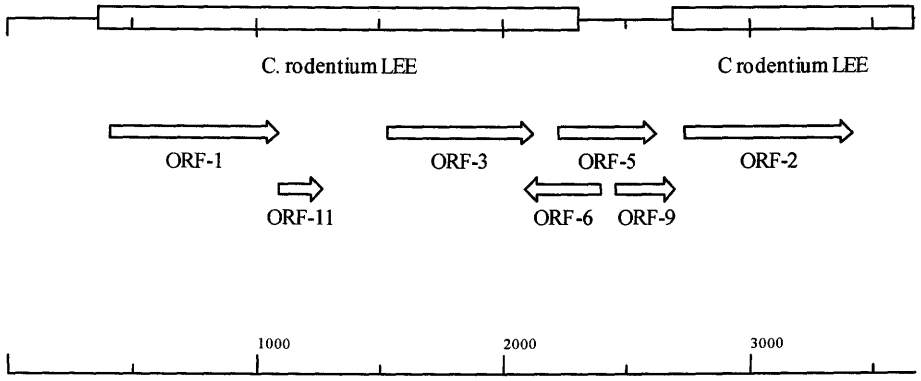
The cosmid 2165 sequence contains 5 ORFs with homology to diverse sequences. One of the ORFs may have a DNA binding function, but the function of the rest of the ORFs could not be predicted.

4.4.8.3.3 Cosmid Containing Fragment 1-63

Fifteen cosmids were identified containing fragment 1-63, including cosmid 2491. A total of 3664 bp of cosmid 2491 were sequenced. The sequence was homologous to a number of pathogenicity islands, including the LEE pathogenicity islands of EHEC strains and *Citrobacter rodentium* (see Figure 4-22) as determined by BlastN searches. The highest homology was to the LEE pathogenicity island of *C. rodentium*, which was 97% homologous to 2912 bp of the sequence. There was also high homology to three virulence plasmids of *Shigella flexneri* and the virulence plasmid pB171 of EHEC. The pB171 is a 69 kb plasmid of the EPEC strain B171, which contains a number of virulence genes, including the *bfp* operon (bundle forming pili) and part of *toxB* of EHEC O157:H7 (Elliott, Wainwright *et al.*, 1998). The homology was to ORFs 29, 30, 31, 49, 50, and 51, which did not code for any known virulence factors, but displayed homology, both with ORFs in the prophage region of the LEE pathogenicity island of EHEC, and to ORFs from a pathogenicity island of uropathogenic *E. coli* strains (UTEC) (Kao, Stucker *et al.*, 1997). Plasmid pB171 has a repeated region, so the ORFs 29, 30, and 31 are very highly homologous to ORFs 49, 50 and 51. The homology to pB171 covers five portions of the sequence for a total of 2020 bp with 93% homology. There was also some homology to ColV plasmid sequences and F plasmids. The sequence from cosmid 2491 had homology to uncharacterized open reading frames of these plasmids and pathogenicity islands.

The sequence was analyzed further by screening for ORFs using Clone Manager, and by screening for putative genes using the Glimmer2 program (Salzberg, Delcher *et al.*, 1998; Delcher, Harmon *et al.*, 1999). Using Clone Manager, 15 ORFs of greater

Figure 4-22: Partial sequence of cosmid 2491 including fragment 1-63.



The ORFs are indicated by the arrows and were determined using both Clone Manager, with ATG as the start codon, and a minimum length of 150bp, and using the Glimmer2 program. The boxes on the restriction site map indicate the regions with homology to the LEE pathogenicity island of *C. rodentium* (97% and 95% homology). The sequences homologous to the ORFs are listed in Table 4-7.

than 50 amino acids, each containing a start codon, were identified. Glimmer2 was used to identify 16 potential genes of greater than 30 amino acids. The ORFs described by Clone Manager were compared to the putative genes identified by Glimmer2; seven of the ORFs screened by Clone Manager were identified as putative genes by Glimmer2 (Figure 4-22). The ORFs homologies are listed in Table 4-7.

ORF1 had 58% homology to ORFs 29 and 49 of pB171, and lower homology to L0013 of the EPEC LEE island and to hp3 of the pathogenicity island of UTEC strain CFT073. The Conserved Domain database search (CD-D), which was automatically done with the BlastP search, was used to identify a 45 amino acid stretch that was homologous to transposase 8. ORF1 and 2 were 97% homologous and had almost identical results for Blast and CD-D searches. These results suggested that this ORF coded for a transposase.

ORF9 has strong homology (77% identity) to a portion of TraC, which is an F-pilus protein (Schandel, Muller *et al.*, 1992). The homology was to 74 of the 79 residues in the ORF, but TraC was 876 residues long, so it was unlikely this ORF had all of the functions of TraC. ORF5 also had homology (94%, 17/18 residues) to TraC. Since the area was so small and the Motif search showed protein motifs with very low scores, the function of this ORF could not be predicted.

Several of the ORFs had low levels of homology to protein sequences or had low score results for Motifs searches so their functions could not be predicted. ORF3 had 90% homology at the amino acid level to parts of ORFs 31 and 51 of pB171, and 60% homology to L0015 of the EPEC LEE island, which are ORF without known functions. The BlastP for ORF6 identified 37% homology to 39 residues of a *Drosophila* gene. The BlastP homology search for ORF11 identified five sequences with very strong

homology. ORFs 30 and 50 from pB171 were 100% matches to the entire sequence of ORF11, but the function of these pB171 ORFs are not known.

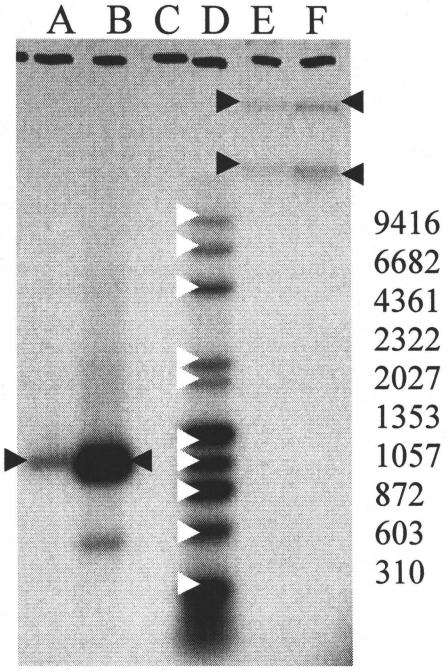
The function of most of the ORFs in the cosmid 2491 sequence could not be predicted. ORFs 1 and 2 had motifs, which suggested they might have functioned as transposons. The fact that this sequence had homology to virulence plasmid and pathogenicity island sequences suggested that it may have some function relating to virulence, but what this function may be is not known.

It was known from previous work that *E. coli* strain EC317 had at least one high molecular weight plasmid. Agarose gel electrophoresis indicated that *E. coli* strain EC2029 also had a high molecular weight plasmid. Most of the cosmid 2491 sequence had high homology to plasmid sequences, so Southern blot analysis was done to determine if plasmids contained fragment 1-63. The large plasmids of *E. coli* strains EC317 and EC 2029 both contained the 1-63 fragment (see Figure 4-23).

4.5 Disruption of Sequences

Disruption of the chromosomal copy of the subtractive hybridization fragments allows further investigation into the function of the sequences. The functions of greatest interest were those involved in pathogenesis. Testing a sequence for a role in pathogenesis involves evaluating a mutant strain for altered virulence in an animal model.

Figure 4-23: The presence of fragment 1-63 on the large plasmid of *E. coli* strains EC317 and EC2029.



A Southern blot was probed with fragment 1-63. Lanes A, B, and C are *Rsa*I digests of chromosomal DNA from *E. coli* strains EC317, EC2029 and MG1655. Lane D is the labeled Drigest III markers. The marker bands are indicated by the white lines, and the sizes in bp are listed to the left of the blot. Lanes E and F are undigested plasmids from *E. coli* strains EC317 and EC2029. The black arrows indicate bands containing fragment 1-63. The presence of the bands in lanes E and F indicate that fragment 1-63 is located on the large plasmids of these two avian *E. coli* strains.

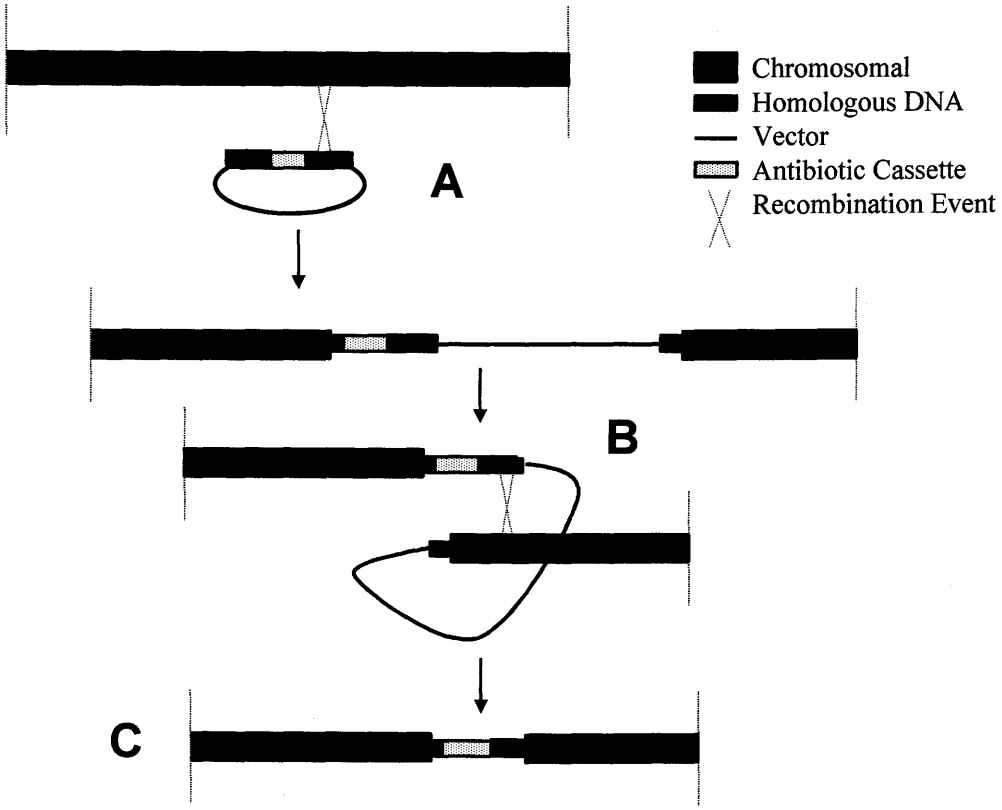
4.5.1 Identification of Sequences for Disruption

Ten subtractive hybridization fragments were chosen for disruption by the insertion of an antibiotic resistance gene. Fragments 28, 347, 358, 453, 3-54, and 3-72 were selected because they did not have significant homology to sequences with known function. Fragment 1-47 had homology to a fimbrial usher protein. Fragment 381 was selected because it had homology to pertactin, which is a virulence factor of *Bordetella* species. Fragment 1-44 was selected due to the homology it shared with a P-pili gene. P-pili are associated with virulence in uropathogenic *E. coli* strains. Fragment 1-63 was chosen due to the homology it had to the LEE pathogenicity island. The subtractive hybridization fragments chosen either had unknown function, or they had homology to virulence factors of other bacterial strains. Of these fragments 28, 358, 381 and 1-44 were found in significantly more of the avian *E. coli* strans then the ECOR strains.

4.5.2 Generation of Knockout Mutants

The general approach for making the knockout mutants was to clone an antibiotic cassette into the target sequences. The disrupted target sequence was moved into a temperature-sensitive plasmid to make the allele replacement plasmid. The allele replacement plasmid was transferred into the pathogenic strain, where it was used to replace the wild type sequence. The allele replacement protocol used the temperature-sensitive nature of the suicide plasmid to select strains that have had the plasmid inserted into the chromosome (see Figure 4-24), and then to identify strains that have undergone

Figure 4-24: Allele replacement to generate a stable mutant.



A) The strain containing the allele replacement plasmid was grown at the non-permissive temperature (42°C) for plasmid replication, which selected for clones with the plasmid integrated into the chromosome. The integration occurred by homologous recombination between the chromosomal sequence and one of the homologous sequences on the plasmid. B) The strain was grown at the permissive temperature for plasmid replication (30°C) to select for the second recombination event where the plasmid was excised. The plasmid was excised by a second homologous recombination event between the chromosomal sequence, and the second homologous sequence in the plasmid. C) The two recombination events caused the replacement of the wild type sequence with the sequence disrupted by the antibiotic resistance gene. The plasmid was excised and lost when the strain was grown at the non-permissive temperature. The second recombination event could have also resulted in the restoration of the wild type allele, but the antibiotic resistance gene in the mutant allele was used to select for the clones with the mutant allele.

the second recombination event. The antibiotic cassette disrupts the target sequence and allows for the selection of the strain containing the mutant allele.

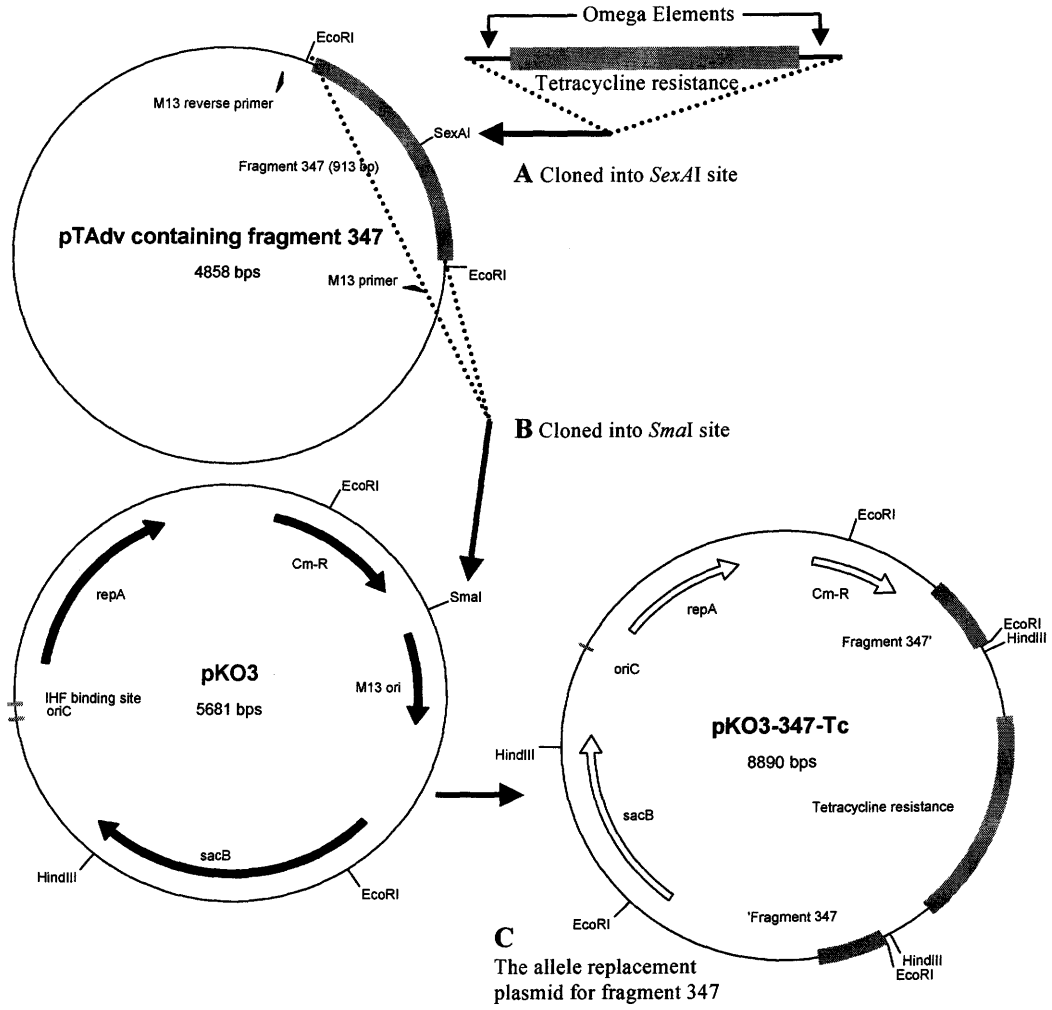
4.5.2.1 Construction of Plasmids Containing the Omega Tetracycline Cassette

The fragments from the subtractive hybridization had been cloned into pT-Adv, which has kanamycin and ampicillin resistance genes. Therefore we chose to use a tetracycline cassette flanked by Ω elements, which are transcriptional and translation stop signals. If an Ω element is inserted into a gene, it will prevent the expression of the portion of the gene that is downstream of the element. When it is used with an antibiotic cassette, it also prevents read though from the antibiotic resistance gene to genes downstream.

The ΩTc cassette was cloned into restriction sites in approximately the middle of the SSH fragments. The SSH fragments were then cloned into the temperature-sensitive plasmid pKO3. The production of the allele replacement vector for fragment 347 was shown in Figure 4-25, as an example of the general cloning plan.

Plasmid pKO3 was a temperature-sensitive suicide vector that has been used for allele replacement in wild type *E. coli* strains (Link, Phillips *et al.*, 1997). The plasmid contained the *sacB* gene that was used to select for the loss of the pKO3 plasmid. This plasmid had been used previously in our laboratory for allele replacement, and it was shown that the *sacB* selection does not work with *E. coli* strain EC317 (unpublished data).

Figure 4-25: The pKO3 allele replacement vector with fragment 347.



Generation of the allele replacement vector containing omega tetracycline in fragment 347. A) The fragment in pT-Adv was digested with *SexA*I, the ends were blunted using T4 polymerase, and the *Sma*I fragment containing the omega tetracycline cassette was inserted. B) The disrupted fragment was removed from pT-Adv using the flanking *EcoR*I sites and cloned into the *Sma*I site of the pKO3 plasmid. C) The resulting plasmid was used for allele replacement.

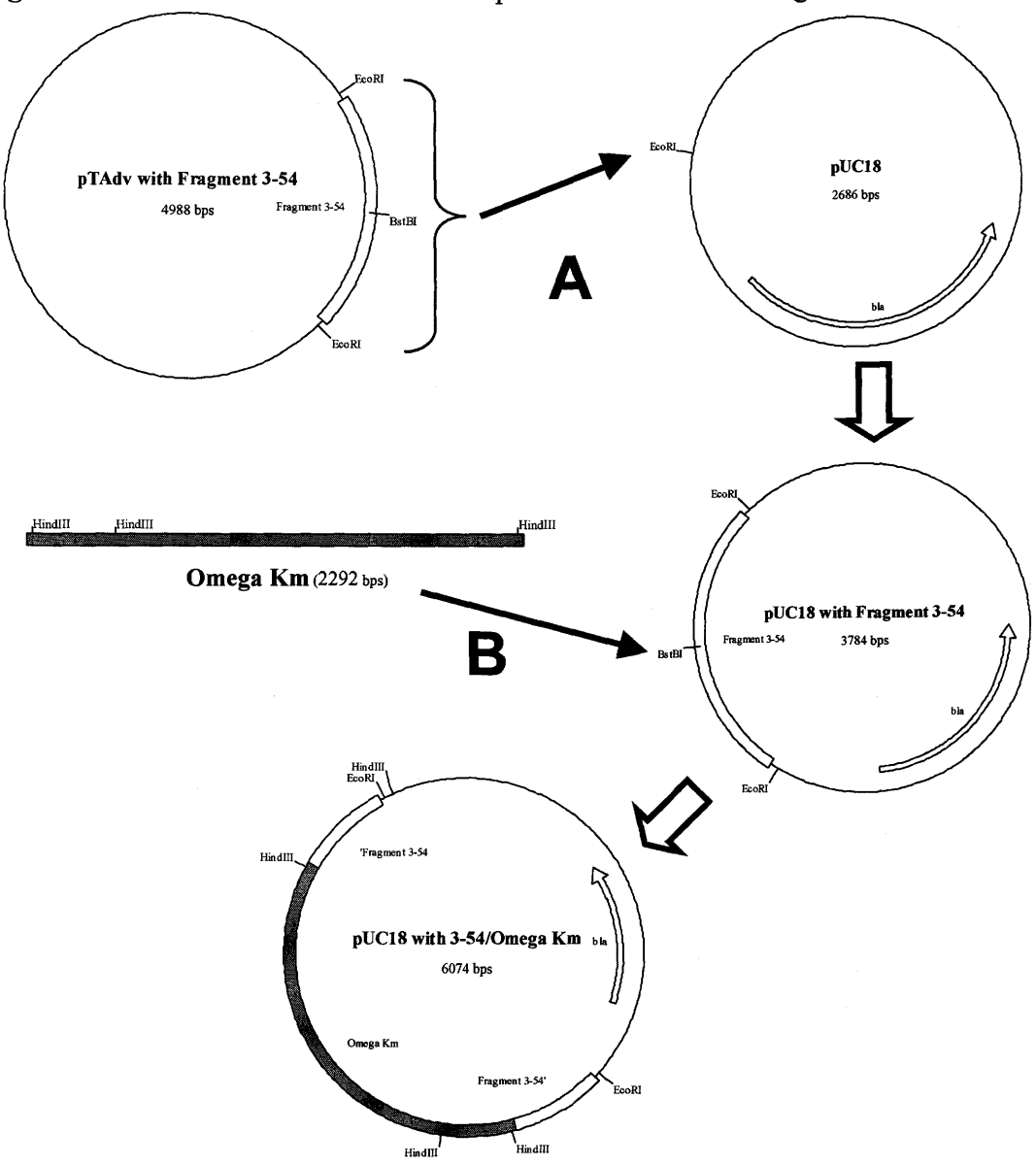
Four of the allele replacement constructs were made using the plasmids containing fragments 28, 347, 1-47 and 3-54. Two of the plasmids (containing fragments 28 and 347) were transferred to *E. coli* strain EC317. *E. coli* EC317 containing the replacement plasmid did not grow well on media containing tetracycline despite the presence of the Ω tetracycline cassette. Other members of our laboratory had similar problems with the Ω tetracycline cassette, so it was decided that the Ω kanamycin cassette should be used instead.

4.5.2.2 Construction of Plasmids Containing the Omega Kanamycin Cassette

To overcome the difficulties encountered using the Ω tetracycline cassette, the allele replacement plasmids were reconstructed using the Ω kanamycin cassette (Figure 4-26). Since the fragments of interest were in the pT-Adv vector, which already contained a kanamycin resistance gene, the fragments were cloned into pUC18. Each of the constructs containing a fragment in pUC18 were digested using appropriate restriction enzymes (Table 4-8) to linearize the plasmid and cut the fragment sequence approximately in half. The Ω kanamycin cassette was cloned into the SSH fragments, and the SSH fragments were cloned into the temperature-sensitive vector pIB307 (Figure 4-27). It was decided that the pIB307 plasmid would be used instead of pKO3 since it had a multiple cloning site.

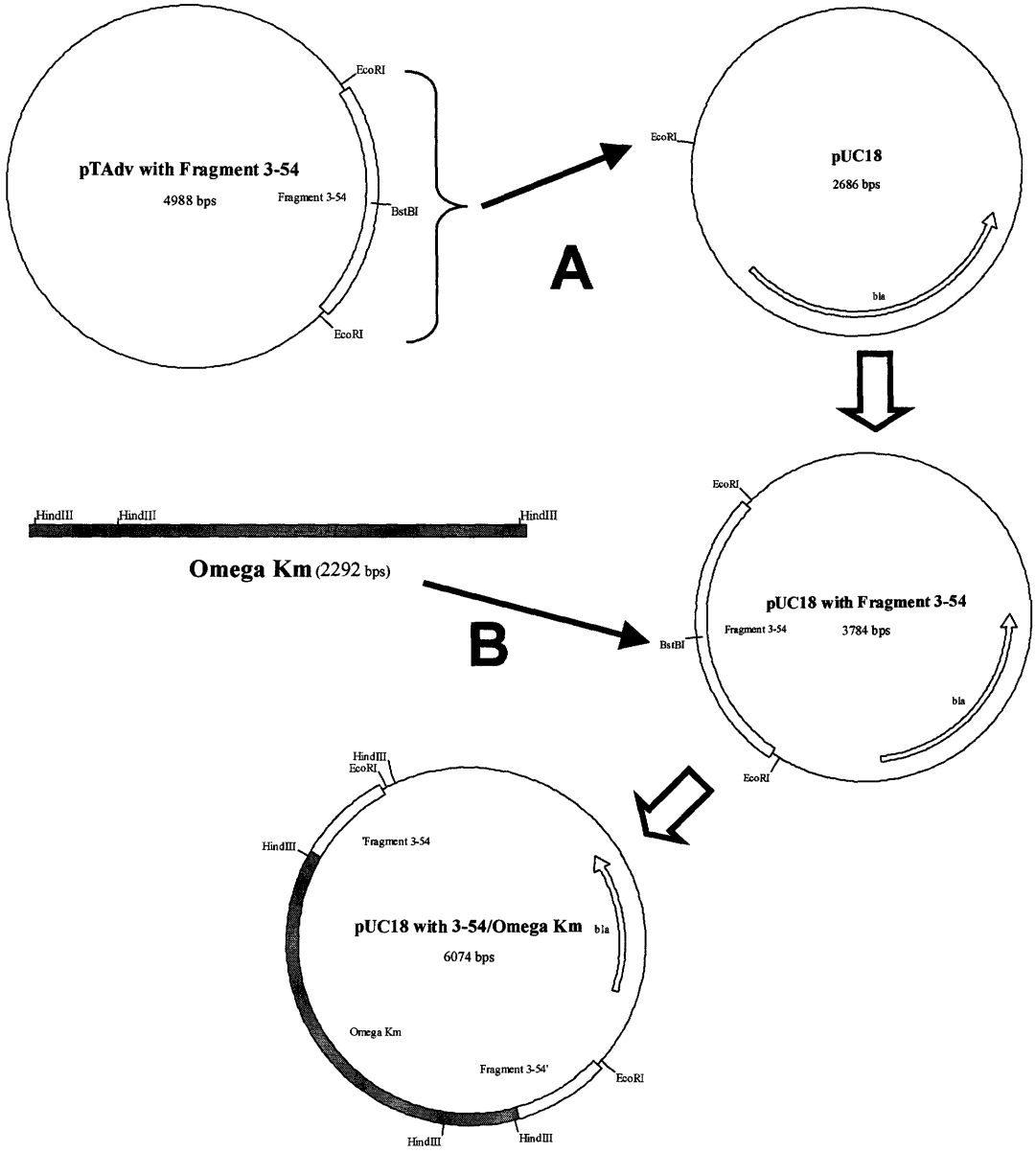
Eight of the ten fragments were cloned into the *EcoR*I site of pUC18. Fragment 3-72 contained an *EcoR*I site, so it required more steps to clone it into pUC18. As a

Figure 4-26: Construction of the allele replacement vector for fragment 3-54.



A) Digestion with *EcoR*I was used to remove fragment 3-54 from the pT-Adv vector. The *EcoR*I fragment was cloned into the *EcoR*I site of pUC18. B) The pUC18 plasmid containing fragment 3-54 was linearized by digestion with *BstB*I, the 5' overhang was removed, and the *Sma*I fragment containing the Ω kanamycin cassette was inserted.
continued on the next page.

Figure 4-26: Construction of the allele replacement vector for fragment 3-54. (continued)

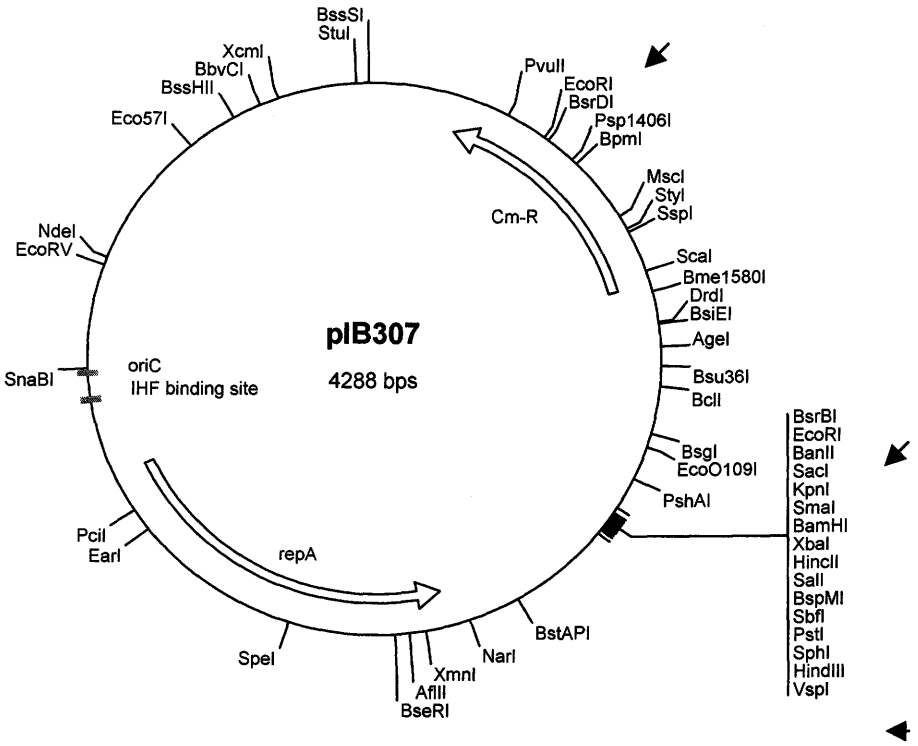


C) Digestion with *EcoR*I was used to generate a piece of DNA containing fragment 3-54 and the Ω kanamycin cassette. The *EcoR*I fragment was cloned into an *EcoR*I site of pIB307. The resulting plasmid (pEC2545) was used for allele replacement in *E. coli* strain EC317.

Table 4-8: The disruption of the subtractive hybridization fragment sequences.

Fragment	Length (bps)	Restriction Enzyme	Location of the site	Comments
28	465	*Bsg*I	257	Completed and transferred to EC317.
347	913	*SexA*I	480	Failed to insert Ω kanamycin.
358	307			Fragment too small.
381	451	*Bsg*I	249	Failed to insert Ω kanamycin.
453	936	*BsaA*I	449	Failed to transfer to EC317.
1-44	941	*Sac*II	518	Completed.
1-47	888	*Hinc*II	481	Failed to insert Ω kanamycin.
1-63	1112	*Sac*II	575	Failed to insert Ω kanamycin.
3-54	1038	*BstB*I	519	Completed and transferred to EC317.
3-72	1175			Failed to clone in to pUC18.

Figure 4-27: The temperature-sensitive plasmid pIB307.



Plasmid pIB307 is a temperature-sensitive plasmid. The RepA protein coded is fully functional at 30°C, but has a greatly reduced function at 42°C. The function of RepA is required for the replication of the plasmid. The plasmid has a chloramphenicol acetyltransferase gene, which encodes for a chloramphenicol resistance phenotype (Cm-R). There are two *EcoR*I sites in the plasmid: one is in the multiple cloning site, and the other is in the chloramphenicol resistance gene. Restriction digestion analysis using *Hind*III was used to identify pIB307 plasmids that contained an insert sequence.

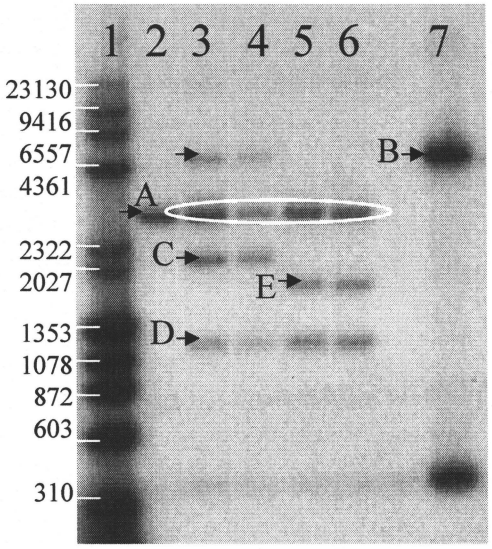
result, the cloning with fragment 3-72 did not proceed as quickly as the other fragments, and the allele replacement vector was not finished. Fragment 358 contained two *Rsa*I fragments ligated together, and it was not pursued further. Table 4-8 lists the fragments and the sites in the fragments where the Ω kanamycin cassette was to be cloned. The Ω kanamycin cassette was successfully cloned into four of the fragments: 28, 453, 1-44, and 3-54. Attempts to clone the cassette into the other fragments failed. The four disrupted SSH fragments were cloned into pIB307. The plasmids were transferred to *E. coli* EC317 for the allele replacement.

The strain used for the allele replacement was *E. coli* EC317, because it had been used previously in attenuation studies and was known to be virulent. Less was known about the virulence of *E. coli* strain EC2029. Fragment 1-44 was not found in *E. coli* EC317, so the allele replacement vector with the fragment 1-44 sequence could not be used with *E. coli* strain EC317. The replacement vectors for fragments 28 and 3-54 were successfully moved into *E. coli* EC317, but attempts to move the fragment 453 vector into *E. coli* EC317 failed.

4.5.2.3 Allele Replacement

The allele replacement was done with the allele replacement plasmids for fragments 28 and 3-54. Two strains were identified that had the antibiotic resistance pattern expected for the fragment 28 mutant. Southern blot analysis of the two disruption mutants was used to confirm that the allele replacement was correct (Figure 4-28). Two strains with the insertion of the plasmid into the chromosomal copy of the

Figure 4-28: Southern blot analysis of *E. coli* EC317 mutants with the insertion in the fragment 28 sequence.



1) Markers
2) Genomic *E. coli* EC317 DNA
3) Single recombination mutant 1
4) Single recombination mutant 2
5) Double recombination mutant 1
6) Double recombination mutant2
7) Replacement vector

The Southern blot was probed with the fragment 28 sequence. Lane 1: radiolabeled Drigest III marker. The sizes of the bands in bp are indicated. Lanes 2-6 are genomic DNA, digested with *Hin*dIII from *E. coli* EC317 and strains derived from it. Lane 7: the replacement vector was expected to give two bands of about 300 bp and 4.6 kb (band B). Lane 2: DNA from *E. coli* EC317. The *Hin*dIII fragment containing the wild type allele of the fragment 28 sequence was about 3 kb (band A). Lanes 3 and 4: DNA from single recombinant mutants. The allele replacement plasmid was inserted into the chromosome. Three bands were expected: the band 4.6 kb was the plasmid sequence (band B), and two bands were from the disrupted chromosomal sequence of fragment 28 (bands C and D). Bands C and D combined were the same size as the wild type allele (band A) plus an additional 450 bp from plasmid sequences, which was expected. Lanes 5 and 6: DNA from the double recombinant mutants. Two bands were expected (bands D and E) to be same size as the wild type fragment when combined. The expected bands were seen for all of the mutant strains that indicates that the allele replacement occurred, but the presence of the wild type allele (shown by the white circle) in all of the strains indicated that there were multiple copies of the fragment 28 sequence in *E. coli* EC317.

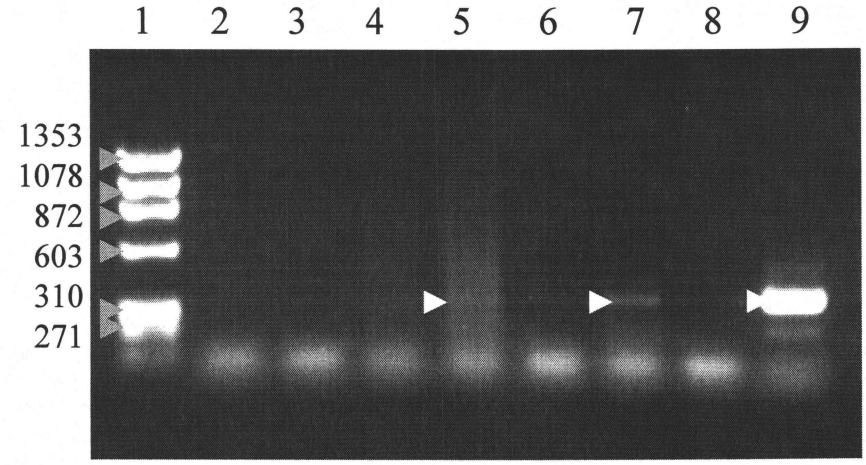
fragment sequence (single recombination event) were used as controls for the blot. The expected bands were present in both the single and double recombination mutants (allele replacement mutants), but bands corresponding to the wild type allele were also present. These results indicate that there are multiple copies of the fragment 28 sequence in the genome of *E. coli* strain EC317. Since the two allele replacement mutants still had intact copies of the fragment 28 sequence, they were not tested for altered virulence.

Reverse transcription PCR (RT-PCR) was done to confirm that ORF1 from fragment 3-54 was expressed *in vivo*. Chickens were infected with *E. coli* EC317 using the cellulitis scratch model of infection. Two days after the challenge had been administered, the liver and spleen were removed from one of the euthanized birds. The tissue was homogenized and the concentration of bacteria was determined using an aliquot of the homogenate. RNA was extracted from the remaining homogenate and used to generate cDNA. The cDNA was also generated from liver and spleen samples from an uninfected bird. The cDNA was used as the template for PCR reactions with two sets of primers (Figure 4-29). The positive control was with primers to *gapA*, which is a constitutively expressed enzyme. Expression of gapA and ORF1 was not seen in the uninfected chicken tissues, which indicates that the primers were specific to the *E. coli* sequences. ORF1 expression was detected in the infected liver and the spleen, though the expression in the liver was barely detectible. The expression of *gapA* was detected in the liver but not in the spleen. These results indicate that ORF1 of fragment 3-54 was expressed under *in vivo* conditions.

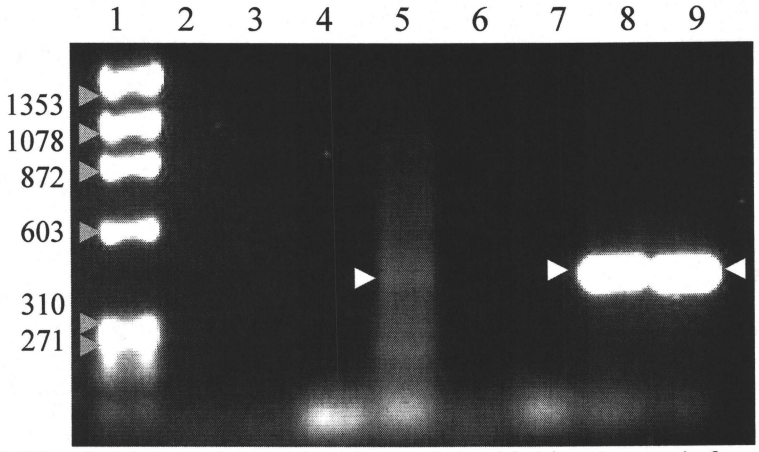
The allele replacement with fragment 3-54 was successful and produced the strain *E. coli* EC2549. PCR was initially used to analyze strain EC2549. The PCR reaction for the mutant had the band found with the allele replacement vector, and did

Figure 4-29: The RT-PCR of fragment 3-54 ORF1.

A)



B)



The RT-PCR of chicken tissue samples, using chicken tissues infected with *E. coli* EC317. Lane1: Phi X-174 restriction fragment DNA digested with *Hae*III. The size of the bands is indicated by the gray arrows, and the sizes in bp are listed. Lanes 2-9 contain the PCR products generated using different templates. Lane 2: RNA of infected chicken liver. Lane 3: RNA of uninfected chicken tissue. Lane 4: RNA of infected chicken spleen. Lane 5: cDNA produced from infected chicken liver. Lane 6: cDNA produced from uninfected liver and spleen tissue. Lane 7: cDNA from infected chicken spleen. Lane 8: chromosomal DNA from *E. coli* K-12. Lane 9: chromosomal DNA from *E. coli* EC317. A) The PCR reaction used primers to ORF1 of fragment 3-54. Lanes 2 to 4 were negative controls testing for DNA contamination of the RNA samples. Lane 6 was a negative control, which shows that the primers to 3-54 do not produce a product from chicken RNA. Lane 9 was the positive control. The bands indicated by the white arrows were the expected product (350 bp) for primers to fragment 3-54. The bands in lanes 5 and 7 indicate fragment 3-54 was transcribed by *E. coli* EC317 under *in vivo* conditions. B) The PCR reaction used primers to *gapA*, which should be expressed constitutively by *E. coli*. The expected band was seen for the liver (416 bp, white arrows). The band was not seen for the spleen sample, as too few bacteria were in the spleen for identification by RT-PCR.

not have the band found in the wild type strain. This result indicated that the fragment 3-54 sequence was replaced by the sequence containing the Ω kanamycin cassette in strain EC2549. The PCR results were confirmed by Southern blot analysis (Figure 4-30), which showed the fragment 3-54 had increased in size by 2.3 kb, the size of the Ω kanamycin cassette. The results from both the PCR analysis and the Southern blot analysis indicate that *E. coli* strain EC2549 contained the Ω kanamycin cassette inserted into the sequence of fragment 3-54.

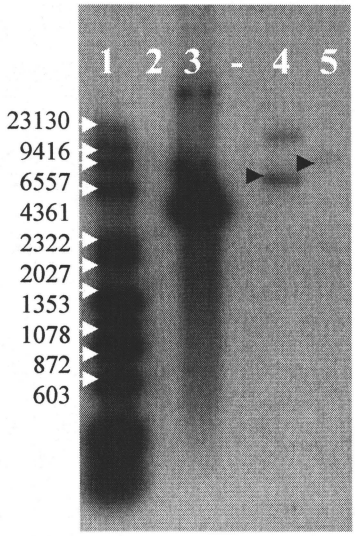
4.5.3 Testing for Attenuation of the Mutant Strain

Two models of infection were used to test *E. coli* strain EC2549 for attenuation. A day-old chick model of infection was used as a general test for virulence. A cellulitis scratch model was used to test for the ability of the strain to cause cellulitis and systemic lesions.

4.5.3.1 Test of the Virulence of Strain EC2549 with the Day-Old Chick Model

The day-old chick model of infection was a general test of virulence in broilers, but did not mimic a natural mode of infection. Three *E. coli* strains were used: EC317, EC2549, and EC106. Strain EC317 was the pathogenic strain, from which the mutant strain EC2549 was derived. *E. coli* strain EC106 was a non-pathogenic avian isolate (Ngeleka, Kwaga *et al.*, 1996). The three strains were tested at two doses,

Figure 4-30: Southern blot analysis of *E. coli* EC317 strain with the insertion in the fragment 3-54 sequence.



1) Markers
2) Vector
3) Replacement Vector
4) Genomic *E. coli* EC 317
5) Double Recombination Mutant

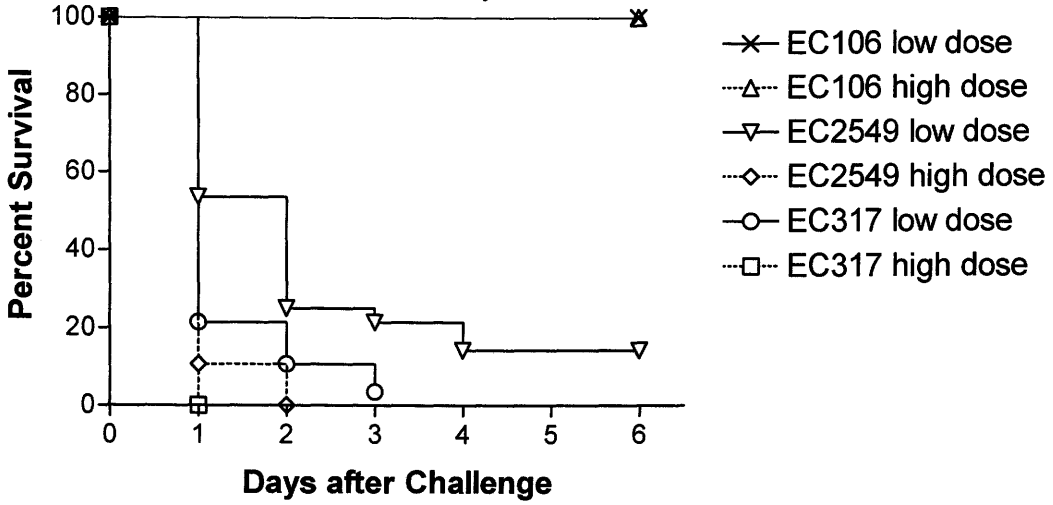
The Southern blot was probed with the 3-54 sequence. The marker lane was radiolabeled Drigest III marker, and the sizes of the bands in bp were indicated. The DNA used was digested with *EcoR*I. Lane 4: genomic DNA from *E. coli* EC317. Lane 5: genomic DNA from *E. coli* strain EC2549, which was the double recombination mutant. Lane 3: The replacement vector, which was expected to give a band of about 3.4 kb. Lane 2: the pIB307 plasmid, which was the negative control. The mutant with the disrupted fragment had a band 2.3 kb larger than the band in the *E. coli* EC317 lane, which was the size of the Ω kanamycin insertion. The black arrows indicated the bands containing the fragment 3-54 sequence. The second band in the *E. coli* EC317 lane was probably due to incomplete digestion of DNA. Other Southern blots have shown that EC317 contains only one copy of fragment 3-54.

approximately $1x10^6$ cfu for the high dose and $5x10^4$ cfu for the low dose. The birds were challenged by subcutaneous injection in the neck and euthanized if they met predetermined criteria for severe illness (Appendix A4).

The survival curves were significantly different when the negative control groups (challenged with strain EC106) were compared to the other groups (Figure 4-31). The group challenged with the low dose of strain EC2549 was significantly different from the group challenged with the low dose of EC317 (P=0.0142). The dead birds were examined for pericardial lesions and the pericardial fluid was cultured. The results of the postmortem examinations are shown in Table 4-9. Three criteria were used to determine if the birds were affected by the challenge: isolation of *E. coli* from the pericardial fluid, presence of systemic lesions, or death of the birds less than five days after the challenge. At both challenge doses, the birds affected were not significantly different for those challenged with *E. coli* EC317 or *E. coli* EC2549 (P= 0.6110 for the low dose). The results of postmortem analysis indicates that there was no difference between the groups challenged with the low dose of *E. coli*, which indicates that the difference seen in the survival curves was not biologically significant. These results indicate *E. coli* strain EC2549 was not attenuated as tested in this model.

Based on these results, we concluded that the mutation of the fragment 3-54 sequence in strain EC2549 did not lead to a reduction in the virulence of the strain using the day-old chick model.

Figure 4-31: Survival of birds in the day-old chick model of infection.



The survival of day-old chickens after subcutaneous challenge on the neck was plotted over time. All of the surviving birds were euthanized and examined for lesions on day six. The last surviving bird in the EC317 low dose group had to be euthanized for humane reasons, but did not have a clinical score that warranted this measure. All of the groups challenged with *E. coli* strains EC317 and EC2549 were significantly different from the groups challenged with *E. coli* EC106. There was a significant difference between the survival of the birds challenged with the low dose of *E. coli* EC317 and *E. coli* EC2549 (P= 0.0142).

Table 4-9: Postmortem examination of birds from the day-old chick challenge.

EC #	Dose (cfu)	Phenotype	Mortality[a]	Lesions[b]	Positive Culture[c]	Birds Affected[d]
106	4.3×10^4	Non-pathogenic	0/28	0/28	0/28	0/28
106	6.0×10^5	Non-pathogenic	0/28	0/28	0/28	0/28
2549	2.8×10^4	Mutant, Km^R	24/28	13/28	25/28	25/28
2549	7.3×10^5	Mutant, Km^R	28/28	3/28	28/28	28/28
317	7.7×10^4	Pathogenic	27/28*	6/28	27/28	27/28
317	1.1×10^6	Pathogenic	28/28	0/28	28/28	28/28

[a] Mortality includes birds that had died and that were euthanized for high clinical scores.
[b] Lesions form after 36 hours, so birds that died or were euthanized before that time would not be expected to have lesions.
[c] Cultures were from pericardial fluid and were plated on both MacConkey Agar and MacConkey agar with kanamycin.
[d] Birds affected were those that had lesions after 36 hours, or a positive pericardial culture any time post challenge.
* One bird was terminated because it was the last remaining animal, but it did not have clinical scores to warrant euthanasia..

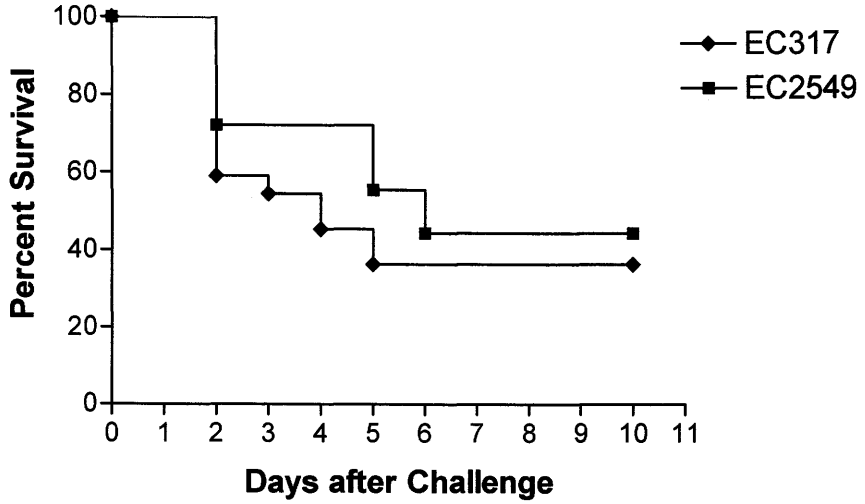
4.5.3.2 Test of the Virulence of Strain EC2549 with the Cellulitis Scratch Model

The cellulitis scratch model was designed to mimic the conditions believed to cause cellulitis (Gomis, Babiuk *et al.*, 2002). This model was also used to determine if *E. coli* strain EC2549 was attenuated.

Cellulitis is caused by the infection of a skin break by *E. coli* strains. The cellulitis scratch model used 25 day-old broilers, which were scratched twice on the abdomen and log phase *E. coli* swabbed onto the scratches. The birds were euthanized if they met predetermined criteria (Appendix A5) and examined for lesions. One group was challenged with *E. coli* EC317, which was the positive control, and one group was challenged with mutant *E. coli* EC2549. The cellulitis lesion, pericardial fluid, and air sacs were cultured, and the size of the cellulitis lesion was measured (Table 4-10). The survival time for each group was shown in Figure 4-32. There was no significant difference between the birds challenged with *E. coli* EC317 and those challenged with *E. coli* EC2549, either for survival time or for the production of lesions.

No alteration in the virulence of the mutant strain tested was observed in the day-old chick model or the cellulitis scratch model. We concluded that ORF1 from fragment 3-54 does not contribute to the pathogenesis of *E. coli* EC317 in the models tested.

Figure 4-32: Survival of birds challenged by the cellulitis scratch model.



The survival of 25 day-old chickens after inoculation of *E. coli* EC317 or *E. coli* EC2549 onto full thickness scratches, was plotted over time. All of the remaining birds were euthanized and examined for lesions, on day ten. The survival of the birds in the two groups was not significantly different.

Table 4-10: Postmortem examination of broilers challenged by the cellulitis scratch model.

Challenge strain	EC2549	EC317
Mortality[a]	56% (10/18)	64% (14/22)
Cellulitis lesions	18/18	22/22
Cultures of cellulitis[c]	16/18	19/24
Average size of cellulitis (cm^2)	14.7	15.4
Pericarditis	11/18	14/22
Cultures of pericardial fluid[c]	10/18	11/22
Air sacculitis	11/18	16/22
Cultures of air sacs[c]	9/18	10/22
Birds affected by systemic lesions	12/18	17/22

[a] Mortality includes birds that died or that were euthanized for high clinical scores.
[b] Birds affected by systemic lesions were defined as those with air sacculitis, pericarditis or other lesions, or those with positive cultures for *E. coli* from pericardial fluid or air sacs.
[c] Cultures were plated on both MacConkey Agar and MacConkey agar with kanamycin.

5 DISCUSSION AND CONCLUSIONS

5.1 Signature Tagged Mutagenesis

We were unable to generate a pool of transposon mutants that were consistently positive by colony blot analysis (Sections 4.1 and 4.2). The reason for this variability is not known. Attempts were made to optimize the generation of the probes, and alterations were made to the colony blot protocol. Unfortunately, these alterations failed to solve the problems encountered with the technique.

Two aspects of the technique were not changed, and these may have been the source of the problems experienced with the colony blots. The two areas that were not altered were the bacterial strain used to make the transposon insertion library, and the tagged transposons used. It is possible that some attribute of *E. coli* EC317 makes it particularly unsuitable for use with STM. The STM technique has been used with *E. coli* K1 (Gonzalez, Lichtensteiger *et al.*, 2001) and with other bacterial strains (Hensel, Shea *et al.*, 1995; Mei, Nourbakhsh *et al.*, 1997; Chiang and Mekalanos, 1998). It seems unlikely that this was the problem, as STM has been successful with a wide variety of bacteria.

The other aspect of the STM technique unaltered in this work was the library of plasmids used to create the transposon insertion library. We obtained the tagged

transposons on a suicide vector from David Holden, who did the original work using *S. typhimurium* (Hensel, Shea *et al.*, 1995). A problem with the construction of the plasmid library could affect the results obtained from using the library. If the primer binding sites, which flank the tagging sequences were altered, the efficiency of the PCR reaction would be greatly reduced. This reduction in the PCR efficiency could in turn affect the results of the colony blot analysis since some of the tags would be amplified to a higher level than others, affecting the signal strength on the colony blots. To determine if the tagging sequences were in some way defective, it would have been necessary to synthesise new tags and make a new library of tagged transposons, then test them in *E. coli* EC317 to see if it functioned better than the library obtained from David Holden. Since this test was not done, it is impossible to know if the technique could be made to work in *E. coli* EC317.

Other researchers have used modifications to the STM technique so that they did not have to use random tagging sequences. The first modification was to identify 96 strong signal tags by colony blot and use them individually to generate the transposon insertion mutants (Mei, Nourbakhsh *et al.*, 1997). Pools of 96 mutants, with each mutant containing a different tag, were used with this STM technique. Another alteration was to identify the members in the pools using PCR analysis instead of using colony blot analysis (Lehoux, Sanschagrin *et al.*, 1999). These modifications to the STM technique suggest that other laboratories also experienced problems using the original STM technique. We started to investigate the first modification to STM, but only identified 36 strongly positive tags before deciding to use the suppression subtractive hybridization technique.

5.2 Subtractive Hybridization Discussion

We used suppression subtractive hybridization to identify sequences found in the genomes of two *E. coli* strains of avian origin, but not in the genome of *E. coli* K-12 (Section 4.4.3). A total of 65 fragments were identified. Three of the fragments were duplicated, so only 62 of the fragments were unique. More than half of the 62 sequences identified were homologous to phage or plasmids sequences (Table 4-3, Section 4.4.4). Both *E. coli* strains contain plasmids, so it was unsurprising that some of the fragments had homology to plasmid sequences. All but one of the phage sequences identified were derived from *E. coli* EC317. The presence of a high number of sequences with homology to phages was also predictable, since this strain is known to be lysogenic for an unidentified phage (B. Allan unpublished data). Similar results were seen when the genomes of *E. coli* O157:H7 Sakai and *E. coli* K-12 were compared. More than half of the *E. coli* O157:H7 specific sequences were phage sequences (Hayashi, Makino *et al.*, 2001).

The subtractive hybridization done in this work identified factors known to be involved in the virulence of avian *E. coli*, as well as sequences not known to be involved in virulence. A fragment with homology to a gene involved in the synthesis of adherence factors was found, and adherence factors have been shown to have a role in the virulence of avian *E. coli* strains (Dho-Moulin and Fairbrother, 1999). In addition, two fragments were identified with homology to *tsh*, which has been previously demonstrated to be associated with avian *E. coli* strains (Ngeleka, Brereton *et al.*, 2002). These results indicate that the subtractive hybridization technique can be used to identify potential virulence factors in *E. coli* strains.

We did not sequence enough of the fragments to identify all of the tester-specific DNA generated using SSH. The sequence analysis of the SSH fragments generated in this work identified three pairs of fragments with the same sequence. If we had sequenced enough fragments to find many of the sequences more then once, we may have approached identification of all the tester specific fragments. The objective of this work was not to identify all of the differences between *E. coli* K-12 and the two avian *E. coli* strains, so it was not considered a priority to sequence the fragments until all of the tester specific DNA had been found.

In a previous study (Brown and Curtiss, 1996), the DNA from an *E. coli* K-12 strain was subtracted from the DNA from an avian *E. coli* strain. The study identified areas of the chromosome where unique DNA sequences were inserted with respect to the *E. coli* K-12 genome. Removal of two of the regions of DNA unique to the avian strain resulted in attenuation. The DNA sequences of the regions were not reported, so their similarity to the fragments we identified cannot be determined. .

There are indications that many of the fragments identified in this study have been acquired by the horizontal transfer of DNA. More than half of the sequences identified had homology to phage or plasmids, both of which are known to be vectors for the horizontal transfer of genes. There were fragments with homology to IS elements, transposases, and pathogenicity islands, which are associated with the horizontal transfer of DNA. When the genomes of *E. coli* O157:H7 Sakai and *E. coli* K-12 were compared, a large number of mobile genetic elements were found, such as prophage sequences and IS elements. It is likely that phage and transposons have played a large role in the evolution of the avian *E. coli* strains, since a high number of these sequences were found in the avian specific DNA found in this study.

Another indication of transfer of DNA between strains is difference in the GC content of areas of the genome. The GC content of some of the fragments varied from the average GC content for *E. coli* strains (50%)(Hayashi, Makino *et al.*, 2001)(Table 4-3). Fragments with differing GC content may have been acquired by horizontal transfer of DNA. An example of this is fragment 3-19, which was 1726 bp long and had a GC content of 32%. Some of the fragments identified in this study may be too small to accurately determine if the GC content is significantly different from that of the genome.

Twenty-five fragments were chosen for further study (Table 4-3). The majority of the fragments chosen did not have known function. The distribution of these fragments in the ECOR collection of reference strains (Section 4.4.6) and in a collection of 41 avian isolates (Section 4.4.5) was determined. The avian isolates contained 39 cellulitis isolates and two turkey septicemia isolates. The turkey isolates were known to cause cellulitis in a disease model (B. Allan, unpublished data). The ECOR collection consists of 72 strains isolated from mammals from diverse geographic locations. The strains were originally assembled to represent the genetic diversity of the *E. coli* species as determined by multilocus enzyme electrophoresis (MLEE)(Ochman and Selander, 1984). Except for 10 strains, which were isolated from women with urinary tract infections (UTI), all 72 of the strains were isolated from feces of mammals. The UTI isolates were chosen to represent pathogenic strains from each major cluster of *E. coli* strains, as defined by Ochman and Selander (Ochman and Selander, 1984). Each of the urinary tract isolates had the same MLEE pattern as one of the fecal isolates.

Sequences that were more highly conserved in avian *E. coli* strains than in non-avian *E. coli* strains may play a role in the virulence of the avian *E. coli*. Each subtractive hybridization fragment was found in some of the ECOR strains. The

distribution of the fragments in the UTI strains did not differ significantly from the other strains in the ECOR collection (P= 0.51). The distribution of fragments in both the UTI strains and in the rest of the ECOR collection was compared to the distribution of fragments in the avian strains. The distribution was significantly different in both cases. It appears that, despite the fact that the UTI strains are pathogenic. they do not contain any more of the fragments derived from *E. coli* that are pathogenic for poultry than the rest of the ECOR strains. It highly possible that some of the ECOR strains contain avian virulence factors, even though they are not known to be pathogenic for poultry.

All of the fragments investigated were found to be present in both the avian collection of strains and the non-avian isolates of the ECOR collection (Tables 4-4 and 4-5). These fragments therefore, were not unique to avian *E. coli* strains. None of the fragments were found in all of the avian *E. coli* strains tested, so it appears that they are not essential for virulence, although some of the fragments may still play a role in the virulence of avian *E. coli* strains. The number of avian strains that contained each fragment varied, which indicates that the genomes of the avian *E. coli* strains are flexible.

Since virulence in avian *E. coli* is multi-factorial, it is expected that not every avian *E. coli* strain will contain every virulence factor. The fact that none of the fragments were found in all of the avian isolates does not eliminate the possibility that the fragments have a role in the virulence of avian *E. coli*. It would, however, be expected that the fragments with a role in virulence would be more strongly associated with the avian *E. coli* strains than the non-avian *E. coli* strains. Eleven fragments were found at higher frequencies in the avian strains than the ECOR strains. This fact suggests that these fragments may be involved in the pathogenesis of poultry infections.

The role of these eleven subtractive hybridization fragments in virulence of *E. coli* in poultry has not been tested. Some of these fragments had sequence homology to known virulence factors. The synthesis of LPS is variable between strains of *E. coli,* so the isolation of LPS genes in an SSH fragment was expected. LPS has been shown to be involved in resistance to serum killing and virulence (Nataro and Levine, 1994). The entire sequence of fragment 296 had 98% homology to the *chuA* gene, which codes for an outer membrane protein involved in heme uptake in EHEC strains (Torres and Payne, 1997). Heme uptake may be important to virulence since the ability to acquire iron has been shown to be important in many bacterial pathogens (Ratledge and Dover, 2000). The identification of this fragment and other fragments with homology to virulence factors indicates that the SSH technique is a valid method for identifying virulence factors in bacteria strains.

Some of the fragments also had homology to pathogenicity islands of other bacterial strains. Fragment 308, which is 542 bp long, had 92% homology over 277 bp to part of orf7 of the *she* pathogenicity island of *S. flexneri* 2a (Al-Hasani, Rajakumar *et al.*, 2001). It also demonstrated 92% homology over 500 bp to a putative transposase in the locus of enterocyte effacement II pathogenicity island (LEE II). Pathogenicity islands contain a high number of virulence factors, but not all of the sequences in the islands are related to the virulence of the strain, so the homology of SSH fragments to ORFs of pathogenicity islands does not indicate that they are involved in virulence.

Three of the twenty-five fragments were found in more of the ECOR strains than the avian *E. coli* strains, which indicates that they may not be virulence factors of avian *E. coli* strains. One of these fragments was number 1-74, which has 117 bp at one end that was 100% homologous to part of the *bfpA* gene. The region of homology was

at the end of fragment 1-74 sequence, so it is possible that the genome contains a complete copy of the *bfpA* gene. BfpA is the structural subunit of bundle forming pili and is a virulence factor of EPEC strains (Bieber, Ramer *et al.*, 1998). The Bfp pili are responsible for initial attachment of EPEC strains to the intestinal epithelium in human disease (Bortolini, Trabulsi *et al.*, 1999). It is not known if Bfp can mediate the attachment of *E. coli* to the intestinal epithelium or other tissues of poultry. The avian strains used for this study were isolated from extraintestinal infections, so it is unlikely that intestinal adherence factors are important during infections with these strains. Since this fragment is more frequent in ECOR than avian strains, it seems probable that this fragment does not function in the virulence of avian *E. coli,* and has only been obtained by horizontal transfer in some of the avian strains.

One of the other fragments that was found in the ECOR collection more often than in the avian *E. coli* strains was fragment 1-47. This fragment had high homology to a putative fimbrial usher protein, so its proposed function would be to aid in the assembly of fimbria. Adhesive factors like fimbria can be virulence factors. The fact that this fragment was not associated with the avian *E. coli* strains suggested this putative usher was required for the assembly of a fimbria, which does not function in the infection of extraintestinal sites of poultry. It is also possible that the ORF codes for a protein with a different function.

Other studies using subtractive hybridization techniques with different types of pathogenic *E. coli* strains have shown results similar to the work presented here. The results of a study using a neonatal meningitis strain (Bonacorsi, Clermont *et al.*, 2000), and a study using an uropathogenic *E. coli* strain (Janke, Dobrindt *et al.*, 2001), identified sequences with homology to LPS synthesis, adherence, and iron metabolism

genes. Our study also identified sequences with homology to genes that code for these functions. Eight of the fragments that were identified in our study had very high homology with sequences derived from the neonatal meningitis strain. These neonatal meningitis sequences did not have known functions. In the study that examined a uropathogenic strain of *E. coli* (UPEC), the prevalence of some of the sequences in a collection of pathogenic and non-pathogenic *E. coli* isolates was determined. Generally, the sequences were more often found in extra-intestinal isolates than in fecal isolates, but two sequences were identified which were highly specific for UPEC strains. Both the UPEC study and our study showed that the fragments identified were present in both pathogenic and non-pathogenic isolates.

Eleven of the SSH fragments had high homology to *E. coli* O157:H7 genomic sequences. *E. coli* O157:H7 is an EHEC strain and is known to contain the LEE pathogenicity island. The LEE pathogenicity island contains the genes for the attaching and effacing phenotype seen in a number of enteric pathogens. These include *Citrobacter rodentium* (Deng, Li *et al.*, 2001), EPEC strains (Elliott, Wainwright *et al.*, 1998), EHEC strains (Perna, Mayhew *et al.*, 1998), and other *E. coli* strains that cause diarrhea in calves, dogs, pigs, and rabbits (Zhu, Agin *et al.*, 2001). Fragment 1-63, which was identified in this work, was homologous to a portion of the LEE pathogenicity island that is a prophage, and not to the part of LEE that is known to code for virulence factors.

The fact that 11 fragments were homologous to *E. coli* O157:H7 DNA and that one fragment had homology to part of a LEE pathogenicity island, suggested that *E. coli* strains pathogenic for poultry might contain some of the virulence factors identified in LEE. The collection of avian *E. coli* isolates were screened for the presence of *eae*,

espA, *espB*, *espD,* and *tir*, genes from the LEE pathogenicity island found in an *E. coli* O157:H7 strain. All of these genes encode proteins that are important to the formation of attaching and effacing lesions. Intimin (*eae*) is expressed on the surface of the bacteria. Tir is inserted into the host cell membrane, where it binds to intimin on the bacteria. Together, intimin and Tir allow the bacteria to adhere tightly to the host cell surface (Kenny, DeVinney *et al.*, 1997). The EspA, EspB, and EspD proteins function in the transfer of Tir into intestinal epithelial cells (Kenny, 2002).

None of the 41 avian isolates used in this study contained the *eae*, *espA*, *espB*, *espD* or *tir* genes. The two strains (*E. coli* EC317 and *E. coli* EC2029) used to generate the subtractive hybridization fragments also did not contain the EHEC genes. Other researchers have shown that nine of 763 *E. coli* isolates from extraintestinal infections of poultry examined contained the *eae* gene (Stordeur, Marlier *et al.*, 2002). The strains used in that study were isolated from chickens, turkeys, and ducks, and the occurrence of the *eae* gene in the isolates in their study was only 1.2%. We did not find *eae* in any of the avian *E. coli* strains, but we investigated a much smaller number of strains than were used in the other study.

It is known that *E. coli* strains that have the attaching and effacing phenotype may be involved in a disease of poultry called poult enteritis-mortality syndrome (Guy, Smith *et al.*, 2000). The study by Guy *et al.* (Guy, Smith *et al.*, 2000) showed that co-infection of turkeys with turkey coronavirus and a turkey EPEC isolate could reproduce the mortality and lesions seen with poult enteritis-mortality syndrome, though other bacteria and viruses have also been associated with the disease (Barnes and Guy, 1997). The EPEC strain used was isolated from a turkey with poult enteritis-mortality syndrome and was able to cause attaching and effacing lesions in turkeys. Attaching and

effacing *E. coli* strains do appear to be involved in enteric poultry disease (Guy, Smith *et al.*, 2000); however they are probably not involved in extraintestinal diseases of poultry.

5.3 Characterization of Selected Fragments

In an attempt to define the limits of the DNA not found in *E. coli* K-12, the DNA flanking some of the fragments in the genome of the avian *E. coli* strains was sequenced. The size of the DNA fragments identified by SSH was not based on the size of the region of DNA found in the avian strain and absent in *E. coli* K-12, but is a reflection of the technique used to identify it. Since the SSH fragments were *Rsa*I restriction digest fragments, the size of the fragments identified was due to the distribution of *Rsa*I sites. PCR was used to amplify the sequences specific to the tester strain (either *E. coli* EC317 or EC2029), and was biased towards amplifying smaller sequences (less then 1kb) (Lisitsyn, Lisitsyn *et al.*, 1993). The use of an *Rsa*I digest to generate the fragments of tester DNA, and the use of PCR to amplify the tester specific sequences, favors the identification of the tester specific regions with frequent *Rsa*I sites. This limitation will prevent the identification of all tester specific DNA, even if large numbers of fragments are sequenced. The fragments identified in this work represented a fraction of the DNA present in *E. coli* EC317 and *E. coli* EC2029, but not in *E. coli* K-12. The fragments were small portions of larger tester specific regions, and the sizes of these regions were not related to the sizes of the SSH fragment. One way to identify more of the DNA found in avian strains, but not in *E. coli* K-12, is to sequence the DNA flanking the SSH generated fragments.

In order to sequence the DNA flanking the SSH fragments, a cosmid library derived from *E. coli* EC317 was screened for the presence of some of the SSH fragments. Initially, 17 fragments were chosen for this work. All of these fragments had been used previously to determine the degree of conservation of the fragments in the avian isolates and ECOR strain collection. The number of fragments chosen from the *E. coli* EC2029 library was limited because it was thought they may have high enough homology to *E. coli* EC317 to be positive in the colony blot analysis and in the Southern blot analysis, but sufficiently different to make sequencing difficult. It was thought that there might be enough differences between the sequences in the two strains that the sequencing primers designed from *E. coli* EC2029 DNA would not bind efficiently to the sequence in *E. coli* EC317. As a result, only two of the 17 fragments used were derived from *E. coli* EC2029. Fragment 1-63 derived from *E. coli* EC2029 was found on a cosmid and successfully sequenced, so this concern was not warranted (Section 4.4.8.2).

The first protocol, using probes made by random priming to screen the cosmid library, generated 68 false positives and 2 true positives. Southern blot analysis of the false positive clones showed cross reactivity between the cosmid vector and the probes (Figure 4-15). The plasmid containing the SSH fragments had an 1828 bp region which was 98% homologous to the cosmid vector. It was believed that contaminating plasmid sequences were labeled with the SSH fragment when the probe was made, which caused the cross-reaction with the cosmid sequences. The level of cross reactivity with the cosmid vector varied for the probes to different SSH fragments, probably due to different levels of vector contaminating the restriction fragment used to generate the

probes. It is likely that the fragment 453 probe contained the least amount of vector DNA, since two cosmids that contained fragment 453 were identified.

For subsequent work, the probes were generated by PCR with radiolabeled nucleotides, and the membranes were blocked with DNA from the pT-Adv, which was the plasmid containing the SSH fragments. Using the modified protocol reduced the number of false positives, and located fragments 296, 358, 361, and 1-63 on cosmids. Due to the expense of sequencing, we chose to sequence the cosmids containing fragments 296, 453 and 1-63. The areas sequenced were from 3000 to 4000 bp, which was not long enough to cover the complete area in the avian *E. coli* strains that was not in *E. coli* K-12. One junction of the fragment 296 region with *E. coli* K-12 DNA was identified, but not the other. The cosmids were not sequenced to the limits of the regions absent in *E. coli* K-12 due to time constrains.

Cosmid 2483 contained fragment 296 and 3033 bp of it was sequenced. A section 3018 bp long was 96% homologous at the DNA level to the *shu* gene cluster of *Shigella dysenteriae* (Figure 4-20). The cosmid sequence did not cover the entire *shu* gene cluster, but contained the *shuS* gene and most of the *shuA* gene. The ShuS protein was known to bind both heme and double stranded DNA (Wilks, 2001). It was thought to function in storage of heme and in the protection of DNA from oxidation induced by high iron levels. The ShuA protein is an outer membrane heme receptor (Wyckoff, Duncan *et al.*, 1998). A *shuA* mutant of *S. dysenteriae* was able to invade and grow in cultured cells, so it was not known what role it may have played in virulence. ChuA was the *E. coli* homologue of ShuA and has been found in *E. coli* strains that cause diarrhea and extraintestinal infections. Both ShuA and ShuS function in the use of heme as an iron source by bacterial strains. Since iron is a limiting nutrient for the growth of *E. coli*

in vivo, the *shu* gene cluster may allow the avian *E. coli* strains an additional method for obtaining iron in the host. The *shu* gene cluster may allow the avian *E. coli* strains to grow more rapidly *in vivo*, increasing the virulence of the *E. coli* strains. Fragment 296 was found in significantly more of the avian *E. coli* strains than the ECOR strains, which lends weight to the theory that the *shu* gene cluster may be important to virulence of the avian *E. coli* strains.

We were unable to demonstrate that the *shu* gene cluster found on the cosmid functioned in heme uptake (Section 4.4.8.3.1). Since the *shu* gene cluster on the cosmid was not sequenced completely, it was possible that a deletion in the gene cluster prevented heme uptake by the strain containing the cosmid. Cosmids are known to be unstable (Sambrook, Fritsch *et al.*, 1989), so it is also possible that part of the *shu* gene cluster was deleted from the cosmid. We were unable to show heme binding with *E. coli* strains EC317 or EC2029 on solid media, but a liquid media assay was not tried. We tested the strains under only *in vitro* conditions, so it is possible that the *shu* gene cluster is functional *in vivo*. More work is necessary to determine if the *shu* gene cluster of avian *E. coli* strains functions *in vivo,* and if it is involved in virulence.

The regions flanking fragment 453 were also sequenced and 3905 bp was identified (Section 4.4.8.3.2, Figure 4-21). Unlike the sequences flanking fragment 296, these sequences did not exhibit very strong homology to sequences with known function. The sequence was analyzed and the results suggested that five sequences were ORFs. The five ORFs identified in the sequence that flanks fragment 453 did not have very high homology to any known proteins. One ORF may have been a phage structural protein. Another ORF had motifs from DNA and RNA binding proteins, so may have functioned in gene regulation. There was also a small ORF with very little homology or

motifs to anything, so it was possible that it did not code for a protein. The other two ORFs contained diverse motifs, thus it was hard to predict what functions they may have possessed. It was possible that this region coded for novel proteins, but without further investigation of the expression of these ORFs, it is impossible to tell what role, if any, they might play in avian *E. coli* strains.

A total of 3664 bp of the cosmid, which contained fragment 1-63, was sequenced (Figure 4-22). The sequence was homologous to parts of some pathogenicity islands. However, the sequences with homology to this area contained putative genes, so it was not known what functions the genes code for, if they code for anything at all. The fact that these sequences were associated with pathogenicity islands suggests two possible roles for the region. The first possibility was that the sequence codes for novel virulence factors; the second, that the sequence codes for the genes for the horizontal transfer of pathogenicity islands. Transposons, IS elements, plasmid and phage sequences have been associated with pathogenicity islands and are thought to be involved in transfer of the regions between strains (Hacker and Carniel, 2001). Pathogenicity islands are known to be transferred horizontally between strains (Johnson, Delavari *et al.*, 2001).

Based on the sequence analysis, it seems most likely that this region codes for genes involved in the transfer of DNA. The analysis of the sequence identified two ORFs with homology to transposons, one ORF with homology to a plasmid transfer gene, and one IS element. This cosmid sequence had homology to a portion of the LEE island of EHEC strains described as a prophage. The fact that there were three ORFs in this cosmid sequence demonstrating homology to transposons and to plasmid mobilization sequences, and that the region had homology to a prophage, was a strong

indication that this sequence functions in horizontal transfer of genes. More work is needed to determine what functions are encoded in this sequence.

In summary, the flanking DNA to three of the fragments was sequenced. The region that flanked fragment 296 was highly homologous to part of the *shu* gene cluster, encoding proteins needed for heme use by *S. dysenteriae*. It was not known if the entire *shu* gene cluster was present in *E. coli* EC317, or if it played a role in the virulence of the strain. The cosmid sequence containing fragment 453 did not have strong enough homology to any known sequences to predict its function with certainty. The sequence was analyzed for ORFs and putative genes. One of the five ORFs had homology to a phage structural gene. Another ORF had homology to DNA binding proteins, so it might have functioned as a regulatory protein. A putative function could not be assigned to the other three ORFs. The cosmid sequence that contained fragment 1-63 had homology to LEE pathogenicity islands. Some of the ORFs had homology to a prophage, a plasmid transfer gene, transposases and IS elements. The ORF homologies indicate that the region may have been involved in the horizontal transfer of DNA. Based on this data we could not conclude that these areas had any specific role in virulence, but that at least two of them may be involved in gene transfer.

5.4 Discussion of the Disruption Mutations

A virulence factor, by definition, is a gene product that is required for virulence. To prove that a gene product is a virulence factor, three things must be shown. First, the gene should be reasonably associated with the pathogen. Second, the disruption of the gene must attenuate the pathogenic strain. Finally, the restoration of the gene should

completely restore the virulence of the strain (Falkow, 1988). The distribution of the subtractive hybridization fragments in the avian *E. coli* strains and the ECOR strains was determined to see if the fragments were associated with the avian *E. coli* strains pathogenic for poultry. For 14 of the fragments, there was a significant difference between the distributions of the fragments in the two groups of strains. Eleven of those fragments were found significantly more often in the avian than in the non-avian strains. This confirms that the fragments are associated with the avian strains.

To prove that a gene codes for a virulence factor, it is necessary to demonstrate that there is a reduction in the virulence of the pathogen if the gene is disrupted. In order to show that the fragments identified in this study using SSH were related to the virulence of avian *E. coli*, we attempted to disrupt the sequence of selected fragments. Two subtractive hybridization fragments were disrupted by allele replacement in *E. coli* EC317. An antibiotic resistance gene was cloned into the fragment to facilitate selection of the disrupted allele. The disruptions were in the fragment 28 and fragment 3-54 sequences (Section 4.5.2.3). The Southern blot analysis of the mutant with a disruption in fragment 28 revealed that there are multiple copies of the fragment 28 sequence in the *E. coli* EC317 genome, and that only one of the copies had been disrupted (Figure 4-28). The mutant was not used for the attenuation studies since it still had at least copy of the wild type allele. A disruption in ORF1 of fragment 3-54 in *E. coli* EC2549 was confirmed by Southern blot analysis (Figure 4-30). ORF1 was shown to be expressed *in vivo* by RT-PCR analysis.

The *E. coli* strain EC2549 was tested for attenuation in two models of infection for poultry. The day-old chick model of infection was a general test of virulence, which did not mimic any specific avian disease. The virulence of the *E. coli* strains was

measured by the number of birds that developed significant disease, and by the presence of lesions in the birds. There was a difference between the survival curves for the birds in the groups challenged with the low dose of the mutant strain, *E. coli* EC2549 and wild type strain *E. coli* EC317 (P=0.0142) (Figure 4-31). This difference was not believed to be biologically significant, since the difference in the number of birds affected by the infection was not significant for the two groups (Table 4-9). It was concluded that *E. coli* strain EC2549 was not attenuated, as tested in this model.

The second animal model was used to test the virulence of the mutant strain was the cellulitis scratch model, which more closely mimicked a natural route of infection. *E. coli* cultures of EC317 and EC2549 were swabbed on full thickness scratches of the abdomen of 25 day-old broilers. The virulence of the *E. coli* strain was determined by the number of birds affected by the infection. The number of birds affected was determined by the number of the birds that developed clinical symptoms and the number of birds with lesions. There was no significant difference between the survival curves for the birds challenged with the two strains (Figure 4-32). There also was not a significant difference in the size or number of lesions produced in the birds of the two groups (Table 4-10). It was concluded that disruption of the 3-54 fragment sequence did not lead to attenuation as tested with the cellulitis scratch model. Fragment 3-54 was therefore not believed to play a role in the virulence of the avian *E. coli* pathogen EC317.

5.5 General Discussion and Conclusions

In this work, 62 DNA fragments not found in *E. coli* K-12 were identified in two avian *E. coli* strains. More than half of the SSH fragments were homologous to phage and transposon sequences, which indicates that the avian *E. coli* strains had been subject to horizontal transfer of genes. The 25 fragments selected for further study were not unique to avian isolates since each fragment was found in at least some of the ECOR strains. The percent of avian strains that contained each fragment ranged from 2 to 92%, indicating a considerable degree of genetic diversity among avian strains. We concluded that the genomes of the avian *E. coli* strains are flexible and contain genes from other *E. coli* strains and other bacteria.

It was expected that by using the SSH technique some of the fragments would encode phage sequences, LPS synthesis genes and previously identified virulence factors. The high number of phage sequences was anticipated, since other *E. coli* strains that have been sequenced showed a high number of phage sequences. As well, we knew that *E. coli* EC317 was lysogenic for a bacteriophage. The identification of LPS genes was not surprising, since the LPS synthesis is variable between *E. coli* strains. The identification of a SSH fragment with homology to the *tsh* gene, which is a known virulence factor of avian *E. coli* strains, confirms that the SSH technique can be used to identify virulence factors in bacterial strains.

Eleven of the fragments were found more frequently in the avian isolates than in the ECOR strains. No fragment was found in all of the avian *E. coli* strains and therefore, none of the fragments were essential for virulence. This does not rule out the possibility that some fragments may increase the virulence of some avian *E. coli* strains,

since the virulence of avian *E. coli* is known to be multifactorial. It seems likely that the virulence of avian *E. coli* strains is not dependent on any one virulence factor, since no one virulence factor has been identified in every avian *E. coli* strain tested.

Sequence analysis of the regions flanking selected SSH fragments revealed a number of sequences with homology to phage, plasmids, and other known sequences, as well as sequences without homology to known sequences. The region around fragment 296 had homology to a heme uptake pathway, but we were unable to prove that this pathway was functional in the two avian strains tested. The region flanking fragment 1-63 had homology to pathogenicity islands, and ORFs in the region had homology to transposons and plasmid transfer sequences. It is possible that these genes functioned in the horizontal transfer of DNA. The last region sequenced flanked fragment 453 and had homology to a diverse group of sequences. The diversity of the region made it difficult to suggest possible functions for the ORFs in the fragment.

The final stage of this work involved the disruption of some of the SSH fragments in the genome of *E. coli* EC317. OFR1 of fragment 3-54 was disrupted in the genomic sequence of *E. coli* EC317 and tested for attenuation in two models of infection. The virulence of the mutant strain and the parent strain *E. coli* EC317 was not significantly different in either of the animal models used. We concluded that the mutant strain was not attenuated in the disease models tested, and that ORF1 of fragment 3-54 of *E. coli* EC317 was not involved in the virulence in poultry.

The role of other fragments identified in the virulence of *E. coli* strain EC317 has not been determined. It would be interesting to investigate the role played in the virulence of *E. coli* EC317 by the eleven fragments more prevalent in the avian *E. coli* strains than in the ECOR strains. Fragment 28 was one of these fragments and attempts

to disrupt the sequence of fragment failed, since there was more then one copy of the fragment 28 sequence in *E. coli* EC317. Further work will be needed before the role of these sequences in the virulence of the avian *E. coli* strains can be determined.

The expression of ORFs on the fragments or on the cosmid sequences could be studied by RT-PCR or Northern blot analysis. It would be valuable to know if any of the fragments with putative functions, or without proposed functions, were expressed. The expression would confirm that the fragments are part of a true gene. Any of the proteins from any of the ORFs that are expressed could be studied so that the function of the protein could be identified.

The locations of the fragments identified in this study on the chromosome of the avian strains was not determined. Once the location of the fragments is known, further sequencing could identify the sequences flanking the fragments, distinguishing the limits of the DNA not found in *E. coli* K-12 and identifying any clusters of fragments. These clusters could indicate the presence of large genomic islands common to avian *E. coli* strains. This work would also improve our understanding of the genomic organization of the avian *E. coli* strains.

The hypothesis for this thesis was: there are DNA fragments which are highly associated with avian pathogenic *E. coli,* and virulence factors not yet identified are present in these DNA fragments. The identification of a known virulence factor of avian *E. coli* proves that the approach used was effective. Move work on the DNA fragments identified will be required to confirm that some of the fragments contain genes that code for virulence factors not yet identified.

6 REFERENCES

Agron, P. G., R. L. Walker, H. Kinde, S. J. Sawyer, D. C. Hayes, J. Wolard and G. L. Andersen (2001). "Identification by subtractive hybridization of sequences specific for *Salmonella enterica* serovar enteritidis." Appl Environ Microbiol **67**: 4984-4991.

Akerley, B. J., E. J. Rubin, A. Camilli, D. J. Lampe, H. M. Robertson and J. J. Mekalanos (1998). "Systematic identification of essential genes by *in vitro* mariner mutagenesis." Proc Natl Acad Sci USA **95**: 8927-8932.

Al-Hasani, K., K. Rajakumar, D. Bulach, R. Robins-Browne, B. Adler and H. Sakellaris (2001). "Genetic organization of the *she* pathogenicity island in *Shigella flexneri* 2a." Microb Pathog **30**: 1-8.

Allan, B. J., J. V. van den Hurk and A. A. Potter (1993). "Characterization of *Escherichia coli* from cases of avian colibacillosis." Can J Vet Res **57**: 146-151.

Altman, D. G. (1991). Practical Statistics for Medical Research. New York, NY, USA, Chapman and Hall.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res **25**: 3389-3402.

Angerer, A. and V. Braun (1998). "Iron regulates transcription of the *Escherichia coli* ferric citrate transport genes directly and through the transcription initiation proteins." Arch Microbiol **169**: 483-490.

Attwood, T. K., D. R. Flower, A. P. Lewis, J. E. Mabey, S. R. Morgan, P. Scordis, J. Selley and W. Wright (1999). "PRINTS prepares for the new millennium." Nucleic Acids Res. **27**: 220-225.

Ausubel, F. M., Ed. (1994). Current Protocols in Molecular Biology, Etobicoke, Ontario, Canada, John Wiley & Sons Canada Ltd.

Badger, J. L., C. A. Wass and K. S. Kim (2000). "Identification of *Escherichia coli* K1 genes contributing to human brain microvascular endothelial cell invasion by differential fluorescence induction." Mol Microbiol **36**: 174-182.

Barnes, H. J. (1994). Colibacillosis in Poultry. Lees Summit, Missouri, U.S.A., North American Animal Health Division, Pfizer inc.

Barnes, H. J. and W. B. Gross (1997). Colibacillosis. Diseases of Poultry. B. W. Calnek. Ames, Iowa, U.S.A., Iowa State University Press: 131-141.

Barnes, H. J. and J. S. Guy (1997). Poult enteritis-mortality syndrome ("spiking mortality") of turkeys. Diseases of Poultry. B. W. Calnek. Ames, Iowa, U.S.A., Iowa State University Press: 1025-1031.

Bateman, A., E. Birney, R. Durbin, S. R. Eddy, R. D. Finn and E. L. Sonnhammer (1999). "Pfam 3.1: 1313 multiple alignments match the majority of proteins." Nucleic Acids Res. **27**: 260-262.

Bellatin, J. A., A. S. Murray, M. Zhao and W. R. McMaster (2002). "Leishmania mexicana: identification of genes that are preferentially expressed in amastigotes." Exp Parasitol **100**: 44-53.

Bieber, D., S. W. Ramer, C. Y. Wu, W. J. Murray, T. Tobe, R. Fernandez and G. K. Schoolnik (1998). "Type IV pili, transient bacterial aggregates, and virulence of enteropathogenic *Escherichia coli*." Science **280**: 2114-2118.

Bjorkman, J., D. Hughes and D. I. Andersson (1998). "Virulence of antibiotic-resistant *Salmonella typhimurium*." Proc Natl. Acad Sci. USA **95**: 3949-3953.

Blomfield, I. C., V. Vaughn, R. F. Rest and B. I. Eisenstein. (1991). "Allelic exchange in *Escherichia coli* using the *Bacillus subtilis sacB* gene and a temperature-sensitive pSC101 replicon." Mol Microbiol **5**: 1447-1457.

Bogush, M. L., T. V. Velikodvorskaya, Y. B. Lebedev, L. G. Nikolaev, S. A. Lukyanov, A. F. Fradkov, B. K. Pliyev, M. N. Boichenko, G. N. Usatova, A. Vorobiev, G. Andersen and E. Sverdlov (1999). "Identification and localization of differences between *Escherichia coli* and *Salmonella typhimurium* genomes by suppressive subtractive hybridization." Mol Gen Genet **262**: 721-729.

Bonacorsi, S. P., O. Clermont, C. Tinsley, I. Le Gall, J. C. Beaudoin, J. Elion, X. Nassif and E. Bingen (2000). "Identification of regions of the *Escherichia coli* chromosome specific for neonatal meningitis-associated strains." Infect Immun **68**: 2096-2101.

Bortolini, M. R., L. R. Trabulsi, R. Keller, G. Frankel and V. Sperandio (1999). "Lack of expression of bundle-forming pili in some clinical isolates of enteropathogenic *Escherichia coli* (EPEC) is due to a conserved large deletion in the bfp operon." FEMS Microbiol Lett **179**: 169-74.

Brinboim, H. C. and J. Doly (1979). "A rapid alkaline extraction procedure for screening recombinant plasmid DNA." Nucleic Acids Res **7**: 1513-1523.

Brown, J. S., A. Aufauvre-Brown, J. Brown, J. M. Jennings, H. J. Arst and D. W. Holden (2000). "Signature-tagged and directed mutagenesis identify PABA synthetase as essential for Aspergillus fumigatus pathogenicity." Mol Microbiol **36**: 1371-1380.

Brown, P. K. and R. I. I. I. Curtiss (1996). "Unique chromosomal regions associated with virulence of an avian pathogenic *Escherichia coli* strain." Proc Natl Acad Sci USA **93**: 11149-11154.

Buchrieser, C., P. Glaser, C. Rusniok, H. Nedjari, H. D'Hauteville, F. Kunst, P. Sansonetti and C. Parsot (2000). "The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella flexneri*." Mol Microbiol **38**: 760-71.

Burland, V., Y. Shoa, N. T. Perna, G. Plunkett, H. J. Sofia and F. B. Blattner (1998). "The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli*O157:H7." Nucleic Acids Res **26**: 4196-4202.

Burns, S. M. and S. I. Hull (1998). "Comparison of loss of serum resistance by defined lipopolysaccharide mutants and an acapsular mutant of uropathogenic *Escherichia coli* O75:K5." Infect Immun **66**: 4244-53.

Burns, S. M. and S. I. Hull (1999). "Loss of resistance to ingestion and phagocytic killing by O(-) and K(-) mutants of a uropathogenic *Escherichia coli* O75:K5 strain." Infec Immun **67**: 3757-62.

Camilli, A., D. T. Beattie and J. J. Mekalanos (1994). "Use of genetic recombination as a reporter of gene expression." Proc Natl Aced Sci USA **91**: 2634-2638.

Camilli, A. and J. J. Mekalanos (1995). "Use of recombinase gene fusions to identify *Vibrio cholerae* genes induced during infection." Mol Microbiol **18**: 671-683.

Canadian Food Inspection Agency (1986-1998). Information Bulletin on Condemnations in Canada., Canadian Food Inspection Agency.

Chiang, S. L. and J. J. Mekalanos (1998). "Use of signature-tagged transposon mutagenesis to identify *Vibrio cholerae* genes critical for colonization." Mol Microbiol **27**: 797-805.

Christiansen, K. H., D. W. Hird, K. P. Snipes, C. Danaye-Elmi, C. W. Palmer, M. D. McBride and W. W. Utterback (1996). "California national animal health monitoring system for meat turkey flocks-1988-89 pilot study: management practices, flock health, and production." Avian Dis **40**: 278-284.

Claus, H., M. Frosch and U. Vogel (1998). "Identification of a hotspot for transformation of *Neisseria meningitidis* by shuttle mutagenesis using signature-tagged transposons." Mol Gen Genet **259**: 363-371.

Cook, J. K. A. (2000). "Avian pneumovirus infections of turkeys and chickens." J Vet **160**: 118-125.

Cormack, B. P., N. Ghori and S. Falkow (1999). "An adhesin of the yeast pathogen *Candida glabrata* mediating adherence to human epithelial cells." Science **285**: 578-582.

Corpet, F., J. Gouzy and D. Kahn (1999). "Recent improvements of the ProDom database of protein domain families." Nucleic Acids Res. **27**: 263-267.

Darwin, A. J. and V. L. Miller (1999). "Identification of *Yersinia enterocolitica* genes affecting survival in an animal host using signature-tagged transposon mutagenesis." Mo. Microbiol **32**: 51-62.

Delcher, A. L., D. Harmon, S. Kasif, O. White and S. L. Salzberg (1999). "Improved microbial gene identification with GLIMMER." Nucleic Acid Res. **27**: 4636-4641.

Deng, W., Y. Li, B. A. Vallance and B. B. Finlay (2001). "Locus of enterocyte effacement from *Citrobacter rodentium*: sequence analysis and evidence for horizontal transfer among attaching and effacing pathogens." Infec Immun **69**: 6323-6335.

DeVinney, R., A. Gauthier, A. Abe and B. B. Finlay (1999). "Enteropathogenic *Escherichia coli*: a pathogen that inserts its own receptor into host cells." Cell Mol Life Sci **55**: 961-976.

DeVinney, R., M. Stein, D. Reinscheid, A. Abe, S. Ruschkowski and B. B. Finlay (1999). "Enterohemorrhagic *Escherichia coli* O157:H7 produces Tir, which is translocated to the host cell membrane but is not tyrosine phosphorylated." Infect Immun **67**: 2389-2398.

Dho-Moulin, M. and J. M. Fairbrother (1999). "Avian pathogenic *Esherichia coli* (APEC)." Vet Res **30**: 299-316.

Diatchenko, L., Y. F. Lau, A. P. Campbell, A. Chenchik, F. Moqadam, B. Huang, S. Lukyanov, K. Lukyanov, N. Gurskaya, E. D. Sverdlov and P. D. Siebert (1996). "Suppression subtractive hybridization: A method of generating differentially regulated or tissue-specific cDNA probes and libraries." Proc Natl Acad Sci USA **93**: 6025-6030.

Diatchenko, L., Y. F. C. Lau, A. P. Campbell, A. Chenchik, F. Moqadam, B. Huang, S. Lukyanov, K. Lukyanov, N. Gurskaya, E. D. Sverdlov and P. D. Siebert (1996). "Suppression subtracitve hybridization: A method for generating differenetially

regulated or tissue-specific cDNA probes and libraries." Proc Natl Acad Sci **93**: 6025-6030.

Donneneberg, M. S. and T. S. Whittam (2001). "Pathogenesis and evolution of virulence in enteropathogenic and enterohemorrhagic *Escherichia coli*." J Clin Invest **107**: 539-548.

Dozois, C. M., M. Dho-Moulin, A. Bree, J. M. Fairbrother, C. Desautels and R. I. I. I. Curtiss (2000). "Relationship between the Tsh autotransporter and pathogenicity of avian *Escherichia coli* and localization and analysis of the Tsh genetic region." Infect Immun **68**: 4145-54.

Dozois, C. M., J. M. Fairbrother, J. Harel and M. Bosse (1992). "*pap*-and *pil*-related DNA sequences and other virulence determinants associated with *Escherichia coli* isolated from septicemic chickens and turkeys." Infect Immun **60**: 2648-2656.

Dubreuil, J. D. (1997). "*Escherichia coli* STb enterotoxin." Microbiol **143**: 1783-1795.

Duguid, J. R. and M. C. Dinauer (1989). "Library subtraction of *in vitro* cDNA libraries to idenitfy differenetially expressed genes in scrapie infection." Nucleic Acids Res. **18**: 2789-2792.

Edelstein, P. H., M. A. C. Edelstein, F. Higa and S. Falkow (1999). "Discovery of virulence genes of *Legionella pneumophila* by using signature tagged mutagenesis in a guinae pig pneumonia model." Proc Natl Acad Sci USA **96**: 8190-8195.

Elliott, S. J., L. A. Wainwright, T. K. McDaniel, K. G. Jarvis, Y. K. Deng, L. C. Lai, B. P. McNamara, M. S. Donnenberg and J. B. Kaper (1998). "The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69." Mol. Microbiol **28**: 1-4.

Emmerth, M., W. Goebel, S. I. Miller and C. J. Hueck (1999). "Genomic subtraction idenifies *Salmonella typhimurium* prophages, F-related plasmid sequences, and a novel fimbiral operon, *stf*, which are absent in *Salmonella typhi*." J. Bacteriol. **181**: 5652-5661.

Falkow, S. (1988). "Molecular Koch's postulates applied to microbial pathogenicity." Rev Infect Dis **10 Suppl 2**: S274-S276.

Feng, P., K. A. Lampel, H. Karch and T. S. Whittam (1998). "Genotypic and phenotypic changes in the emergence of *Escherichia coli* O157:H7." J Infec Dis **177**: 1750-1753.

Fields, P. I., R. V. Swanson, C. G. Haidaris and F. Heffron (1986). "Mutants of *Salmonella typhimurium* that cannot survive within the macrophage are avirulent." Proc Natl Acad Sci U S A **83**: 5189-5193.

Foulongne, V., G. Bourg, C. Cazevieille, S. Michaux-Charachon and D. O'Callaghan (2000). "Identification of *Brucella suis* genes affecting intracellular survival in an *in vitro* human macrophage infection model by signature-tagged transposon mutagenesis." Infect Immun **68**: 1297-1303.

Fuller, T. E., R. J. Shea, B. J. Thacker and M. H. Mulks (1999). "Identification of in vivo induced genes in *Actinobacillus pleuropneumoniae*." Microb Pathog **27**: 311-327.

Gomis, S., L. Babiuk, D. Godson, B. Allan, P. Wilson, R. Hecher and A. Potter (2002). "Stimulation of the non-specific immune system of chickens against *Escherichia coli* infections by bacterial DNA containing CpG-ODN motifs." submitted to Infec Immun.

Gomis, S. M., A. I. Gomis, N. U. Horadagoda, T. G. Wijewardene, B. J. Allan and A. A.Potter (2000). "Studies on cellulitis and other disease syndromes caused by *Escherichia coli* in broilers in Sri Lanka." Trop Anim Health Prod **32**: 341-51.

Gomis, S. M., T. Watts, C. Riddell, A. A. Potter and B. J. Allan (1997). "Experimental reproduction of *Escherichia coli* cellulitis and septicaemia in broiler chickens." Avian Dis **41**: 234-240.

Gonzalez, M. D., C. A. Lichtensteiger and E. R. Vimr (2001). "Adaptation of signature-tagged mutagenesis to *Escherichia coli* K1 and the infant-rat model of invasive disease." FEMS Microbiol Lett **198**: 125-128.

Griffiths, E. (1994). Iron acquisition systems in *Escherichia coli*. *Escherchia coli* in Domestic Animals and Humans. C. L. Gyles. Wallingford UK, CAB international: 533-566.

Gross, W. B. (1991). Colibacillosis. Diseases of Poultry. B. w. Calnek, H. J. Barnes, C. W. Beard, W. M. Reid and H. W. Yoder. Ames, Iowa, U.S.A., Iowa State University Press: 138-144.

Gross, W. G. (1994). Diseases due to *Escherichia coli* in poultry. *Escherchia coli* in Domestic Animals and Humans. C. L. Gyles. Wallingford UK, CAB international: 237-259.

Guy, J. S., L. G. Smith, J. J. Breslin, J. P. Vaillancourt and H. J. Barnes (2000). "High mortality and growth depression experimentally produced in young turkeys by dual infection with enteropathogenic *Escherichia coli* and turkey coronavirus." Avian Dis **44**: 105-113.

Guzzo, A. and M. S. DuBow (1994). "Identification and characterization of genetically programmed responses to toxic metal exposure in *Escherichia coli*." FEMS Microbiol Rev **14**: 369-374.

Gyles, C. L. (1994). *Escherichia coli* Enterotoxins. *Escherichia coli* in Domestic Animals and Humans. C. L. Gyles. Wallingford UK, CAB International: 337-363.

Gyles, C. L. (1994). *Escherichia coli* verotoxin and other cytotoxins. *Escherichia coli* in Domestic Animals and Humans. C. L. Gyles. Wallingford UK, CAB International: 365-398.

Hacker, J. and E. Carniel (2001). "Ecological fitness, genomic islands and bacterial pathogenicity." EMBO Rep **2**: 376-381.

Hara, E., T. Kato, S. Nakada, S. Sekiya and K. Oda (1991). "Subtractive cDNA cloning using oligo(dT)30-latex and PCR: isolantion of cDNA clones specific to undifferentiated human embryonal carcinoma cells." Nucleic Acids Res. **19**: 7097-7104.

Hayashi, T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C. G. Han, E. Ohtsubo, K. Nakayama, T. Murata, M. Tanaka, T. Tobe, T. Iida, H. Takami, T. Honda, C. Sasakawa, N. Ogasawara, T. Yasunaga, S. Kuhara, T. Shiba , M. Hattori and H. Shinagawa (2001). "Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12." DNA Res **8**: 11-22.

Hebrok, M., S. K. Kim and D. A. Melton (1999). "Screening for novel pancreatic genes expressed during embryogenesis." Diabetes **48**: 1550-1556.

Hedrick, S. M., D. I. Cohen, E. A. Nielsen and M. M. Davis (1984). "Isolation of cDNA clones encoding T cell-specific membrane-associated proteins." Nature **308**: 149-153.

Henikoff, J. G., S. Henikoff and S. Pietrokovski (1999). "New features of the Blocks database servers." Nucleic Acids Res. **27**: 226-228.

Hensel, M., J. E. Shea, C. Gleeson, M. D. Jones, E. Dalton and D. W. Holden (1995). "Simultaneous identification of bacterial virulence genes by negative selection." Science **269**: 400-403.

Herzer, P. J., S. Inouye, M. Inouye and T. S. Whittam (1990). "Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Esherichia coli*." J Bacteriol **172**: 6175-6181.

Hofmann, K., P. Bucher, L. Falquet and A. Bairoch (1999). "The PROSITE database, its status in 1999." Nucleic Acids Res. **27**: 215-219.

Holden, D. W., J. W. Kronstad and S. A. Leong (1989). "Mutation in a heat-regulated hsp70 gene of *Ustilago maydis*." EMBO **8**: 1927-1934.

Holt, J. G., N. R. Krieg, P. H. A. Sneath, J. T. Staley and S. T. Williams, Eds. (1994). Bergey's Manual of Determinative Baceriology. Baltimore, Maryland, USA, Williams and Wilkins.

Janke, B., U. Dobrindt, J. Hacker and G. Blum-Oehler (2001). "A subtractive hybridization analysis of genomic differences between the uropathogenic *E. coli* strain 536 and the *E. coli* K-12 strain MG1655." FEMS Microbiol Lett **199**: 61-66.

Jiang, S. Q., G. Q. Yu, Z. G. Li and J. S. Hong (1988). "Genetic evidence for modulation of the activator by two regulatory proteins involved in the exogenous induction of phosphoglycerate transport in *Salmonella typhimurium*." J Bacteriol **170**: 4304-4308.

Johnson, J. R., P. Delavari, M. Kuskowski and A. L. Stell (2001). "Phylogenetic distribution of extraintestinal virulence-associated traits in *Escherichia coli*." J Infect Dis **183**: 78-88.

Joya, J. E., T. Tsuji, A. V. Jacalne, M. Arita, T. Tsukamoto, T. Honda and T. Miwatani (1990). "Demonstration of enterotoxigenic *Escherichia coli* in diarrheic broiler chicks." Eur J Epidemiol **6**: 88-90.

Kao, J. S., D. M. Stucker, J. W. Warren and H. L. Mobley (1997). "Pathogenicity island sequences of pyelonephritogenic *Escherichia coli* CFT073 are associated with virulent. uropathogenic strains." Infec. Immun. **65**: 2812-2820.

Karlyshev, A. V., P. C. Oyston, K. Williams, G. C. Clark, R. W. Titball, E. A. Winzeler and B. W. Wren (2001). "Application of high-density array-based signature-tagged mutagenesis to discover novel *Yersinia* virulence-associated genes." Infect Immun **69**: 7810-7819.

Karp, P. D., M. Riley, M. Saier, I. T. Paulsen, S. M. Paley and A. Pellegrini-Toole (2000). "The EcoCyc and MetaCyc databases." Nucleic Acids Res **28**: 56-9.

Kauffmann, F. (1944). "Zur serologie der Coli-Gruppe." Acta Pathologicaet Microbiologica Scandinavica **21**: 20-45.

Kenny, B. (2002). "Mechanism of action of EPEC type III effector molecules." Int J Med Microbiol **291**: 469-477.

Kenny, B., R. DeVinney, M. Stein, D. J. Reinscheid, E. A. Frey and B. B. Finlay (1997). "Enteropathogenic *E. coli* (EPEC) transfers its receptor for intimate adherence into mammalian cells." Cell **91**: 511-520.

Kim, J., J. Nietfeldt, J. Ju, J. Wise, N. Fegan, P. Desmarchelier and A. K. Benson (2001). "Ancestral divergence, genome diversification, and phylogeographic variation in subpopulations of sorbitol-negative, beta-glucuronidase-negative enterohemorrhagic *Escherichia coli* O157." J Bacteriol **183**: 6885-6897.

Kleinbaum, D. G., L. L. Kupper and K. E. Muller (1988). Applied Regression Analysis and Other Multivariable Methods. Belmont, CA, USA, PWS-KENT Publishing Company.

Kleyn, P. W., W. Fan, S. G. Kovats, J. J. Lee, J. C. Pulido, Y. Wu, L. R. Berkemeier, D. J. Misumi, L. Holmgren, O. Charlat, E. A. Woolf, O. Tayber, T. Brody, P. Shu, F. Hawkins, B. Kennedy, L. Baldini, C. Ebeling, G. D. Alperin, J. Deeds, N. D. Lakey, J. Culpepper, H. Chen, M. A. Glucksmann-Kuis and K. Moore, *et al.* (1996). "Identification and characterization of the mouse obesity gene tubby: a member of a novel gene family." Cell **85**: 281-290.

Knoll, L. J., G. L. Furie and J. C. Boothroyd (2001). "Adaptation of signature-tagged mutagenesis for *Toxoplasma gondii*: a negative screening strategy to isolate genes that are essential in restrictive growth conditions." Mol Biochem Parasitol **116**: 11-16.

Kunkel, L. M., A. P. Monaco, W. Middlesworth, H. D. Ochs and S. A. Latt (1985). "Specific cloning of DNA fragments absent from the DNA of a male patient with an X chromosome deletion." Proc Natl Acad Sci USA **82**: 4778-47882.

La Ragione, R. M., W. A. Cooley and M. J. Woodward (2000). "The role of fimbriae and flagella in the adherence of avian strains of *Escherichia coli* O78:K80 to tissue culture cells and tracheal and gut explants." J Med Microbiol **49**: 327-38.

La Ragione, R. M., A. R. Sayers and M. J. Woodward (2000). "The role of fimbriae and flagella in the colonization, invasion and persistence of *Escherichia coli* O78:K80 in the day-old-chick model." Epidemiol Infect **124**: 351-363.

Lafont, J. P., M. Dho, H. M. D'Hauteville, A. Bree and P. J. Sansonetti (1987). "Presence and expression of aerobactin genes in virulent avian strains of *Escherichia coli*." Infec Immun **55**: 193-197.

Leboffe, M. and B. E. Pierce (1999). Exercises for the microbiology laboratory, Morton Pub Co.

Lehoux, D. E. and R. C. Levesque (2000). "Detection of genes essential in specific niches by stignature-tagged mutagenesis." Curr Opin Biotechnol **11**: 434-439.

Lehoux, D. E., F. Sanschagrin and R. C. Levesque (1999). "Defined oligonucleotide tag pools and PCR screening in signature-tagged mutagenesis of essential genes from bacteria." Biotechniques **26**: 473-480.

Li, Q. and P. R. Reeves (2000). "Genetic variation of dTDP-L-rhamnose pathway genes in S*almonella enterica*." Microbiol **146**: 2291-307.

Li, Y., E. Frey, A. M. Mackenzie and B. B.Finlay (2000). "Human response to *Escherichia coli* O157:H7 infection: antibodies to secreted virulence factors." Infec Immun **68**: 5090-5095.

Liang, P. and A. B. Pardee (1992). "Differential display of eukaryotic messenger RNA by means of polymerase chain reaction." Science **257**: 967-971.

Linde, K., G. C. Fthenakis and A. Fichtner (1998). "Bacterial live vaccines with graded level of attenuation achieved by antibiotic resistance mutations: transduction experiments on the functional unit of resistance, attenuation and further accompanying markers." Vet Microbiol **62**: 121-134.

Link, A. J., D. Phillips and G. M. Church (1997). "Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: applications to open reading frame characterization." J. Bacteriol. **179**: 6228-6237.

Lior, H. (1994). Classification of *Escherichia coli*. *Escherchia coli* in Domestic Animals and Humans. C. L. Gyles. Wallingford UK, CAB international**:** 31-72.

Lisitsyn, N., N. Lisitsyn and M. Wigler (1993). "Cloning the differences between two complex genomes." Science **259**: 946-951.

Lukomski, S., R. A. Hull and S. I. Hull (1996). "Identification of the O antigen polymerase (*rfc*) gene in *Escherichia coli* O4 by insertional mutagenesis using a nonpolar chloramphenicol resistance cassette." J Bacteriol **178**: 240-247.

Lukyanov, K. A., N. G. Gurskaya, E. A. Bogdanova and S. A. Lukyanov (1999). "Selective suppression of polymerase chain reaction." Bioorg Khim **25**: 141-147.

Mahan, M. J., J. M. Slauch and J. J. Mekalanos (1993). "Selection of bacterial virulence genes that are specifically induced in host tissues." Science **259**: 686-688.

Mahan, M. J., J. W. Tobias, J. M. Slauch, P. C. Hanna, R. J. Collier and J. J. Mekalanos (1995). "Antibiotic-baced selection for bacterial genes that are specifically induced during infection of a host." Proc Natl Acad Sci USA **92**: 669-673.

McClelland, M., K. E. Sanderson, J. Spieth, S. W. Clifton, P. Latreille, L. Courtney, S. Porwollik, J. Ali, M. Dante, F. Du, S. Hou, D. Layman, S. Leonard, C. Nguyen, K. Scott, A. Holmes, N. Grewal, E. Mulvaney, E. Ryan, H. Sun, L. Florea, W. Miller, T. Stoneking, M. Nhan, R. Waterston and R. K. Wilson (2001). "Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2." Nature **413**: 852-856.

Mei, J. M., F. Nourbakhsh, C. W. Ford and D. W. Holden (1997). "Identification of *Staphylococcus aureus* virulence genes in a murine model of bacteraemia using signature-tagged mutagenesis." Mol Microbiol **26**: 399-407.

Miyamoto, H., W. Nakai, N. Yajima, A. Fujibayashi, T. Higuchi, K. Sato and A. Matsushiro (1999). "Sequence analysis of Stx2-converting phage VT2-Sa shows a great divergence in early regulation and replication regions." DNA Res **6**: 235-40.

Murooka, Y., T. Higashiura and T. Harada (1978). "Genetic mapping of tyramine oxidase and arylsulfatase genes and their regulation in intergeneric hybrids of enteric bacteria." J Bacteriol **136**: 714-722.

Nakazato, M., R. Hanada, N. Murakami, Y. Date, M. S. Mondal, M. Kojima, H. Yoshimatsu , K. Kangawa and S. Matsukura (2000). "Central effects of neuromedin U in the regulation of energy homeostasis." Biochem. Biophys. Res. Commun. **277**: 191-194.

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

Nataro, J. P. and M. M. Levine (1994). *Escherichia coli* Diseases in humans. *Escherichia coli* in Domestic Animals and Humans. C. L. Gyles. Wallingford UK, CAB International**:** 285-333.

Nelson, R. T., J. Hua, B. Pryor and J. K. Lodge (2001). "Identification of virulence mutants of the fungal pathogen *Cryptococcus neoformans* using signature-tagged mutagenesis." Genetics **157**: 935-947.

Ngeleka, M., L. Brereton, G. Brown and J. M. Fairbrother (2002). "Pathotypes of avian *Escherichia coli* as related to tsh-, pap-, pil-, and iuc-DNA sequences, and antibiotic sensitivity of isolates from internal tissues and the cloacae of broilers." Avian Dis **46**: 143-152.

Ngeleka, M., J. K. Kwaga, D. G. White, T. S. Whittam, C. Riddell, R. Goodhope, A. A. Potter and B. Allan (1996). "*Escherichia coli* cellulitis in broiler chickens: clonal relationships among strains and analysis of virulence-associated factors of isolates from diseased birds." Infect Imm **64**: 3118-3126.

Norton, R. A. (1997). "Avian cellulitis." World's Poultry Sci J **53**: 334-349.

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

Olkowski, A. A., L. Kumor, D. Johnson, M. Bielby, M. Chirino-Trejo and H. L. Classen (1999). "Cellulitis lesions in commercial turkeys identified during processing." Vet Rec **145**: 228-229.

Otto, B. R., S. J. van Dooren, J. H. Nuijens, J. Luirink and B. Oudega (1998). "Characterization of a hemoglobin protease secreted by the pathogenic *Escherichia coli* strain EB1." J Exp Med **188**: 1091-1103.

Parker, L. L. and B. G. Hall (1990). "Characterization and nucleotide sequence of the cyrptic cel operon of *Escherichia coli* K12." Genetics **124**: 455-71.

Peighambari, S. M., D. B. Hunter, P. E. Shewen and C. L. Gyles (2002). "Safety, immunogenicity, and efficacy of two *Escherichia coli cya crp* mutants as vaccines for broilers." Avian Dis **46**: 287-297.

Peighambari, S. M., J.-P. Vaillancourt, R. A. Wilson and C. L. Gyles (1995). "Characteristics of *Escherichia coli* isolates from avian cellulitis." Avian Dis **39**: 116-124.

Perna, N. T., G. F. Mayhew, G. Posfai, S. Elliott, M. S. Donnenberg, J. B. Kaper and F. R. Blattner (1998). "Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7." Infec Immun **66**: 3810-3807.

Perna, N. T., G. I. I. I. Plunkett , V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Posfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamousis, J. Apodaca, T. S. Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A.

Welch and F. R. Blattner (2001). "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7." Nature **409**: 529-533.

Pfaff-McDonough, S. J., S. M. Horne, C. W. Giddings, J. O. Ebert, C. Doetkott, M. H. Smith and L. K. Nolan (2000). "Complement resistance-related traits among *Escherichia coli* isolates from apparently healthy birds and birds with colibacillosis." Avian Dis **44**: 23-33.

Polissi, A., A. Pontiggia, G. Feger, M. Altieri, H. Mottl, L. Ferrari and D. Simon (1998). "Large-scale identification of virulence genes from *Streptococcus pneumoniae*." Infec Immun **66**: 5620-5629.

Pontarollo, R. A., C. R. Rioux and A. A. Potter (1997). "Cloning and characterization of bacteriaphage-like DNA from *Haemophilus somnus* homologous to phages P2 and HP1." J. Bacteriol **179**: 1872-1879.

Prestridge, D. S. (1991). "SIGNAL SCAN: A computer program that scans DNA sequences for eukaryotic transcriptional elements." CABIOS **7**: 203-206.

Randall, C. J., P. A. Meakins, M. P. Harris and D. J. Watt (1984). "A new skin disease in broilers?" Vet Rec **114**: 246.

Ratledge, C. and L. G. Dover (2000). "Iron metabolism in pathogenic bacteria." Annu Rev Microbiol **54**: 881-941.

Reckseidler, S. L., D. DeShazer, P. A. Sokol and D. E. Woods (2001). "Detection of bacterial virulence genes by subtractive hybridization: identification of capsular polysaccharide of *Burkholderia pseudomallei* as a major virulence determinant." Infec. Immun. **69**: 34-44.

Reese, M. G. and F. H. Eeckman (1995). Novel Neural Network Algorithms for Improved eukaryotic promoter site recognition. Genome Science and Technology, Vol. 1 No. 1: p.45, the seventh international genome sequencing and analysis conference, Hyatt Regency, Hilton Head Island, South Carolina, USA.

Reeves, P. R., M. Hobbs, M. A. Valvano, M. Skurnik, C. Whitfield, D. Coplin, N. Kido, J. Klena, D. Maskell, C. R. Raetz and P. D. Rick (1996). "Bacterial polysaccharide synthesis and gene nomenclature." Trends Microbiol **4**(12): 495-503.

Rogers, H. J. (1973). "Iron-binding catechol and virulence in *Escherichia coli*." infect immun **7**: 445-456.

Salyers, A. A. and D. D. Whitt (1994). Bacterial pathogenesis : a molecular approach. Washington, DC, USA, ASM Press, pg 30-46.

Salzberg, S. L., A. L. Delcher, S. Kasif and O. White (1998). "Microbial gene identification using interpolated Markov models." Nucleic Acids Res. **27**: 544-548.

Sambrook, J., E. F. Fritsch and T. Maniatis (1989). Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Press.

Sauer, F. G., M. A. Mulvey, J. D. Schilling, J. J. Martinez and S. J. Hultgren (2000). "Bacterial pili: molecular mechanisms of pathogenesis." Curr Opin Microbiol **3**: 65-75.

Schandel, K. A., M. M. Muller and R. E. Webster (1992). "Localization of TraC, a protein involved in assembly of the F conjugative pilus." J. Bacteriol. **174**: 3800-3806.

Schatz, D. G. and M. Hubank (1994). "Identifying differences in mRNA expression by representational difference analysis of cDNA." Nucleic Acids Res **22**: 5640-5648.

Seidman, C. E. (1994). Introduction of plasmid DNA into cells. Current Protocols in Molecular Biology. F. M. Ausubel. Boston, USA, Green Publishing Associates Inc., John Wiley and Sons Inc. **1:** 1.8.1-1.8.3.

Slauch, J. M. and A. Camilli (2000). "IVET and RIVET: use of gene fusions to identify bacterial virulence factors specifically induced in host tissues." Methods Enzymol **326**: 73-96.

Slauch, J. M., M. J. Mahan and J. J. Mekalanos (1994). "*In vivo* expression technology for selection of bacterial genes specifically induced in host tissues." Methods Enzymol **235**: 481-492.

Sompayrac, L., S. Jane, T. C. Burn, D. G. Tenen and K. J. Danna (1995). "Overcoming limitations of the mRNA differential display technique." Nucleic Acids Res **23**: 4738-4739.

Sramek, S. J. and F. E. Frerman (1975). "Purification and properties of *Escherichia coli* coenzyme A-transferase." Arch Biochem Biophys **171**: 14-26.

Stordeur, P., D. Marlier, J. Blanco, E. Oswald, F. Biet, M. Dho-Moulin and J. Mainil (2002). "Examination of *Escherichia coli* from poultry for selected adhesin genes important in disease caused by mammalian pathogenic *E. coli*." Vet Microbiol **84**: 231-241.

Sturm, S., B. Jann, P. Fortnagel and K. N. Timmis (1986). "Genetic and biochemical analysis of *Shigella dysenteriae* 1 O antigen polysaccharide biosynthesis in *Escherichia coli* K-12: structure and functions of the rfb gene cluster." Microb Pathog **1**: 307-324.

Sussman, M., Ed. (1997). *Escherichia coli* Mechanisms of Virulence. New York, NY, U.S.A., Cambridge University Press.

Tabor, S. and K. Struhl (1994). DNA-Dependent DNA Polymerases. Current Protocols in Molecular Biology. F. M. Ausubel. Boston, USA, Green Publishing Associates Inc., John Wiley and Sons Inc. **1:** 3.5.9.

Takeuchi, T., H. Nishimatsu, T. Ueki, T. Kajiwara, H. Fukuhara, T. Ishida, N. Moriyama and T. Kitamura (2000). "Differentially expressed mRNAs in androgen-independent but not androgen-dependent Shionogi carcinoma." Urol Res **28**: 82-85.

Tatlow, D., R. Brownlie, L. A. Babiuk and P. Griebel (2000). "Differential display analysis of gene expression during the induction of mucosal immunity." Immunogenetics **52**: 73-80.

Tinsley, C. R. and X. Nassif (1996). "Analysis of the genetic differences between *Neisseria meningitidis* and *Neisseria gonorrhoeae*: two closely related bacteria expressing two different pathogenicities." Proc. Natl. Acad. Sci USA **93**: 11109-11114.

Torres, A. G. and S. M. Payne (1997). "Haem iron-transport system in enterohaemorrhagic *Escherichia coli* O157:H7." Mol Microbiol **23**: 825-833.

Torres, A. G., P. Redford, R. A. Welch and S. M. Payne (2001). "TonB-dependent systems of uropathogenic *Escherishia coli*: aerobactin and heme transport and TonB are required for virulence in the mouse." Infec Immun **69**: 6179-6185.

Tseng, C. C., J. J. Wu, H. L. Liu, J. M. Sung and J. J. Huang (2002). "Roles of host and bacterial virulence factors in the development of upper urinary tract infection caused by *Escherichia coli*." Am J Kidney Dis **39**: 744-752.

Tsolis, R. M., S. M. Townsend, E. A. Miao, S. I. Miller, T. A. Ficht , L. G. Adams and A. J. Baumler (1999). "Idenification of a putative *Salmonella enterica* serotype Typhimurium host range factor with homology to IpaH and YopM by signature-tagged mutagenesis." Infect Immun **67**: 6385-6393.

Valdivia, R. H. and S. Falkow (1996). "Bacterial genetics by flow cytometry: rapid isolation of *Salmonella typhimurium* acid-induced promoters by differential fluorescence induction." Mol Microbiol **22**: 367-378.

Valdivia, R. H. and S. Falkow (1997). "Fluorescence-based isolateion of bacterial genes expressed within host cells." Science **277**: 2007-2011.

Valdivia, R. H., A. E. Hromockyj, D. Monack, L. Ramakrishnan and S. Falkow (1996). "Applications for green fluorescent protein (GFP) in the study of host-pathogen interactions." Gene **173**: 47-52.

Venkatesan, M. M., M. B. Goldberg, D. J. Rose, E. J. Grotbeck, V. Burland and F. R. Blattner (2001). "Complete DNA sequence and analysis of the large virulence plasmid of *Shigella flexneri*." Infect Immun **69**: 3271-85.

Wang, L., W. Qu and P. R. Reeves (2001). "Sequence analysis of four *Shigella boydii* O-antigen loci: implication for *Escherichia coli* and *Shigella* relationships." Infect Immun **69**: 6923-6930.

Welsh, J., K. Chada, S. S. Dalal, R. Cheng, D. Ralph and M. McClelland (1992). "Arbitrarily primed PCR fingerprinting of RNA." Nucleic Acids Res **20**: 4965-4970.

Wilks, A. (2001). "The ShuS protein of *Shigella dysentariae* is a heme-sequestering protein that also binds DNA." Arch. Biochem. Biophys. **387**: 137-142.

Wilson, K. (1994). Preparation of genomic DNA from bacteria. Current Protocols in Molecular Biology. F. M. Ausubel. Boston, USA, Green Publishing Associates Inc., John Wiley and Sons Inc. **1:** 2.4.1-2.4.2.

Wooley, R. E., L. K. Nolan, J. Brown, P. S. Gibbs, C. W. Giddings and K. S. Turner (1993). "Association of K-1 capsule, smooth lipopolysaccharides, traT gene, and colicin V production with complement resistance and virulence of avian *Escherichia coli*." Avian Dis **37**: 1092-1096.

Wyckoff, E. E., D. Duncan, A. G. Torres, M. Mills, K. Maase and S. M. Payne (1998). "Structure of the *Shigella dysenteriae* haem transport locus and its phylogenetic distribution in enteric bacteria." Mol. Microbiol. **28**: 1139-1152.

Yogaratnam, V. (1995). "Analysis of the causes of high rates of carcase rejection at a poultry processing plant." Vet Rec **137**: 215-217.

Zhang, Y. L., C. T. Ong and K. Y. Leung (2000). "Molecular analysis of genetic differences between virulent and avirulent strains of *Aeromonas hydrophila* isolated from diseased fish." Micro. **146**: 999-1009.

Zhou, C., Y. Yang and A. Y. Jong (1990). "Mini-prep in ten minutes." Biotechniques **8**: 172-3.

Zhu, C., T. S. Agin, S. J. Elliott, L. A. Johnson, T. E. Thate, J. B. Kaper and E. C. Boedeker (2001). "Complete nucleotide sequence and analysis of the locus of enterocyte effacement from rabbit diarrheagenic *Escherichia coli* RDEC-1." Infect Immun **69**: 2107-2115.

A. APPENDIX

A1: The Strains and Plasmids used.

Strain	Phenotype/Genotype	Source/Reference
CC118 λpir	Δ[*ara-leu*], *araD*, Δ*lacX*74,*galE*, *galK*, phoA20, *thi*-1, *rpsE*, *rpoB*, *argE* (Am^R), *recA*1, λpir phage lysogen	(Hensel, Shea *et al.* 1995)
DH5α	*supE*44 Δ*lacU*169 f80 *lacZDM*15 *hsdR*17 *recA*1 '*endA*1 *gyrA*96 *thi*-1 *relA*1	N/A
EC106	serogroup O1, Liver isolate from turkey colibacillosis	(Ngeleka, Kwaga *et al.* 1996)
EC317	ATCC 55346, Avian strain non-motile O2:HNM, Liver isolate from turkey colisepticaemia	C. Riddell, Western College of Veterinary Medicine
EC1347	Spontaneous rifampicin resistant mutant of EC317	This study
EC2029	Serotype O2, cellulitis isolate from Sri Lanka	(Gomis, Gomis *et al.* 2000)
EC2503	HB101 Tn5::ent$^-$	S. Payne University of Texas (Wyckoff, Duncan *et al.* 1998)
EC2524	EC317 with p2516	This study
EC2525	EC317 with p2517	This study
EC2526	EC317 with p2518	This study
EC2527	EC317 with p2519	This study
EC2547	EC317 with ΩKn in sequence for fragment 28	This study
EC2548	EC317 with ΩKn in sequence for fragment 28	This study
EC2549	EC317 with ΩKn in sequence for fragment 3-54	This study
EDL 933	O157:H7, EHEC, STX1, STX2	Brett Finlay, University of British Columbia
HB101	F- *hsd*S20 (r_B^-, m_B^-) *thi*-1 *sup*E44 *ara*14 *gal*K2 *lac*Y1 *pro*A2 *rps*L20 (Str^R) *xyl*-5 *mtl*-1 *rec*A13 *mcr*B *leu*B6	N/A

Strain	Phenotype/Genotype	Source/Reference
MG1655	*E. coli* K-12, F^-, lambda$^-$, *ilvG*$^-$, *rfb*-50 *rph*$^-$1, OR:H48:K$^-$	*E. coli* genetic stock center
pRAP117 in JM105	Hmb+ pHC79 derivative with a 30 kb Sau3A1 fragment of *Haemophilus somnus* strain HS25 genomic DNA	(Pontarollo, Rioux *et al.* 1997)
S17 λpir	*tpr*, Sm^R, *recA*, *thi*, *pro*, *hsdR*-, M+, RP4:2-Tc:Mu:Km:Tn7, λpir	(Hensel, Shea *et al.* 1995)
TOP 10F'	F′ {*lac*I^q *Tn10*(Tet^R)} *mcr*A .Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*Z ΔM15 .*lac*X74 *deo*R *rec*A1, *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G	BD Biosciences Clontech Palo Alto, CA, USA

Plasmid	Description	Reference or source
pT-Adv	Positive selection cloning vector; ColE1 *ori*, f1 *ori*, Km^R, Am^R	BD Biosciences Clontech Palo Alto, CA, USA
pHTL106	*Bgl*II/*Eco*RI fragment containing the entire *shu* gene cluster cloned into the *BamH*I/*Eco*RI sites of pWKS30	(Wyckoff, Duncan *et al.* 1998)
pIB307	*rep*A (pSC101-ts) *cat* (Cm^R) *bla* (Amp^R) M13ori	(Blomfield, Vaughn *et al.* 1991)
pKO3	*rep*A (pSC101-ts) *cat* (Cm^R) *sac*B *bla* (Amp^R) M13ori	(Link, Phillips *et al.* 1997)
pUC4Ωkm2	Am^R, ΩKm,	Jose Perez-Casal, VIDO
pUC18	Am^R	Gibco/BRL
pUTmini-Tn5Km2	Am^R, Km^R	(Hensel, Shea *et al.* 1995)
pHP45ΩTc	Tc^R, Am^R	Jose Perez-Casal, VIDO
pEC1605	pT-Adv with subtractive hybridization fragment 1-4	This study
pEC1606	pT-Adv with subtractive hybridization fragment 1-44	This study
pEC1608	pT-Adv with subtractive hybridization fragment 1-47	This study
pEC1609	pT-Adv with subtractive hybridization fragment 1-63	This study
pEC1611	pT-Adv with subtractive hybridization fragment 1-74	This study
pEC1616	pT-Adv with subtractive hybridization fragment2-40	This study
pEC1619	pT-Adv with subtractive hybridization fragment 2-76	This study
pEC1627	pT-Adv with subtractive hybridization fragment 3-19	This study

Plasmid	Description	Reference or source
pEC1629	pT-Adv with subtractive hybridization fragment 3-54	This study
pEC1634	pT-Adv with subtractive hybridization fragment 3-72	This study
pEC1924	pGex-6p-1 with *esp*A	(Li, Frey *et al.* 2000)
pEC1925	pGex-6p-1 with *esp*B	(Li, Frey *et al.* 2000)
pEC1926	pGex-6p-1 with *esp*D	(Li, Frey *et al.* 2000)
pEC1927	pGex-6p-1 with *tir*	(Li, Frey *et al.* 2000)
pEC1928	pGex-6p-1 with truncated intimin (*eae*)	(Li, Frey *et al.* 2000)
pEC2067	pT-Adv with subtractive hybridization fragment 28	This study
pEC2074	pT-Adv with subtractive hybridization fragment 201	This study
pEC2077	pT-Adv with subtractive hybridization fragment 219	This study
pEC2080	pT-Adv with subtractive hybridization fragment 240	This study
pEC2082	pT-Adv with subtractive hybridization fragment 282	This study
pEC2084	pT-Adv with subtractive hybridization fragment 296	This study
pEC2086	pT-Adv with subtractive hybridization fragment 305	This study
pEC2088	pT-Adv with subtractive hybridization fragment 308	This study
pEC2090	pT-Adv with subtractive hybridization fragment 315	This study
pEC2096	pT-Adv with subtractive hybridization fragment 347	This study
pEC2100	pT-Adv with subtractive hybridization fragment 358	This study
pEC2101	pT-Adv with subtractive hybridization fragment 361	This study
pEC2107	pT-Adv with subtractive hybridization fragment 400	This study
pEC2110	pT-Adv with subtractive hybridization fragment 453	This study
pEC2115	pT-Adv with subtractive hybridization fragment 235	This study
pEC2158	sCos-1 containing *E. coli* EC317 genomic DNA and fragment 453	This study
pEC2165	sCos-1 containing *E. coli* EC317 genomic DNA and fragment 453	This study
pEC2181	sCos-1 containing *E. coli* EC317 genomic DNA and thought to contain fragment 1-74	This study

Plasmid	Description	Reference or source
pEC2186	sCos-1 containing *E. coli* EC317 genomic DNA known not to contain fragment 453	This study
pEC2196	SuperCos 1 Am^R Km^R	Stratagene
pEC2483	sCos-1 containing *E. coli* EC317 genomic DNA contains fragment 296	This study
pEC2491	sCos-1 containing *E. coli* EC317 genomic DNA contains fragment 1-63	This study
pEC2487	pT-Adv with subtractive hybridization fragment 381	This study
pEC2512	p2067 with ΩTc *Sma*I fragment in the *Bsg*I site	This study
pEC2513	p2096 with ΩTc *Sma*I fragment in the *SexA*I site	This study
pEC2514	p1606 with ΩTc *Sma*I fragment in the *Sac*II site	This study
pEC2515	p1608 with ΩTc *Sma*I fragment in the *Hinc*II site	This study
pEC2516	*EcoR*I fragment from p2512 in *Sma*I site of pKO3	This study
pEC2517	*EcoR*I fragment from p2513 in *Sma*I site of pKO3	This study
pEC2518	*EcoR*I fragment from p2515 in *Sma*I site of pKO3	This study
pEC2519	p1629 with ΩTc *Sma*I fragment in the *BstB*I site	This study
pEC2528	*EcoR*I fragment from p2110 in *EcoR*I site of pUC18	This study
pEC2529	*EcoR*I fragment from p2067 in *EcoR*I site of pUC18	This study
pEC2530	*EcoR*I fragment from p2096 in *EcoR*I site of pUC18	This study
pEC2531	*EcoR*I fragment from p2100 in *EcoR*I site of pUC18	This study
pEC2532	*EcoR*I fragment from p2487 in *EcoR*I site of pUC18	This study
pEC2533	*EcoR*I fragment from p1606 in *EcoR*I site of pUC18	This study
pEC2534	*EcoR*I fragment from p1608 in *EcoR*I site of pUC18	This study
pEC2535	*EcoR*I fragment from p1609 in *EcoR*I site of pUC18	This study
pEC2536	*EcoR*I fragment from p1629 in *EcoR*I site of pUC18	This study
pEC2537	p2529 with ΩKn *Sma*I fragment in the *Bsg*I site	This study

Plasmid	Description	Reference or source
pEC2538	p2533 with ΩKn *Sma*I fragment in the *Sac*II site	This study
pEC2539	p2536 with ΩKn *Sma*I fragment in the *BstB*I site	This study
pEC2540	p2528 with ΩKn *Sma*I fragment in the *BsaA*I site	This study
pEC2541	pIB307 partial digest with *EcoR*I with *EcoR*I fragment from p2537 in the site, temperature sensitive, Cm, Amp, Km	This study
pEC2544	pIB307 partial digest with *EcoR*I with *EcoR*I fragment from p2538 in the site, temperature sensitive, Cm, Amp, Km	This study
pEC2545	pIB307 partial digest with *EcoR*I with *EcoR*I fragment from p2538 in the site, temperature sensitive, Cm, Amp, Km	This study
pEC2546	pIB307 partial digest with *EcoR*I with *EcoR*I fragment from p2541 in the site, temperature sensitive, Cm, Amp, Km	This study
SuperCos 1 (sCos 1)	*bla* (Am^R), *npt*II (Neo^R)	Stratagene

A2: Collection of Avian *E. coli* Isolates

Strain	Description*	Reference or source
EC222	ATCC 55348, O1, Liver isolate from turkey colibacillosis	D. Ouderka
EC234	V-078, O78:K80:H9, Liver isolate from turkey colisepticemia	Larry Arp Iowa State
EC419	O20:HNM, chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC420	O(NT):H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC421	O78:HNM, chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC422	O78:H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC423	O78:H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC424	O(NT):H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC425	O25:H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC426	O25:H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC427	O25:H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC428	O25:H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC429	O32,83:H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC430	O(NT):HNM, chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC435	O(NT):HNM, chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC436	O(NT):HNM, chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC437	O(NT):HNM, chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC438	O(NT):H49, chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC439	O(NT):HNM, chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC440	O(NT):HNM, chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC441	O(NT):H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC442	O(NT):HNM, chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC443	O(NT):H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC444	O29:H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC445	O(NT):H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC447	O83:H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC448	O25:H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC449	O(NT):HNM, chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC450	O25:H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC469	O(NT):H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC470	O25:H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC471	O(NT):H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC472	O(NT):H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC473	O(NT):H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC474	O(NT):H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC475	O(NT):H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC476	O(NT):H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC477	O(NT):H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC478	O(NT):H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC479	O(NT):H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)
EC480	O(NT):H(ND), chicken cellulitis	(Ngeleka, Kwaga *et al.*, 1996)

* The designation of O(NT) indicates that the serotype of the strain could not be determined. The designation of H(ND) indicates that the flagellar antigen could not be determined. The designation of HNM indicates that the strain was non-motile.

A3: The ECOR collection of strains:

They are described by Ochman *et al.* (Ochman and Selander 1984).

Strain	Description	Strain	Description
ECOR1	Human female UTI	ECOR37	Marmoset
ECOR2	Human male	ECOR38	Human female
ECOR3	Dog	ECOR39	Human female
ECOR4	Human female	ECOR40	Human female UTI
ECOR5	Human female	ECOR41	Human male
ECOR6	Human male	ECOR42	Human male
ECOR7	Orangutan	ECOR43	Human female
ECOR8	Human female	ECOR44	Cougar
ECOR9	Human female	ECOR45	Pig
ECOR10	Human female	ECOR46	Celebese ape
ECOR11	Human female UTI	ECOR47	Sheep
ECOR12	Human female	ECOR48	Human female UTI
ECOR13	Human female	ECOR49	Human female
ECOR14	Human female	ECOR50	Human female
ECOR15	Human female UTI	ECOR51	Human infant
ECOR16	Leopard	ECOR52	Orangutan
ECOR 17	Pig	ECOR53	Human female
ECOR18	Celebese ape	ECOR54	Human
ECOR19	Celebese ape	ECOR55	Human female
ECOR20	Steer	ECOR56	Human female, UTI
ECOR21	Steer	ECOR57	Gorilla
ECOR22	Steer	ECOR58	Lion
ECOR23	Elephant	ECOR59	Human male
ECOR24	Human female	ECOR60	Human female UTI
ECOR25	Dog	ECOR61	Human female
ECOR26	Human infant	ECOR62	Human female UTI
ECOR27	Giraffe	ECOR63	Human female
ECOR28	Human female	ECOR64	Human female UTI
ECOR29	Kangaroo rat	ECOR65	Celebese ape
ECOR30	Bison	ECOR66	Celebese ape
ECOR31	Leopard	ECOR67	Goat
ECOR32	Giraffe	ECOR68	Giraffe
ECOR33	Sheep	ECOR69	Celebese ape
ECOR34	Dog	ECOR70	Gorilla
ECOR35	Human male	ECOR71	Human female UTI
ECOR36	Human female	ECOR72	Human female

A4: Description of Scores for Day-Old Chicks

Score	description
0	Normal, active when stimulated, may go back to sleep soon after (especially in 1 and 2 day old chicks). Scratch and forage for food. Preen feathers. Drink from water bowl. Move away from an object placed in sight path. Good balance, may flap wings. Peck at each other.
0.5	Bird does not appear normal. May be slow to move or peck at feed. Something just does not appear to be right. These birds are examined closely at next evaluation time.
1	Go back to sleep within a few seconds after stimulation. Do not forage for food readily, may take a drink and peck some. Depressed, head hanging down. Still aware of environment when stimulated and will move away from object in sight path. Balance may not be perfect when walking. Hesitate to move. Tire quickly.
2	Unable to stand, or remain standing. May get up momentarily but sit down again quickly. May sit back on hocks with neck extended and mouth breath. Difficulty holding head up. Unable to forage for food effectively, may make a few attempts at pecking. Balance is poor, may stumble. Wings often extended. Do not move away from object placed in sight path.
3	Bird is found dead in pen.

A5: Description of Scores for Chickens in Both Cellulitis Models of Infection

SCORE	DESCRIPTION
0	Normal. Birds active when you enter the room. Walk around, forage for food. Move away quickly when approached. Preen feathers. Drink from water bowl. Move away from a person when in the room. Good balance and curious.
0.5	Bird does not appear normal. May be slow to move or peck at feed. May have mild diarrhea. Something just does not appear to be right. These birds are examined closely at next evaluation time.
1	Plumage is unpreened and has a ruffled appearance especially around the head. Bird is usually sitting and reluctant to stand. Bird hesitates to move, when it does move it may be lame in one of it's legs. Do not forage for food readily, may take a drink and peck some. May have moderate diarrhea. Depressed but still responsive to environment. These birds may be euthanized at the end of afternoon observation.
2	Unable to stand, or remain standing. Head is usually extended, with bird mouth breathing. Bird is usually very lame and will vocalize when is has to move. May have severe diarrhea. Very depressed and non responsive to environment. These birds are euthanized.
3	Bird is found dead.

A6: Accession Numbers for Subtractive Hybridization Fragments and Cosmid Sequences.

Fragment or Cosmid	accession number	Fragment or Cosmid	accession number
17	BH854384	341	BH854417
28	BH854385	347	BH854418
50	BH854386	350	BH854419
53	BH854387	356	BH854420
101	BH854388	358	BH854421
103	BH854389	361	BH854422
152	BH854390	366	BH854423
199	BH854391	373	BH854424
201	BH854392	375	BH854425
202	BH854393	381	BH854426
214	BH854394	390	BH854427
219	BH854395	400	BH854428
230	BH854396	401	BH854429
233	BH854397	410	BH854430
235	BH854398	453	BH854431
238	BH854399	1-4	BH854432
240	BH854400	1-44	BH854433
243	BH854401	1-47	BH854434
254	BH854402	1-63	BH854435
266	BH854403	1-74	BH854436
268	BH854404	2-42	BH854437
282	BH854405	2-69	BH854438
289	BH854406	2-76	BH854439
296	BH854407	3-8	BH854440
303	BH854408	3-12	BH854441
305	BH854409	3-13	BH854442
307	BH854410	3-19	BH854443
308	BH854411	3-54	BH854444
314	BH854412	3-72	BH854445
315	BH854413	Cosmids	
317	BH854414	2165	AF524932
338	BH854415	2483	AF524931
340	BH854416	2491	AF524930