This is to certify that the
dissertation entitled

EVOLUTION OF INVASIVENESS IN *ESCHERICHIA COLI* AND *SHIGELLA*

presented by

Alyssa Courtney Bumbaugh

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Genetics

Thomas L. Whitham
Major Professor's Signature

Aug. 4, 2003
Date

EVOLUTION OF INVASIVENESS IN *ESCHERICHIA COLI* AND *SHIGELLA*

By

Alyssa Courtney Bumbaugh

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Graduate Program in Genetics

2003

ABSTRACT

EVOLUTION OF INVASIVENESS IN *ESCHERICHIA COLI* AND *SHIGELLA*

By

Alyssa Courtney Bumbaugh

Enteroinvasive *Escherichia coli* and *Shigella* are invasive enteric pathogens that are responsible for over 1.1 million cases of illness each year. These organisms have the ability to invade mucosal epithelia and cause dysentery. The acquisition of a large virulence plasmid conferring invasive ability has been a major factor influencing the evolution of these pathogens. Additionally, the spread of mobile blocks of virulence genes known as pathogenicity islands, and loss-of-functions caused by large genomic deletions (so called "black holes") have enhanced virulence. The basis of this research is to: 1) establish a phylogenetic framework for enteroinvasive *E. coli* and *Shigella* in order to test hypotheses about the evolution of virulence attributes, and the extent and timing of gene losses and acquisitions; 2) assess the within- and between-group variation in the components of fitness and virulence, including invasiveness, intracellular replication, and spread; 3) and develop and test a method for genomically screening pathogenic *E. coli* and *Shigella* for large insertions and deletions and to determine the impact of gene loss on adaptation and virulence. To address these goals, 15 housekeeping loci were sequenced in enteroinvasive *E. coli* and *Shigella* isolates in order to establish a phylogenetic framework. Additionally, all isolates were screened for six virulence loci associated with pathogenicity islands and a large virulence plasmid (pINV) to determine

patterns of gene acquisition and loss over evolutionary time. Virulence attributes, specifically, invasion, intracellular multiplication, and spread to adjacent cells, were measured in isolates representative of clonal groups in the phylogenetic framework. In order to gain a broader knowledge of genome evolution within these pathogens, two genome comparison techniques were employed. Suppression subtraction hybridization allows for the identification of strain-specific DNA by comparing two genomes. Because this technique is not informative with regard to location within the genome and only allows for the immediate comparison of two genomes, a paired-end sequencing mapping approach was developed. This new approach allows for the detection of insertions and/or deletions in the genome of an isolate with an unsequenced genome by comparison to a closely related isolate with a completely sequenced genome. This method can be employed to identify new virulence factors encoded by pathogenicity islands or black holes. The synthesis of results from each project will aid in understanding the evolution of invasive *Escherichia coli* and *Shigella* with consideration to both genotype and phenotype.

DEDICATION

This work is dedicated to my loving family and friends for their continuing support, patience, and encouragement throughout my graduate career. To my beloved grandmothers: Jennie Irene Goes and Sarah Jane Bumbaugh who passed away during my time in Michigan, I know in my heart that you are sharing in this accomplishment with me and are so very proud. I present this work to all of you with love and gratitude.

ACKNOWLEDGEMENTS

I wish to thank my mentor, Thomas S. Whittam for his invaluable guidance and support. I thank my committee members, Cindy Grove Arvidson, Michael Bagdasarian, Robert Brubaker, and Vincent Young for their suggestions and guidance. I also thank the director of Graduate Program in Genetics, Helmut Bertrand for his direction and Jeannine Lee for her secretarial support.

Many of my fellow lab members have been instrumental to the completion of this research by providing suggestions and support. I thank Cheryl Tarr, David Lacher, Teresa Large, Seth Walk, Adam Nelson, Lindsay Ouellette, Katie Hyma, Lukas Wick, and Shanda Birkeland. I also wish to thank Mahdi Saeed and his lab members, Hany Elsheikha and Paul Dabney, for providing instruction with cell culture. I thank Anne Plovanich-Jones for her friendship, guidance, and strength of spirit.

I also wish to acknowledge the many collaborators that have provided bacterial isolates and guidance with protocols: Shelley Payne and Liz Wyckoff at the University of Texas at Austin; Anthony Maurelli and Rachel Binet at the Uniformed Services University of the Health Sciences; Kaisar Talukder at the International Centre for Diarrhoeal Diseases Research, Bangladesh; and Nancy Strockbine at the Centers for Disease Control and Prevention.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

LITERATURE REVIEW

Enteric bacterial pathogens are the causative agents of gastroenteritis and enteric fevers in both humans and animal and include *Campylobacter*, *Shigella*, *Salmonella*, *Vibrio, Yersinia,* and certain pathovars of *Escherichia coli*. To be successful, enteric pathogens must be able to colonize the intestinal tract, adhere to or efface the epithelium, and deliver cytotoxins or enterotoxins in order to cause disease in the host (42). Even though these organisms can inhabit the same ecological niche, they differ and are distinguishable by virulence attributes, metabolic functions and biochemical properties.

Escherichia, *Shigella*, and *Salmonella* are closely related members of the Enterobacteriaceae; however, they each cause a clinically distinct disease. *E. coli* is typically a harmless commensal of the gut but it can cause human disease ranging from watery diarrhea to hemolytic uremic syndrome (112). These pathogens are commonly transmitted to humans via contaminated food and water with many being extremely acid tolerant allowing for survival both inside the host and in the harsh external environment. *Shigella* and *Salmonella* differ from the majority of the *E. coli* in that they are able to invade the cells of the intestinal tract. This offers a selective advantage as the pathogens can evade the host immune system and utilize the intracellular resources. The genes underlying this phenotype encode a type III secretion system and have been acquired on mobile genetic elements.

This dissertation focuses on *Shigella* and enteroinvasive *E. coli* (EIEC), enteric pathogens that have evolved the ability to invade epithelial cells and cause severe intestinal illness. For historical reasons, these bacteria are referred to as belonging to two genera, *Shigella* and *Escherichia.* The bacteria are similar Gram-negative coliforms but the *Shigella* do not ferment lactose whereas *Escherichia* do. *Shigella* species are

pathogenic bacteria that are invasive and cause bacillary dysentery, whereas within *Escherichia*, only certain serotypes of *E. coli*, the enteroinvasive strains, have the same ability. Recent molecular evidence indicates that the classification of these bacteria is artificial and it is the main objective of this research to investigate the evolution of these specialized pathogens.

There are four species of *Shigella; S. boydii, S. dysenteriae, S. flexneri,* and *S. sonnei,* that have been recognized historically because of the severity of disease and their clinical importance. The four species are identified and distinguished by biochemical traits and the expression of specific somatic antigens. Within each species there is some antigenic variation with the number of serotypes ranging from 18 of *S. boydii,* to a single type of *S. sonnei.* Molecular evidence indicates the *S. sonnei* strains belong to a single widespread clone (55).

The prevalence of the various strains in disease varies geographically and has changed historically. At the present time, *S. flexneri* 2A is the most prevalent endemic strain in developing countries, whereas *S. sonnei* infections continue to account for most shigellosis in industrialized nations. Worldwide, shigellosis is responsible for the death of more than 1.1 million people each year (56) and in the United States it causes more than 400,000 cases of illness per year (70).

Infections are transmitted by the fecal-oral route usually as a result of direct person-to-person transfer or through contact with or ingestion of contaminated food and water (24). The infectious dose is very low with ingestion of as few as 10 bacteria causing symptomatic infections (24). Shigellae colonize only humans and non-human primates so there are no alternative species of animal reservoirs.

Enteroinvasive *E. coli* strains are similar to *Shigella* and were first identified in Italy in the 1940s (29). Isolates within this pathovar of *E. coli* have been found to harbor the same virulence plasmid as the shigellae (48). EIEC have been involved in several large outbreaks of acute gastroenteritis in the United States (39, 47, 109, 121) and have also been implicated in traveler's diarrhea (128). In the developing world, EIEC infections contribute to endemic rates of diarrheal disease: enteroinvasive strains are typically isolated in 1 - 5 % of the cases of acute diarrhea in children (25, 32, 63, 92, 117, 120), although incidence rates vary with season (92) and socio-economic conditions (120).

Invasion. Shigellae and enteroinvasive *E. coli* have a characteristic form of pathogenesis involving invasion of the mucosal epithelial cells of the large intestine. The molecular and cellular events underlying epithelial cell invasion by Shigellae have been intensively studied and reviewed (35, 44, 83, 100, 102). An overview of the invasion process is shown in Figure 1. Briefly, invasion occurs via bacterium-directed phagocytosis with the major events as follows: contact of bacteria with the surface of the epithelial cell induces rearrangements of the cyto-skeleton, local membrane ruffling, and uptake of the bacteria (17). This is depicted as step 1 in Figure 1. Inside the cell, the bacteria escape from the endosomal vacuole by lysing the membrane, enter the cytoplasm (step 2), and multiply there (step 3). The intracellular bacteria move through the cytoplasm by polymerizing actin filaments (step 4). This movement results in protrusions from an infected cell's membrane that contains bacterial cells at the tip, which can then be engulfed by adjacent cells. In this way, the invasive bacteria can multiply and spread from cell-to-cell without being exposed to the extracellular environment.

The components underlying the invasive phenotype are encoded on a large (~200 kb) pINV plasmid. The pINV plasmids vary in size and composition, but in general, they include an entry region containing 35 genes organized into at least 4 transcriptional units (83). These include the secretory machinery, secreted proteins, molecular chaperones, and regulators encoded by *virB-ipgD*, *icsB-mxiE*, *mxiM-spa13*, and *spa47-spa40*. The entry region genes are homologous to the genes of the SPI-1 island of *Salmonella* (37). The pINV plasmid also carries genes for actin-based motility of shigellae inside the cell, a variety of plasmid antigens, and other suspected virulence-related proteins. Although most of the research has been conducted on *S. flexneri*, it is clear that many of the genes on pINV are critical to cell invasion and are required for full virulence of enteroinvasive strains.

Regulation of the plasmid-borne loci is temperature dependent (30). Extensive work has been done to identify the regulatory pathway that ultimately results in the expression of a type III secretion system. A chromosomal locus, *virR* (68), binds upstream of a plasmid locus, *virF* and causes a conformational change in the topology of the DNA. VirF is then responsible for the transcriptional regulation of *virB* (3), also located on the large virulence plasmid. VirB then transcribes the genes encoding the type III secretion system as well as additional effector molecules. In some instances, the plasmid can become integrated into the chromosome resulting in a reduction of *virB* transcription and ultimately, a non-invasive phenotype (14).

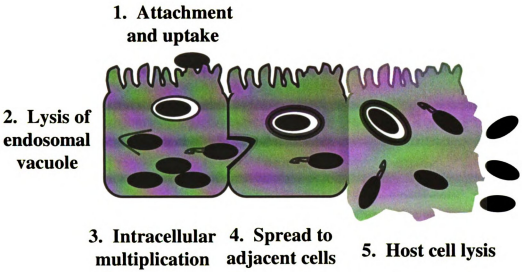

Figure 1. Bacterial mediated invasion in eukaryotic cells. The invasive process begins with bacterial mediated attachment and uptake into the host cell. The bacterium lyses the endosomal vacuole and is free to multiply in the intracellular environment. Polymerization of actin filaments allows for the bacteria to spread to adjacent cells. Ultimately, the host cell will lyse due to the compromised membrane.

Evolutionary relationships among *E. coli* and *Shigella* strains. In addition to their ability to invade epithelial cells, *Shigella* species and enteroinvasive *E. coli* strains often share other phenotypic properties: they usually do not decarboxylate lysine and, with a few exceptions, they are nonmotile (8, 108). In addition, invasive ability is associated with a limited number of serotypes. Together these observations encouraged the notion that invasive strains were evolutionarily related and represented a specialized natural group of bacteria. Application of the methods of evolutionary genetics began to elucidate this issue. Ochman et al. (81) used multilocus enzyme electrophoresis (MLEE) to assess the amount and structure of genetic variation at enzyme encoding genes in a diverse, global collection of *E. coli* and *Shigella*. The results of this study showed that in terms of genetic distance, there is a very close affinity between *Shigella* and *E. coli,* and that the assignment of *Shigella* to distinct species was unwarranted from an evolutionary standpoint. In 1997, Pupo et al. supported and extended these findings by examining 32 strains including representatives of the major pathovars (EPEC, EHEC, ETEC) as well as 12 *Shigella* and 5 enteroinvasive *E. coli* strains (87). The bacteria were characterized by MLEE for 10 enzyme-encoding genes and nucleotide sequence for part of the *mdh* gene. Independently, both groups found that *Shigella* fell within the diversity of *E. coli* and that there were at least two distinct clusters with invasive strains, one including *S. boydii* serotypes and the other comprised of *S. flexneri* serotypes.

Recently, Reeves and colleagues (88) published the most extensive evolutionary analysis of *Shigella* spp. at the DNA sequence level. In this work, they determine the nucleotide sequence of a total 7,160 bp representing 8 housekeeping genes from 4 regions of the genome. Phylogenies constructed separately for each region were very similar in

topology with all but five of the *Shigella* strains falling into one of three main clusters. There was only a small amount of nucleotide diversity within clusters and most of the divergence occurred between clusters. Because the same genetic relationships were seen for the genes in each genomic region, and the clusters were all supported by high bootstrap confidence limits, Reeves and colleagues conclude that these clusters are robust and mark distinct phylogenetic groups (88). A summary of the serotypes found within each phylogenetic group is summarized in Table 1.

Taken together, three main conclusions are supported by the molecular evolutionary analysis. First, bacteria that belong to four traditional species of *Shigella* (*S. boydii*, *S. dysenteriae*, *S. flexneri*, and *S. sonnei*) fall within the genetic diversity of *E. coli*. Second, *Shigella* strains do not form a single lineage within *E. coli*, that is, they are not a monophyletic group but instead have multiple origins within *E. coli*. Finally, the recognized species of *Shigella* do not represent natural subgroups within *E. coli* but instead belong to genetically distinct clusters that do not concord with the phenotypic and antigenic properties that define the species classification.

Genomic evolution: gene acquisition and loss. The acquisition of new genes by horizontal transfer has played a major role in the adaptation and ecological specialization of bacterial lineages (61). It has been estimated, for example, that ~18% of the current genome of *E. coli* K-12 represents foreign DNA acquired by horizontal transfers since the divergence of *E. coli* and *Salmonella enterica* (62). Gene acquisitions have also contributed to the variation in virulence among strains and closely related bacterial species (43, 96). In *E. coli* and *S. enterica*, blocks of virulence genes, called pathogenicity islands, have been acquired at different times, thus generating a

Table 1. Classification of *Shigella* spp. serotypes into phylogenetic groups (Reeves Groups). This classification is based on the nucleotide sequence from four chromosomal regions (88).

Phylogenetic group	Serotypes
Group 1	B1, B2, B3, B4, B6, B8, B10, B14, B18
	D3, D4, D5, D6, D7, D9, D11, D12, D13
	F6, F6A
Group 2	B5, B7, B9, B11, B15, B16, B17
	D2
Group 3	B12
	F1A, F1B, F2A, F2B, F3A, F3B, F3C, F4A, F4B,
	F5, FX, FY
Others	B13
	D1, D8, D10
	SS

Abbreviations: B=Boydii, D=Dysenteriae, F=Flexneri, SS=Sonnei

variety of pathogens with distinct virulence genes and mechanisms of pathogenesis (41, 79, 80).

In the evolution of enteroinvasive *E. coli* and *Shigella*, gene acquisition has been important in two ways: first with the spread of the pINV plasmid that encodes invasive ability, and second with the presumed acquisition of a variety of mobile pathogenicity islands. Based on the sequence analysis of three genes, Lan et al. (58) found that the *Shigella* invasion plasmid can be classified into two homogeneous sequence types, called pINVA and pINVB. The plasmid sequence types have an interesting alignment with the Reeves groups: pINVA occurs in Group 1 strains whereas pINVB occurs in Group 3 strains. Both types are found in Groups 2 strains, several EIEC, as well as in Sonnei and Dysenteriae serotypes 1 and 10. This pattern supports the hypothesis that there have been several lateral transfers of the pINV plasmids to create new invasive lineages.

In addition, there have been five pathogenicity islands (PAIs) identified and characterized among invasive strains. One encodes a *Shigella* enterotoxin, three carry operons involved in iron scavenging, and the fifth specifies O-antigen modification (Table 2). SHI-1 (*Shigella* island 1; formerly known as *she*) is a 46.6 kb PAI located at the 3' end of the *pheV* tRNA gene in *S. flexneri* (Table 2). The island encodes several virulence factors including ShET1 (*Shigella* enterotoxin 1) whose activity was originally isolated in culture filtrates of a plasmid-cured strain which caused significant fluid accumulation in rabbit ileal loops (31). ShET1 is encoded by *set1*A and *set1*B and is associated predominantly with Flexneri 2a strains (5, 7, 77, 91). SHI-1 also encodes SigA, a cytopathic protease that contributes to intestinal fluid accumulation and Pic, a protease with mucinase and hemagglutinin activities. Interestingly, *set1* and *pic* have

overlapping reading frames. The island contains many intact and truncated mobile genetic elements, plasmid-related sequences, and several open reading frames (ORFs) with high sequence similarity to those found on O-islands in the *E. coli* O157:H7 genome.

SHI-2 is located on the chromosome near the *selC* tRNA locus, the site of insertion of pathogenicity islands in several other enteric pathogens. SHI-2 occurs in both Flexneri 2A and 5A strains (73, 124). The two versions of the island that have been studied differ in length but both encode an aerobactin system for iron acquisition, immunity to colicin V, and several other proteins (Table 2). It has been hypothesized that proteins produced by SHI-2 enhance fitness by facilitating bacterial survival under low iron conditions and promote survival in competitive situations with other bacteria.

SHI-3 is a 21 kb PAI that also carries genes for the synthesis and transport of aerobactin, as well as a P4 prophage-like integrase gene and numerous IS elements (89). This island was identified in a *S. boydii* B5 strain and is located at the *pheU* tRNA locus in some *S. boydii* isolates but not in others. Although the aerobactin operon is thought be advantageous in certain environments, an *S. boydii* aerobactin synthesis mutant (0-1392 *iucB*) did not differ from wild type in tissue culture assays of invasion and intercellular spread (89).

SHI-4 (or SRL for *Shigella* resistance locus) is a 66 kb island comprised of antibiotic resistance genes (*tet, cat, oxa-1*, and *aadA1*) closely linked to the *fec* operon, a ferric dicitrate iron transport system. The PAI is embedded in a larger (~99 kb) genetic element that is capable of precise excision (64). SHI-4 appears to be widely disseminated among *Shigella* although its distribution in light of the emerging phylogenetic

classification of shigellae and EIEC is unknown. The *fec* system is one of several iron uptake systems whose primary role in virulence may be in the uptake of iron from the intestinal lumen where exogenous citrate is available for chelation of iron (64).

In addition to the above islands, Adhikari (1) discovered that the serotype conversion genes in a Flexneri 1A strain occur on a unique segment of the chromosome and have many of the characteristics of a PAI, and has thus been referred to as the SHI-O island (53). Sequence analysis suggests that the present transposon-like structure of SHI-O was originally part of a bacteriophage that integrated near the *thrW-proA* attachment site (~ 6 min) in the K-12 genome. Interestingly, the opposite end of the element shows homology to the *dsdC* gene in *E. coli*, which maps to minute 53, suggesting that the Flexneri 1A chromosome has undergone additional genomic rearrangements.

In addition to gene acquisition, there is growing evidence that gene loss has been important in adaptive radiation and the evolution of bacterial virulence. For example, Maurelli and coworkers (67) present evidence that the universal deletion of the lysine decarboxylase gene (*cadA*) has enhanced the virulence of *Shigella* species because cadaverine, a product of the reaction catalyzed by lysine decarboxylase, inhibits the activity of the *Shigella* enterotoxin. Maurelli and coworkers refer to such large, universal deletions that enhance virulence as "black holes", the loss-of-function counterpart to pathogenicity islands.

Black hole formation is one example of pathogenicity-adaptive, or pathoadaptive, mutation (111). These genetic alterations represent a mechanism for enhancing bacterial virulence without horizontal transfer of specific virulence factors (111). Pathoadaptive mutations include, for example, increases in bacterial virulence by random functional

Table 2. Pathogenicity islands identified in *Shigella* strains.

Island	Serotype (strain)	tRNA (min)	Size in kb (no. ORFs)	Virulence factors	Reference
SHI-1	F2a (YSH6000)	*pheV* (67)	46.6 (31)	ShET1 enterotoxin (*set1*), autotransporter protease (*sigA*), mucinase (*pic*)	(6, 91)
SHI-2	F2a (SA100)	*selC* (82)	~30.0 (35)	aerobactin operon (*iucA-D*), aerobactin receptor (*iutA*), ColV immunity (*shiD*)	(124)
	F5a (M90T)	*selC* (82)	23.8 (24)		(73)
SHI-3	B5 (O-1392)	*pheU* (94)	20.9 (16)	aerobactin operon (*iucA-D*), aerobactin receptor (*iutA*)	(89)
SHI-4 (SRL)	F2a (YSH6000)	*serX* (24)	66.3 (59)	multiple antibiotic resistance genes, ferric dicitrate transport (*fec operon*)	(64)
SHI-O	F1a (Y53)	*thrW* (6)	10.6 (5)	serotype F1a O-antigen modification	(1)

mutations in a commensal trait that are adaptive for a pathologic environment, such as found for the FimH variants of uropathgenic *E. coli* (110).

Purpose. The overall objective of this research is to test hypotheses about the order and nature of gene acquisition and loss in the evolution of invasive *E. coli* and Shigellae. First, an evolutionary framework based on sequence polymorphisms in conserved housekeeping genes will be developed. This will provide a phylogenetic perspective for the divergence of the "backbone" of the genomes. The questions to be addressed are: To what extent has recombination created new alleles and multilocus genotypes? How tree-like is the divergence of clonal frames? What is the quality of the phylogenetic signal and does the rate of divergence fit the molecular clock hypothesis?

Second, the distribution of known and suspected virulence factors will be compared to the phylogenetic framework. The factors include, for example, genes that mark pINV and known pathogencity islands. The information will be incorporated into an evolutionary model that minimizes the number of acquisition events. The questions here include: How often have particular islands been gained or lost? Is there evidence of parallel changes in multiple groups? What component of the variation in virulence (invasiveness) is explained by the combination of acquired factors? The inferred evolutionary model can make predictions about the order and age of acquired elements which can be a tested by new data based on sequencing and phylogenetic analysis of the elements themselves.

Third, evidence for the formation of new black holes and novel islands will be investigated by developing a genomic method for finding major insertions and deletions. This method is based on the concept of paired-end sequencing and makes use of known

genomic sequences. It is expected that the application of this method will provide insights into the genomic alterations and molecular adaptations that accompany the shift to intracellular invasion and multiplication.

CHAPTER 2

PHYLOGENETIC RELATIONSHIP OF *ESCHERICHIA COLI* AND *SHIGELLA*

SUMMARY

Enteroinvasive *Escherichia coli* (EIEC) and *Shigella* species are bacteria that invade the mucosal epithelia of the intestine and are a major cause of dysentery worldwide. To determine the evolutionary relationships of these invasive pathogens to other *E. coli* pathovars, genetic variation was assessed by DNA sequencing of 15 housekeeping genes in 42 strains. The analysis reveals levels of nucleotide polymorphism ranging from 1.8% to 12.4% across loci with an average of 5.2%. Phylogenetic analysis indicates that most *Shigella* serotypes fall into one of three groups. *S. sonnei* and the *S. dysenteriae* serotypes 1 and 10 are distinct lineages independent of the other *Shigella* groups and *S. boydii* serotype 13 is a highly divergent lineage. The analysis also reveals distinct phylogenetic groups of EIEC with one strain (serotype O144:H-) clustering at the base of the Group 1 *Shigella*. A second cluster of EIEC includes serotypes O28, O29, O124, and O152 and appears to be closely related to *E. coli* O111:H21, an atypical enteropathogenic clone whose virulence mechanisms are poorly understood. Other EIEC serotypes fall outside of these clusters and are most closely related to Shiga-toxin producing *E. coli*. The analysis yielded identical groups of *Shigella* serotypes as those reported by Pupo and colleagues based on 8 different genes sequenced in 4 regions of the genome. The concordance of two independent studies based on different isolates and different genes shows that the approach is robust and indicates that recombination has not eliminated the phylogenetic signal in the history of divergence of the chromosomal backgrounds. In addition to the housekeeping loci, all strains were assayed for the presence of six virulence loci. Using the phylogenetic

framework, the timing of gene acquisitions and losses within the clonal groups could be inferred.

INTRODUCTION

There are four species of *Shigella; S. boydii, S. dysenteriae, S. flexneri,* and *S. sonnei,* that have been recognized historically because of the severity of disease and their clinical importance. The four species are identified and distinguished by biochemical traits (or lack thereof) and the expression of specific somatic antigens (reviewed in (26) and (38)). Within each of the four species, there is some antigenic variation with the number of serotypes ranging from 18 of *S. boydii,* to a single type of *S. sonnei.* Molecular evidence indicates the *S. sonnei* strains belong to a single widespread clone (55). Overall, there are 46 recognized serotypes of *Shigella* (59).

The enteroinvasive *E. coli* (EIEC) were first identified in Italy in the 1940s (29) and are similar to *Shigella* in that they can cause dysentery and exhibit an invasive phenotype. They have been incriminated in several large outbreaks of acute gastroenteritis in the United States (39, 47, 109, 121) and have also been implicated in traveler's diarrhea (128). In the developing world, EIEC infections contribute to endemic rates of diarrheal disease; enteroinvasive strains are typically isolated in 1 - 5 % of the cases of acute diarrhea in children (25, 32, 63, 92, 117, 120), although incidence rates vary with season (63) and socio-economic conditions (120). There is variation in the somatic antigens among EIEC strains, and eleven distinct O-types have been identified (O28, O29, O112, O124, O136, O143, O144, O147, O152, O164, and O167). With the exception of O124:H30 strains, EIEC are nonmotile and do not express flagellar antigens.

Application of the methods of evolutionary genetics began to elucidate the notion that invasive strains were evolutionarily related and represented a specialized natural group of bacteria. Ochman et al. (81) used multilocus enzyme electrophoresis (MLEE) to

assess the amount and structure of genetic variation at enzyme encoding genes in a diverse, global collection of *E. coli* and *Shigella*. The method revealed extensive protein polymorphisms for 12 enzymes, and resolved 3 major sub-specific groups of *E. coli* and 23 electrophoretic types (ETs) among 123 *Shigella* strains. The *Shigella* ETs fell within the diversity of the *E. coli* species as a whole. There were two distinct clusters of *Shigella* ETs, one cluster comprised mostly of strains of *S. flexneri* but also included ETs of *S. boydii* and *S. dysenteriae*. The other cluster contained strains belonging to all four species (81). This study demonstrated that in terms of genetic distance, there is a very close affinity between *Shigella* and *E. coli,* and that the assignment of *Shigella* to a distinct species was unwarranted from an evolutionary standpoint.

Pupo et al. (87) supported and extended these findings by examining 32 strains including representatives of the major pathovars (EPEC, EHEC, ETEC) as well as 12 *Shigella* and 5 enteroinvasive *E. coli* strains. The bacteria were characterized by MLEE for 10 enzyme-encoding genes. They also sequenced part of the *mdh* gene to infer the genetic relationships of pathogenic strains to isolates of the *E. coli* Reference collection (ECOR) set. They found that *Shigella* fell within the diversity of *E. coli* and that there were at least two distinct clusters with invasive strains, one including *S. boydii* serotypes and the other comprised of *S. flexneri* serotypes.

Recently, Reeves and colleagues (88) published an evolutionary analysis of *Shigella* spp. at the DNA sequence level. In this work, they determined the nucleotide sequence of a total 7,160 bp representing 8 housekeeping genes from 4 regions of the genome. Comparison of the sequences among 46 *Shigella* strains revealed substantial DNA polymorphism with the identification of more than 150 informative sites.

Phylogenies constructed separately for each region were very similar in topology with all but five of the *Shigella* strains falling into one of three main clusters. There was only a small amount of nucleotide diversity within clusters and most of the divergence occurred between clusters. Because the same genetic relationships were seen for the genes in each genomic region, and the clusters were all supported by high bootstrap confidence limits, Reeves and colleagues concluded that these clusters are robust and mark distinct phylogenetic groups (88).

A crucial result from the evolutionary analysis is that the phylogenetic groups contain serotypes of different *Shigella* species (Table 1). Group 1 includes 9 serotypes of *S. boydii*, 9 serotypes of *S. dysenteriae*, and 2 serotypes of *S. flexneri.* Group 2 includes 7 serotypes of *S. boydii* and *S. dysenteriae* type 2. Group 3 contains 12 *S. flexneri* serotypes as well as *S. boydii* type 12. Five *Shigella* serotypes do not fall within these clusters and are distinct from one another. These distinct clones are *S. dysenteriae* types 1, 8, and 10, *S. sonnei*, and *S. boydii* type 13.

Taken together, three main conclusions are supported by the previous molecular evolutionary analyses. First, bacteria that belong to four traditional species of *Shigella* (*S. boydii*, *S. dysenteriae*, *S. flexneri*, and *S. sonnei*) fall within the genetic diversity of *E. coli.* Second, *Shigella* strains do not form a single lineage within *E. coli*, that is, they are not a monophyletic group but instead have multiple origins within *E. coli.* Finally, the recognized species of *Shigella* do not represent natural subgroups within *E. coli* but instead belong to genetically distinct clusters that are not concordant with the phenotypic and antigenic properties that define the species classification.

Previous work in our laboratory has focused on developing a phylogeny based on seven housekeeping loci within EPEC and EHEC strains (94). This phylogeny was used as the framework to demonstrate the hypothesized time of virulence gene acquisition in these strains. Here, a similar approach is used to investigate the genetic relatedness of enteroinvasive *E. coli* and *Shigella* strains.

MATERIALS AND METHODS

Bacterial strains. All strains were grown overnight in LB broth at 37°C. The strains examined include 27 *Shigella* and 15 enteroinvasive *E. coli*. The *Shigella* strains were obtained from the CDC reference collection and include serotypes from each of the traditional species. The designation of the *Shigella* serotypes will follow the designation set forth by Pupo (88). This study includes: 8 *S. boydii* of serotypes B2 (4444-74), B4 (3594-74), B5 (3408-67), B9 (291-75), B11 (5254-60), B13 (C-425), B14 (2770-51), B15 (965-58), and B17 (3615-53); 7 *S. dysenteraie* of serotypes D1 (1007-74, 3823-69), D2 (155-74), D3 (225-75), D7 (3470-56), D10 (5514-56), and D12 (3341-55); 8 *S. flexneri* of serotypes F1 (2702-71), F2A (2457T, 2747-71), F5 (2794-71, 1170-74), and F6 (1043-82, 1485-50, 3138-88), and 3 *S. sonnei* (4822-66, 3226-85, 3233-85). CDC or Dr. Luis Trabulsi (66) supplied the EIEC strains which included serotypes O28:H21 (1758-70), O29:H27 (1827-79), O29:NM (1885-77), O124:H- (929-78), O124:H30 (5898-71, 202-72), O28:H- (LT-15, LT-26), O136:H- (LT-41), O143:H- (LT-62), O144:H- (LT-68), O152:H- (LT-99), O164:H- (LT-91), O167:H- (LT-82), and O-:H- (LT-94). Additionally, 20 strains from a previous study by Reid et. al. (94) along with O26:NM (395-2, EPEC1), O119:H6 (277-84, EHEC2), O44:H18 (O42, EAEC), and O11:H21 (5338-66, atypical EPEC) were used to survey the relationships of the various pathovars. Freezer stocks were made for each isolate and stored at -70°C.

DNA isolation. In preparation for sequencing, genomic DNA was isolated from 1 ml of overnight culture using the Puregene DNA isolation kit (Gentra Systems Inc., Minneapolis, MN). DNA preparations were then electrophoresed on 0.8% agarose gels, stained with ethidium bromide, and the DNA concentrations were determined using the

LasPro software. DNA preparations were diluted to a final concentration of 100 ng/µl and stored at 4°C

PCR amplification. Oligonucleotide primers were designed to amplify internal fragments for 15 housekeeping genes (Table 3). Six of these genes were shown to be useful for identifying clonal frames in a previous study of pathogenic *E. coli* (94). These primers were used to amplify *arcA, aroE, aspC, clpX, cyaA, dnaG, fadD, grpE, icdA, lysP, mdh, mtlD, mutS, rpoS,* and *uidA* in the *Shigella* and EIEC isolates. Each amplification reaction included primers at a final concentration of 0.5 mM, 0.2mM of each dNTP, 3 U of Amplitaq Gold (Applied Biosystems, Foster City, CA), and 100 ng of template. PCR was performed for 35 cycles under the following conditions: 1 min of denaturation at 92°C, 1 min of primer annealing at 57°C, and 15 sec of extension at 72°C with an initial denaturing step of 94°C for 10 min. Amplicons were electrophoresed on a 0.8% agarose gel and visualized. PCR products were purified using the Qiaquick PCR Purification Kit (Qiagen, Valencia, CA). Purified products were electrophoresed on a 0.8% agarose gel and the concentration was determined.

Nucleotide sequencing and alignment. Cycle sequencing reactions were performed with CEQ dye terminator cycle sequencing kits (Beckman-Coulter, Fullerton, CA) with approximately 50 fmol of template and a final primer concentration of 2 µM. The thermal cycle was run for 30 cycles with the following parameters: 96°C, 20 sec; 50°C, 20 sec; and 60°C, 4 min. Reactions were purified using Sephadex columns and dried under vacuum centrifugation at room temperature. The samples were then rehydrated in 40 µl of formamide and sequenced using a Beckman CEQ 2000XL (Beckman-Coulter, Fullerton, CA) capillary sequencer.

Table 3. Primer sequences, positions, and amplicon sizes for 15 housekeeping loci.

Locus	Primer	Primer sequence	Position in gene	Size of amplicon
arcA	arcA-F1	5' - GACAGATGGCGCGGAAATGC - 3'	99..118	552 bp
	arcA-R2	5' - TCCGGCGTAGATTCGAAATG - 3'	631..650	
aroE	aroE-F1	5' - GCGTTGGCTGGTGCTGTTA - 3'	238..256	362 bp
	aroE-R2	5' - GGGATCGCCGGAATATCACC - 3'	580..599	
aspC	aspC-F4	5' - GTTTCGTGCCGATGAACGTC - 3'	57..76	594 bp
	aspC-R7	5' - AAACCCTGGTAAGCGAAGTC - 3'	631..650	
clpX	clpX-F6	5' - CTGGCGGTCGCGGTATACAA - 3'	262..281	672 bp
	clpX-R1	5' - GACAACCGGCAGACGACCAA - 3'	914..933	
cyaA	cyaA-F3	5' - CTCGTCCGTAGGGCAAAGTT - 3'	312..331	571 bp
	cyaA-R3	5' - AATCTCGCCGTCGTGCAAAC - 3'	863..882	
dnaG	dnaG-F9	5' - ACCGCCGATCACATACAACT - 3'	868..887	512 bp
	dnaG-R6	5' - TGCACCAGCAACCCTATAAG - 3'	1360..1397	
fadD	fadD-F6	5' - GCTGCCGCTGTATCACATTT - 3'	768..787	580 bp
	fadD-R3	5' - GCGCAGGAATCCTTCTTCAT - 3'	1328..1347	
grpE	grpE-F1	5' - CCCGGAAGAAATTATCATGG - 3'	39..58	488 bp
	grpE-R4	5' - TCTGCATAATGCCCAGTACG - 3'	507..526	
icdA	icd-F2	5' - CTGCGCCAGGAACTGGATCT - 3'	352..371	669 bp
	icd-R2	5' - ACCGTGGGTGGCTTCAAACA - 3'	1001..1020	
lysP	lysP-F1	5' - CTTACGCCGTGAATTAAAGG - 3'	36..55	628 bp
	lysP-R8	5' - GGTTCCCTGGAAAGAGAAGC - 3'	644..663	
mdh	mdh-F3	5' - GTCGATCTGAGCCATATCCCTAC - 3'	130..152	650 bp
	mdh-R4	5' - TACTGACCGTCGCCTTCAAC - 3'	760..779	
mtlD	mtlD-F2	5' - GCAGGTAATATCGGTCGTGG - 3'	22..41	658 bp
	mtlD-R3	5' - CGAGGTACGCGGTTATAGCAT - 3'	659..679	
mutS	mutS-F1	5' - GGCCTATACCCTGAACTACA - 3'	1488..1507	596 bp
	mutS-R1	5' - GCATAAAGGCAATGGTGTC - 3'	2065..2083	
rpoS	rpoS-F3	5' - CGCCGGATGATCGAGAGTAA - 3'	274..293	618 bp
	rpoS-R1	5' - GAGGCCAATTTCACGACCTA - 3'	872..891	
uidA	uidA-277F	5' - CATTACGGCAAAGTGTGGGTCAAT - 3'	277..300	658 bp
	uidA-934R	5' - CCATCAGCACGTTATCGAATCCTT - 3'	911..934	

Sequences were concatenated and aligned with the SeqMan module in the DNAStar Lasergene (Lasergene, Madison, WI) computer software package. Sequences were aligned individually using the K-12 sequence as a reference. The 15 loci from the published genome sequences of *E. coli* K-12 (9) and O157:H7 (49, 85), *Salmonella typhimurium* LT-2 (69), and *S. flexneri* Sf301 (54) were added to the data set. Consensus sequences were aligned with ClustalX (119) and the output files were modified for use in MEGA2 (57).

Sequence analysis. Phylogenetic trees were constructed using the neighbor-joining algorithm (99) with the MEGA2 program (57). Trees were based on synonymous distance (d_S) calculated by the modified Nei-Gojobori method (76) with a Jukes-Cantor correction.

Mannitol utilization. Because some isolates had an absent or larger than expected amplicon for the *mtlD* locus, the extent of this loss was examined using PCR and mannitol utilization assays. *S. dysenteriae* isolates of serotypes D1 (1007-74 and 3823-69), D2 (155-74), D3 (225-75 and 2415-49), D4 (1112-78 and 2045-75), D6 (852-59 and 3514-76), D7 (3470-56), D9 (653-82), D10 (5514-56), D11 (3873-50), D12 (3341-55) and *S. flexneri* serotype F6 (1043-82, 1141-81, 1148-83, 1485-80, 3138-88, 3469-89, 3500-89, 3638-77, and 910-81) were grown overnight at 37°C on MacConkey agar containing mannitol. Strains with the ability to utilize mannitol exhibit pink colony morphology while the strains that are unable to utilize mannitol grow as white colonies (Figure 2). DNA was isolated as described earlier. Oligonucleotide primers were designed to amplify the entire *mtlD* locus and the adjacent loci, *mtlA* and *mtlR*. The primers sequences are: mtlD2160 5' – TTGGCGCAGGTAATATCGGT – 3', mtlD3252

5' – ACCTCGCTGTTGGCATCAAG – 3', mtlA122 5' – GGTGGTTACCGAACGAGACG, mtlA1909 5' – TACGACCTGCCAGCAGTTCC – 3', mtlR3382 5' – CGTGTGCTTGAGCGTCTGAA – 3', and mtlR3819 5' – CATTGTTGAGCGCACAGCCT – 3'. PCR amplification was performed as described above under the following conditions: 92°C for 1 min, 54°C for 1 min, 72°C for 3 min for 35 cycles with an initial denaturing step of 94°C for 10 min. Amplicons were purified and prepared for nucleotide sequencing as described above.

Virulence loci. To examine the distribution of previously identified virulence loci, isolates were screened by PCR for the presence or absence of the following loci: *pic*, *senA*, *she*, *shuA*, *iucD*, and *iutA*. The amplifications were performed as described above using the primers in Table 4 with the annealing temperature ranging from 51°C to 61°C depending on the locus. Products were electrophoresed, purified, sequenced and analyzed as described above.

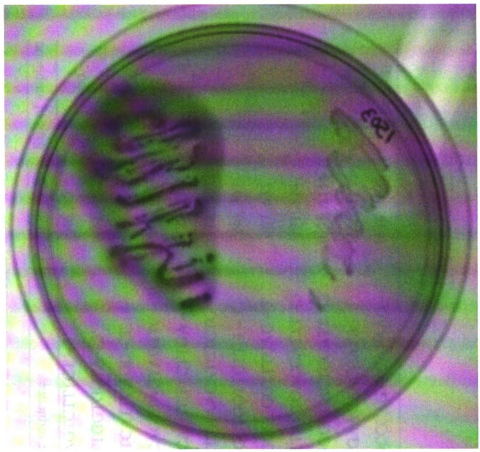

Figure 2. Bacterial growth on mannitol MacConkey agar. Mannitol positive (2415-49, left) and negative (225-75, right) isolates are shown after overnight growth at 37°C.

Table 4. Primer sequences, positions, and amplicon sizes for *Shigella* virulence loci.

Locus	Primer	Primer sequence	Position in gene	Size of amplicon
pic	pic46	5' – ACGGTTTTCCCAGTCTTTCT – 3'	46..65	436 bp
	pic481	5' – TATATCCCCCTTTGGTGGTA – 3'	481..462	
senA	senA486	5' – GGGGGATTTTGTCATTCAGC – 3'	486..505	490 bp
	senA975	5' – CATTCCTTCCGCAGTTAGTAGTTC – 3'	975..952	
she	she2722	5' – GCGTTTGCCGGCAGTCTTCTTGA – 3'	2722..2744	397 bp
	she3118	5' – GGGTGCCGGCCTTAAACATCTCTG – 3'	3118..3095	
shuA	shuA286	5' – CATCGCGGCGTGCTGGTTCTTG – 3'	286..307	372 bp
	shuA657	5' – CTCGTCATTCGGCGCGGTTTCAC – 3'	657..635	
iucD	iucD-F2	5' – CTGCCGATTCTTTACTGACCATTT – 3'	791..814	324 bp
	iucD-R2	5' – TATTTTCTTTCGGGCCACTCCATT – 3'	1091..1114	
iutA	iutA-F4	5' – TCCGGAGGGCTGAACAACCATA – 3'	1165..1186	475 bp
	iutA-R1	5' – TAAAGCGCCAGCCCAGTTCATAAG – 3'	1616..1639	

RESULTS

Amplification of housekeeping loci. All housekeeping loci were amplified in *Shigella* and EIEC isolates with the exception of the *mtlD* locus, which was absent in *S. dysenteriae*, and *S. flexneri* serotype 6 isolates. Interestingly, the Dysenteriae 3 isolate had an amplicon larger in size, whereas the amplicon was absent in other Dysenteriae strains. Further investigation of this anomaly identified the presence of an IS2-like element inserted near the 5' end of the *mtlD* gene.

Phylogenetic analysis. The phylogenetic analysis of the *Shigella* and EIEC isolates identifies 5 distinct clusters of invasive isolates (Figure 3). Three of these clusters were previously identified by an independent investigation by Pupo (88). The Group 1 *Shigella* (Figures 3 and 4) is comprised of serotypes B2, B4, B14, D3, D7, D12, and F6. Group 2 *Shigella* (Figures 3 and 5) includes serotypes B5, B9, B11, B15, B17, and D2. The Group 3 *Shigella* (Figures 3 and 6) are predominately Flexneri isolates (F1, F2A, F5) with the exception of a single Sonnei isolate that clusters with these strains. *Shigella* of serotypes D1, D10, B13 and Sonnei fell outside of the identified groups (Figure 3). The D10 isolate appears to be related to and EAEC strain, O42, and the B13 isolate is highly divergent falling between the *Salmonella* and *Escherichia* lineages. The D1 isolates form a cluster with two EIEC isolates suggesting the rise of an additional invasive group.

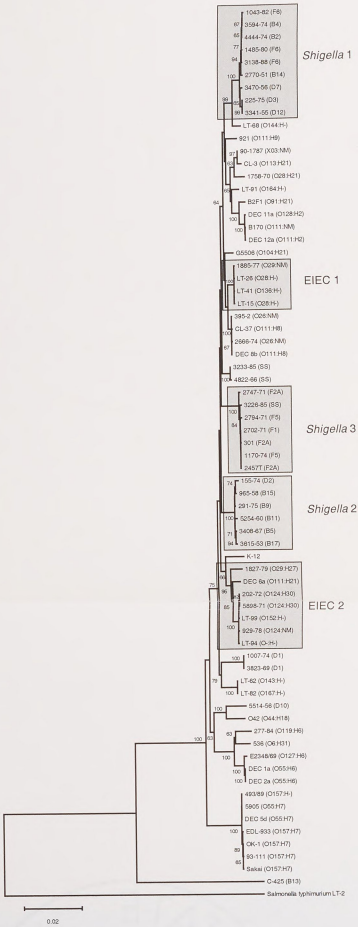

Figure 3. Neighbor-joining tree of *Escherichia coli* isolates based on 15 housekeeping loci. Bootstrap values are indicated at the internal nodes. The branch lengths are measured as the number of synonymous substitutions per site. The main groups of *Shigella* and EIEC are indicated with shaded boxes.

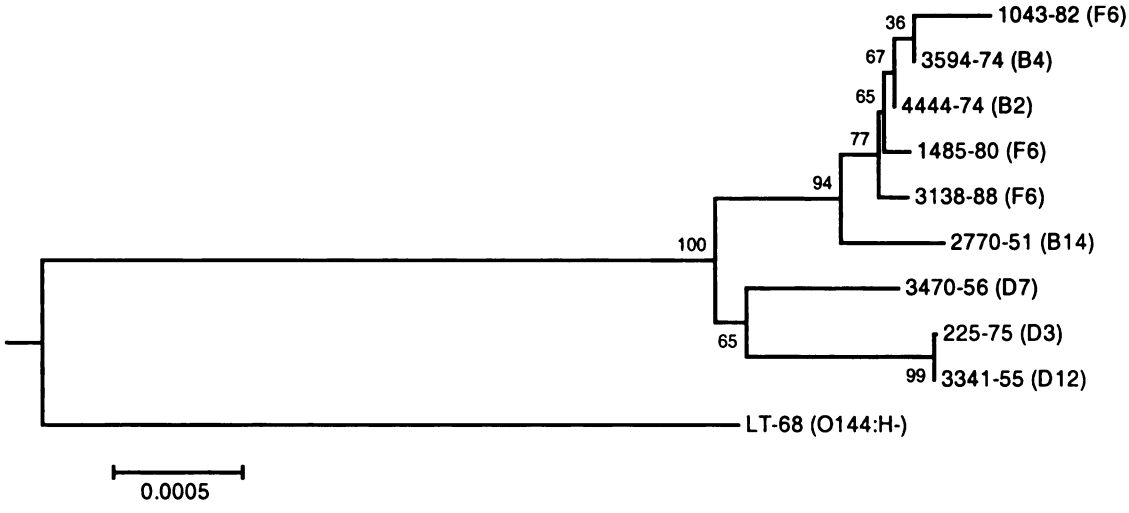

Figure 4. Subtree of the Group 1 *Shigella* isolates. The serotype is indicated in parentheses. The grouping of the *Shigella* isolates is supported by a bootstrap value of 100. An EIEC isolate, LT-68, falls just outside of this group of *Shigella*. The branch lengths are measured as the number of synonymous substitutions per site.

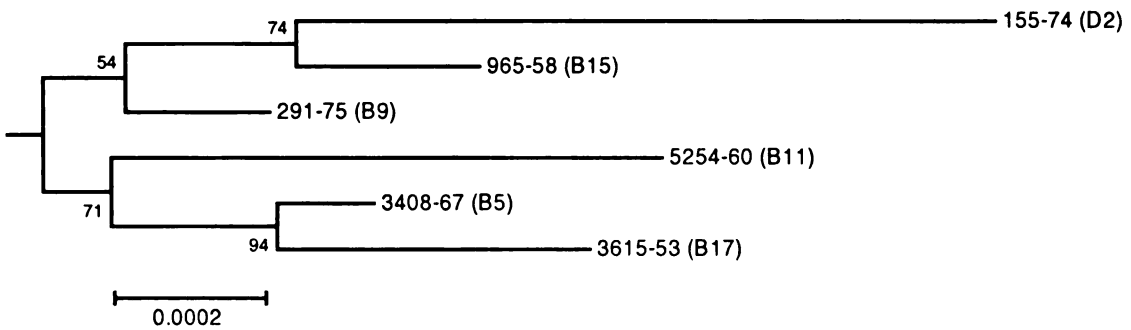

Figure 5. Subtree of the Group 2 *Shigella* isolates. The serotypes are indicated in parentheses. This cluster consists predominately of *S. boydii* serotypes but also contains the *S. dysenteriae* 2 serotype. The branch lengths are measured as the number of synonymous substitutions per site.

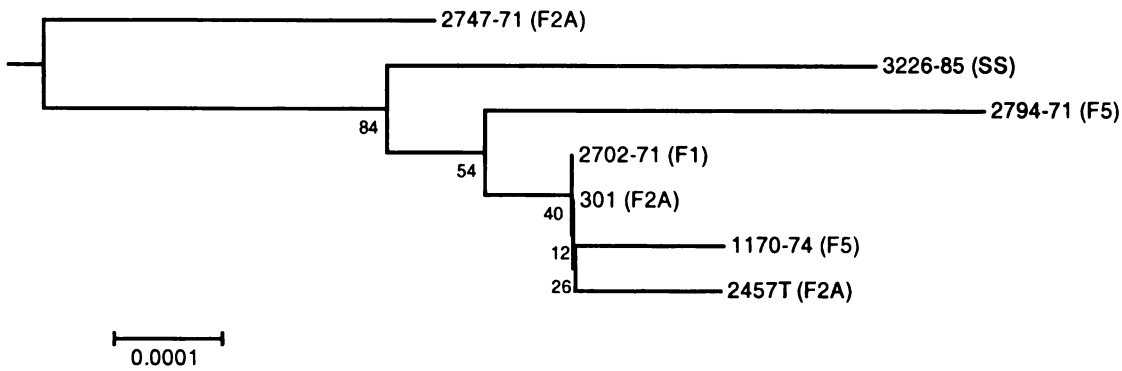

Figure 6. Subtree of the Group 3 *Shigella* isolates. The serotypes are indicated in parentheses. This group consists of *S. flexneri* serotypes along with a single *S. sonnei* isolate. Branch lengths are measured as the number of synonymous substitutions per site.

The EIEC also show clustering patters like the Shigellae (Figure 3). Two distinct groups of EIEC are identified with Group 1 EIEC (Figure 7) consisting of serotypes O28, O29, and O136. This group is related to reference isolate G5506 (O104:H21), a shiga toxin-producing *E. coli*. Another group of EIEC isolates (Group 2 EIEC) include serotypes O29, O124, O152, and O- (Figure 8). Interestingly, an atypical EPEC isolate (O111:H21) falls into this group. The Group 2 EIEC isolates are most closely related to the *E. coli* K-12 reference isolate. Like the Reeves classification of the Shigellae, there are EIEC isolates that can be classified as "other" due to the fact that they are not associated with the major groups. Three EIEC isolates are associated with clusters of *Shigella* isolates. An EIEC isolate of serotype O144:H- falls just outside of the Group 1 *Shigella* (Figure 4) while 2 EIEC serotypes (O167:H- and O143:H-) cluster with Dysenteriae 1 strains (Figure 9). Two EIEC isolates with serotypes O28:H21 and O164:H- appear to be most closely related to STEC and EPEC reference isolates.

Variation in housekeeping loci. Out of a total of 7,452 nucleotide bases, there were 765 variable sites in the *Shigella* and EIEC housekeeping loci. When the highly divergent Boydii 13 isolate was excluded, 390 variable sites result. The percentage of polymorphic sites ranges from 1.8% to 12.4% with an average of 5.2% across the 15 housekeeping loci (Table 5). There are no changes at the amino acid level for ArcA or LysP, while 14 amino acid substitutions occur in UidA. A plot of nucleotide changes against amino acid changes identifies 4 outlying loci, *aroE, fadD, mutS*, and *uidA* (Figure 10). Of these loci, *mutS* has the highest number of nucleotide changes however many are synonymous changes resulting in only 3 changes at the amino acid level.

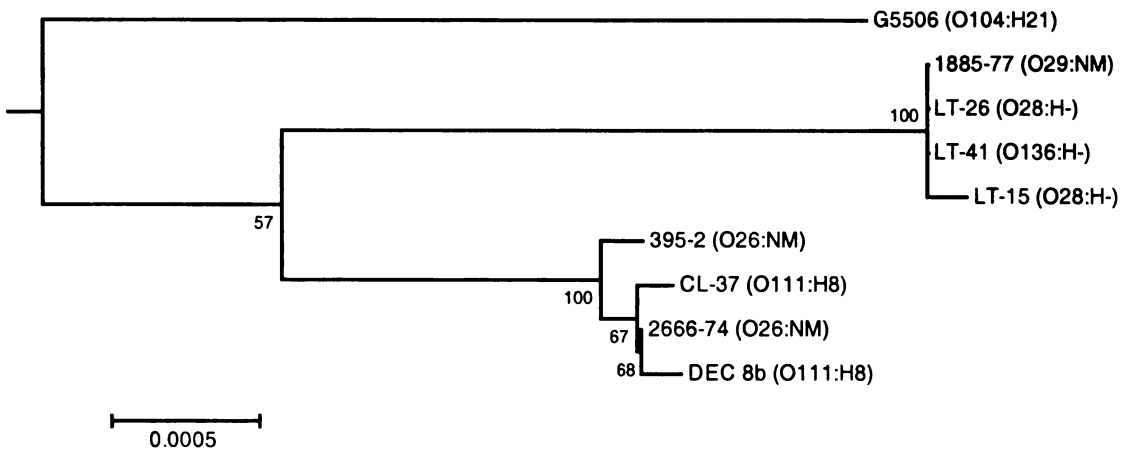

Figure 7. Subtree of the Group 1 EIEC isolates. The serotypes of each isolate are indicated in parentheses. This grouping of four is supported by a bootstrap value of 100. The branch lengths are measured as the number of synonymous substitutions per site.

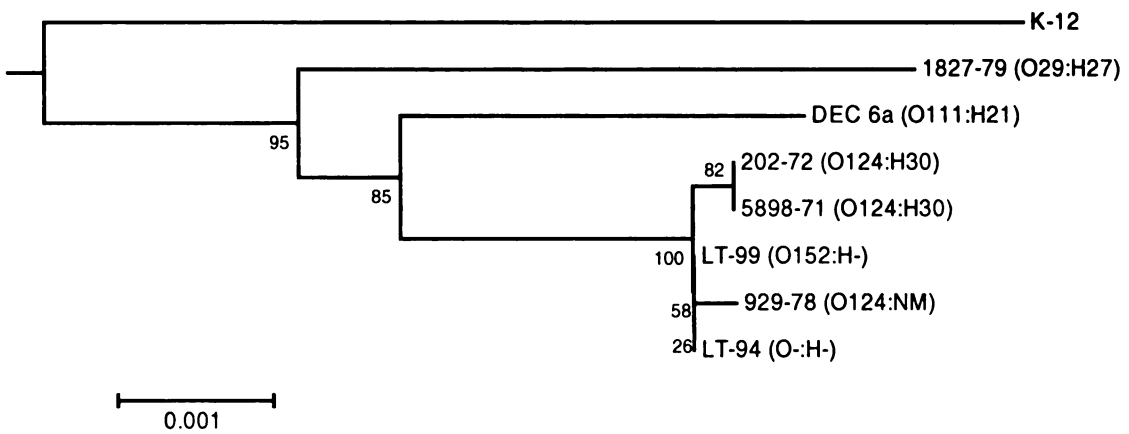

Figure 8. Subtree of the Group 2 EIEC isolates. The serotypes are indicated in parentheses. This clustering is supported by a bootstrap value of 86. The phylogenetic analysis shows this group being most closely related to the laboratory strain, K-12, and also contains an atypical enteropathogenic *E. coli* isolate, DEC 6a. Branch lengths are measured as the number of synonymous substitutions per site.

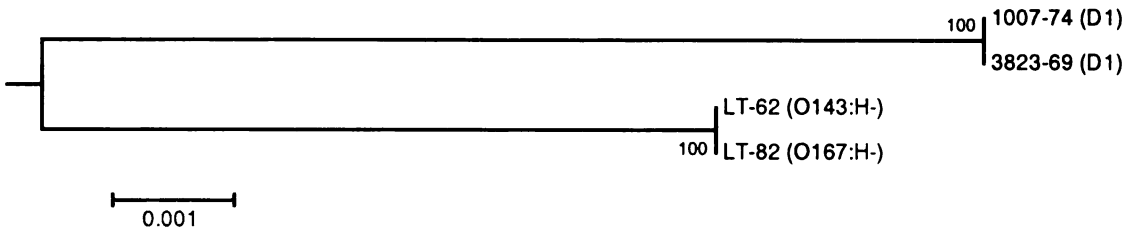

Figure 9. Subtree showing the relationship between *S. dysenteriae* type 1 and EIEC isolates. Serotypes of each isolate are indicated in parentheses. The Dysenteriae 1 isolates are phylogenetically distinct from the other *Shigella*. Branch lengths are measured as the number of synonymous substitutions per site.

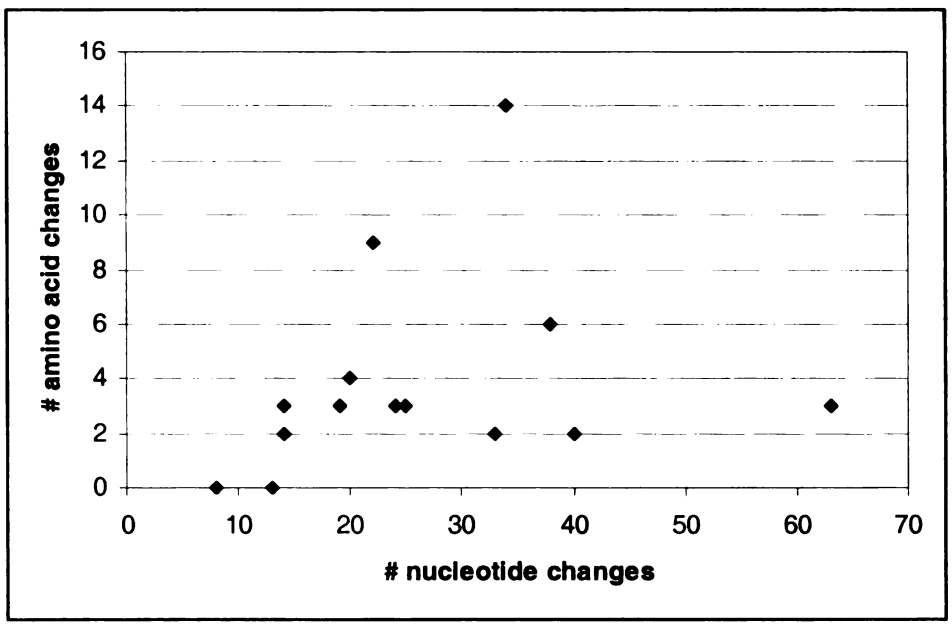

Figure 10. Plot of the number of nucleotide changes versus the number of amino acid changes for the 15 housekeeping loci. MutS shows the most nucleotide changes with a relatively low number of amino acid changes; however, UidA has the most amino acid changes followed by AroE and FadD.

Table 5. Variability within each housekeeping locus. Variation is measured at both the nucleotide and predicted amino acid level. The highly variable Boydii 13 isolate is not included in the analysis.

Locus	Gene Product	Total Sites	Variable Sites (%)	Amino Acid Variation (%)
arcA	Aerobic respiratory control protein	435	8 (1.8)	0
aroE	Shikimate dehydrogenase	291	22 (7.6)	9 (9.2)
aspC	Aspartate amino transferase	513	24 (4.7)	3 (1.8)
clpX	ATP-binding subunit of *clp* protease	567	40 (7.1)	2 (1.1)
cyaA	Adenylate cyclase	498	25 (5.0)	3 (1.8)
dnaG	Primase to initiate DNA replication	444	14 (3.2)	3 (2.0)
fadD	Acyl-CoA synthetase	492	38 (7.7)	6 (3.7)
grpE	Heat shock protein	417	14 (3.4)	2 (1.4)
icdA	Isocitrate dehydrogenase	567	33 (5.8)	2 (1.1)
lysP	Lysine-specific permease	477	13 (2.7)	0
mdh	Malate dehydrogenase	549	20 (3.6)	4 (2.2)
mtlD	Mannitol-1-phosphate dehydrogenase	540	24 (4.4)	3 (1.7)
mutS	DNA mismatch repair protein	507	63 (12.4)	3 (1.8)
rpoS	RNA polymerase subunit sigma-38	567	19 (3.4)	3 (1.6)
uidA	Beta-glucuronidase	588	34 (5.8)	14 (7.1)

Genetic diversity within *Shigella* and EIEC. The *Shigella* groups defined by the phylogenetic analysis were used to examine genetic diversity. The percentage of variable sites within the three *Shigella* groups averages 0.27% (Table 6). The EIEC 1 group has 51 variable sites, while the variability within the EIEC 2 group is more conservative with only one site of nucleotide variation. By adding the EIEC isolate LT-68 to the analysis of the Group 1 *Shigella*, the percentage of variable sites increases slightly to 0.8%. The putative invasive group of Dysenteriae 1 and EIEC isolates (LT-62 and LT-82) has the highest percentage of variable sites (1.2%) resulting in 9 changes at the amino acid level.

Loss of mannitol transport and utilization. The amplification of housekeeping loci provided evidence that the amplicon for the mannitol-1-phosphate dehydrogenase locus (*mtlD*) is either absent or larger than expected for the Dysenteriae and Flexneri 6 strains. When the pattern is examined phylogenetically, two interesting observations are apparent. First, a loss or change in the *mtlD* locus occurred independently at four different stages in the evolution of the Shigellae. Second, the *S. dysenteriae* isolates with larger amplicons fall exclusively within Group 1 of the Reeves classification, and in addition, some *S. flexneri* serotype 6 isolates of Group 1 are also mannitol negative. The *mtlA* locus was absent in all serotypes examined and the *mtlR* locus was present only in the D2 and D3 serotypes. These observations suggest that natural selection has favored inactivation of the mannitol operon.

Nucleotide sequencing was used to address the phenomenon of the larger than expected amplicon for the *mtlD* locus in the Group 1 *S. dysenteriae* isolates. Five additional isolates of serotypes of Reeves Group 1 were also analyzed. One D3 isolate

Table 6. Genetic diversity within the derived phylogenetic groups of *Shigella* and EIEC. The measurement of diversity is expanded to include the closely related EIEC isolates (LT-68 with Group 1; LT-62 and LT-82 with the Dysenteriae type 1).

	Number of Isolates	Variable Sites (%)	d_S x 100	d_N x 100	Amino Acid Variation (%)
Group 1	9	20 (0.3)	0.2 ± 0.1	0.1 ± 0.0	0
Group 2	6	23 (0.3)	0.2 ± 0.1	0.1 ± 0.0	0
Group 3	6	15 (0.2)	0.1 ± 0.0	0.0 ± 0.0	0
Sonnei	2	2 (0)	0.0 ± 0.0	0.0 ± 0.0	0
EIEC 1	4	51 (0.7)	0.0 ±0.0	0.0 ± 0.0	0
EIEC 2	6	1 (0)	0.8 ± 0.1	0.0 ± 0.0	0
Group 1, LT-68	10	56 (0.8)			0
D1, LT-62, LT-82	4	90 (1.2)			9 (0.4)

has an intact *mtlD* locus with 28 nucleotide changes (5 amino acid changes) compared to the published K-12 sequence. Seven isolates (D3, D4, D6, D9, and D11) have an IS2-like element inserted near the 5' end of the gene. The insertion site occurs between bases 208 and 209. The insertion element is 1336 bp and has homology to *tpnG* and *tpnF* of the SHI-2 pathogenicity island (Genbank AF141323) and *int* loci of bacteriophage SfX (Genbank BXU82084). The first 208 bp of *mtlD* differs from the published K-12 sequence by only one nucleotide, however, after the interruption, the gene differs by 29 nucleotide differences with most of the variation occurring close to the 3' end of the locus. The molecular analysis of these isolates correlates with the results obtained in the phenotypic assay using mannitol MacConkey agar as an indicator of a functional mannitol operon and is summarized in Table 7.

Acquisition of virulence loci. The distribution of known pathogenicity islands was addressed by PCR detection of associated virulence genes. This information is useful in devising and testing an evolutionary model for the acquisition of mobile virulence elements in invasive strains. Primers were designed for the PCR detection of 6 putative virulence genes that have been associated with pathogenicity islands. The virulence loci include: *set,* ShET1 enterotoxin (SHI-1, Group 3, F2A) (31); *pic,* mucinase and hemagglutinin activity (SHI-1, Group 3, F2A) (77); *iutA* and *iucD*, aerobactin transport genes (SHI-2, SHI-3) (73, 89); and *shuA*, heme binding gene (D1)(71). The *senA* gene was used to detect the presence of the pINV plasmid (75).

All invasive strains used for multilocus sequencing as well as some additional isolates representative of the inferred phylogenetic groups (52) were screened for the presence of these virulence loci. A summary of the findings is presented

Table 7. Mannitol genotypes and phenotypes in Dysenteriae and Flexneri 6 isolates. The presence (+) or absence (-) of the mannitol dehydrogenase locus and insertion element was determined by PCR amplification. Overnight growth on mannitol MacConkey agar was used as a phenotypic assay to indicate a functional (+) or non-functional (-) mannitol operon.

Species or pathovar	TW Number	Strain	Serotype	Locale	Year	Source	*mltD* locus	Insert	Mannitol MacConkey Agar
S. dysenteriae	07448	1112-78	D4	Panama (Canal Zone)	1978	CDC	+	+	-
	07585	2054-75	D4	USA (Massachusetts)	1975	CDC	+	+	-
	01514	852-59	D6	Russia (Moscow)	1959	CDC	+	+	-
	07584	3514-76	D6	USA (New Jersey)	1976	CDC	+	+	-
	01513	653-82	D9		1982	CDC	+	+	-
	01503	225-75	D3	India (Bombay)	1975	CDC	+	+	-
	01504	2415-49	D3		1949	CDC	+	-	+
	01506	3341-55	D12	USA (Arizona)	1955	CDC	-		-
	01507	3470-56	D7	No Data	1956	CDC	-		-
	02609	1007-74	D1	USA (California)	1974	CDC	-		-
	02630	3823-69	D1	Guatemala	1969	CDC	-		-
	02615	155-74	D2	USA (California)	1974	CDC	-		-
	02637	5514-56	D10	Rhodesia	1956	CDC	-		-
	01508	3873-50	D11	Mexico (Mexico City)	1950	CDC	+	+	-
S. flexneri	07572	3138-88	F6	USA (Massachusetts)	1988	CDC	+	-	+
	07573	1485-80	F6	USA (Michigan)	1980	CDC	+	-	+
	07574	1148-83	F6	USA (Nevada)	1983	CDC	+	-	+
	07575	1141-81	F6	USA (Virginia)	1981	CDC	+	-	+
	07569	3638-77	F6		1969	CDC	-		-
	07570	3500-89	F6	USA (Pennsylvania)	1989	CDC	-		-
	07571	3469-89	F6	USA (Illinois)	1989	CDC	-		-
	07576	1043-82	F6	USA (Colorado)	1982	CDC	-		-
	07577	910-81	F6	USA (Arizona)	1981	CDC	-		-

in Table 8. The *senA* locus is present in the majority of the *Shigella* and EIEC isolates and has 14 variable sites with all of the changes being synonymous. The *senA* locus is absent in two Group 3 isolates, the highly divergent B13 serotypes, DEC6a, Albert 10457, and 3097-02, a recent *Shigella* isolate of unknown serotype. Two loci, *setlA* and *pic,* were present and sequenced in 6 isolates. These loci occur in mainly in Group 3 but also occur in the 3097-02 *Shigella* isolate and three EIEC isolates, LT-15 (EIEC1), LT-94 (EIEC2), and 929-78 (EIEC2). There is little variation at the nucleotide level with only 2 variable sites in *setlA*. The heme transport locus, *shuA* is present in Dysenteriae type 1 and type 10 strains, two EIEC isolates (LT-62 and LT-82) and two Boydii 13 isolates (3556-77 and 3054-94). The two EIEC isolates are noteworthy in that they group with the Dysenteriae type 1 isolates (Figure 3 and 9). A phylogenetic analysis based on a subset of the housekeeping loci used in this study shows a relatively close relationship between the Dysenteriae type 1/EIEC cluster and a cluster containing some atypical Boydii 13 isolates (52). The distribution of the SHI-2 and SHI-3 pathogenicity islands was examined by screening the isolates for the *iucD* and *iutA* loci. These findings were variable within and between the groups. Both loci were sequenced in six isolates to determine the origin of the island. There are 5 variable sites in *iucD* that correspond to the SHI-2 sequence; however, the 3 variable sites identified in *iutA* are not unique to either island. Overall, the nucleotide sequences suggest that these virulence genes are highly conserved among the invasive clones. The distributions of the aerobactin loci indicate that *Shigella* groups 1, 2, and 3 have acquired the SHI-2, SHI-3 or possibly a previously unidentified aerobactin island. The presence of these two loci is variable in all other invasive groups. Figure 11 uses the phylogenetic framework to show the

hypothesized timing of gene acquisition and loss in the evolutionary history of the invasive *E. coli*.

Table 8. Acquisition of virulence loci in *Shigella* and EIEC. PCR assays were used to detect the presence (+) or absence (-) of known *Shigella* virulence loci in *Shigella*, EIEC and phylogenetically related isolates.

Clonal Group	TW Number	Strain	Serotype	Locale	Year	Source	Virulence Loci					
							iutA	*iucD*	*pic*	*set1A*	*shuA*	*senA*
Shigella 1	01510	4444-74	B2	USA (Idaho)	1974	CDC	+	+	-	-	-	+
	01154	3594-74	B4	USA (Colorado)	1974	CDC	+	+	-	-	-	+
	07576	1043-82	F6	USA (Colorado)	1982	CDC	+	+	-	-	-	+
	07573	1485-50	F6	USA (Michigan)	1980	CDC	+	+	-	-	-	+
	07572	3138-88	F6	USA (Massachusetts)	1988	CDC	+	+	-	-	-	+
	01142	2770-51	B14	USA (California)	1951	CDC	+	+	-	-	-	+
	01503	225-75	D3	India (Bombay)	1975	CDC	+	+	-	-	-	+
	01506	3341-55	D12	USA (Arizona)	1955	CDC	+	+	-	-	-	+
	01507	3470-56	D7	No Data	1956	CDC	+	+	-	-	-	+
	01504	2415-49	D3		1949	CDC	+	+	-	-	-	+
	08830	K-66	F6	Bangladesh		Talukder	+	+	-	-	-	+
	08831	K-313	F6	Bangladesh		Talukder	+	+	-	-	-	+
	08835	K-730	B	Bangladesh		Talukder	+	+	-	-	-	+
	08836	K-2085	B	Bangladesh		Talukder	+	+	-	-	-	+
	07585	2054-75	D4	USA (Massachusetts)	1975	CDC	+	+	-	-	-	+
Shigella 2	01175	965-58	B15	USA (Minnesota)	1958	CDC	+	+	-	-	-	+
	01155	3615-53	B17	Vietnam (Hanoi)	1953	CDC	+	+	-	-	-	+
	02615	155-74	D2	USA (California)	1974	CDC	+	+	-	-	-	+
	01151	3408-67	B5	USA (Maryland)	1967	CDC	+	+	-	-	-	+
	01146	291-75	B9	USA (California)	1975	CDC	+	+	-	-	-	+
	01162	5254-60	B11	Antilles	1960	CDC	+	+	-	-	-	+
	07547	5216-82	B17	Bulgaria	1963	CDC	+	+	-	-	-	+
	07550	513-84	B15			CDC	+	+	-	-	-	+
Shigella 3	02622	2702-71	F1	USA (Montana)	1971	CDC	+	+	+	+	-	+
	06299	2457T	F2A	No Data		CVD	+	+	+	+	-	+
	02623	2747-71	F2A	USA (California)	1971	CDC	+	+	+	+	-	+
	01143	2794-71	F5	USA (California)	1971	CDC	+	+	+	-	-	+
	01130	1170-74	F5	USA (Massachusetts)	1974	CDC	+	+	+	+	-	+
	08837	SA100	F2a			Payne	+	+	+	+	-	+
	01149	3226-85	SS	USA (Oklahoma)	1985	CDC	+	+	-	-	-	+
	08828	K-482	F1c	Bangladesh		Talukder	+	+	-	-	-	+
	08833	K-147	F4	Bangladesh		Talukder	+	+	-	-	-	+
	07554	3390-91	B12	USA (Florida)	1991	CDC	+	+	-	-	-	-
	01144	2850-71	F3a	USA (New Jersey)	1971	CDC	+	+	-	-	-	-

Table 8 (continued).

Clonal Group	TW Number	Strain	Serotype	Locale	Year	Source	Virulence Loci					
							iutA	*iucD*	*pic*	*set1A*	*shuA*	*senA*
Shigella other	08881	3556-77	B13			CDC	-	-	-	-	+	-
	08889	3054-94	B13			CDC	-	-	-	-	+	-
	02630	3823-69	D1	Guatemala	1969	CDC	-	-	-	-	+	+
	02609	1007-74	D1	USA (California)	1974	CDC	-	-	-	-	+	+
	02637	5514-56	D10	Rhodesia	1956	CDC	+	+	-	-	+	+
	01161	4822-66	SS	USA (Arizona)	1966	CDC	+	+	-	-	-	+
	01150	3233-85	SS	USA (Florida)	1985	CDC	+	+	-	-	-	+
	08839	12032	B13			ATCC	+	+	-	-	-	-
	08891	3097-02	Unknown			CDC	+	+	+	+	-	-
	07625	Albert 10457		USA (California)		Janda	-	-	-	-	-	-
	08884	2046-51	B13			CDC	-	-	-	-	-	-
EIEC 1	06117	LT-15	O28:H-	Brazil	1983	Trabulsi	-	-	+	+	-	+
	06129	LT-26	O28:H-	Japan	1978	Trabulsi	+	+	-	-	-	+
	01095	1886-77	O29:NM	No Data	1977	CDC	+	+	-	-	-	+
	06139	LT-41	O136:H-	Bangladesh	1983	Trabulsi	+	+	-	-	-	+
	06186	LT-91	O164:H-	Japan	1981	Trabulsi	+	+	-	-	-	+
EIEC 2	03204	1827-70	O29:H27	USA (Virginia)	1979	CDC	-	-	-	-	-	+
	01116	929-78	O124:H-	No Data	1978	CDC	+	+	+	+	-	+
	01110	5898-71	O124:H30	No Data	1971	CDC	+	+	-	-	-	+
	01096	202-72	O124:H30	No Data	1972	CDC	+	+	-	-	-	+
	06192	LT-99	O152:H-	Brazil	1968	Trabulsi	+	+	-	-	-	+
	06189	LT-94	O-:H-	Brazil	No Data	Trabulsi	-	-	+	+	-	+
	00073	5338-66	O111:H21	USA (New Jersey)	1966	CDC	+	+	-	-	-	-
	08882	5216-70	B13			CDC	+	+	-	-	-	+
EIEC other	06162	LT-68	O144:H-	Brazil	1984	Trabulsi	-	-	-	-	-	+
	03203	1758-70	O28:H21	USA (Tennessee)	1970	CDC	+	+	-	-	-	+
	06177	LT-82	O167:H-	Brazil	1981	Trabulsi	-	-	-	-	+	+
	06158	LT-62	O143:H-	Japan	1965	Trabulsi	+	+	-	-	+	+

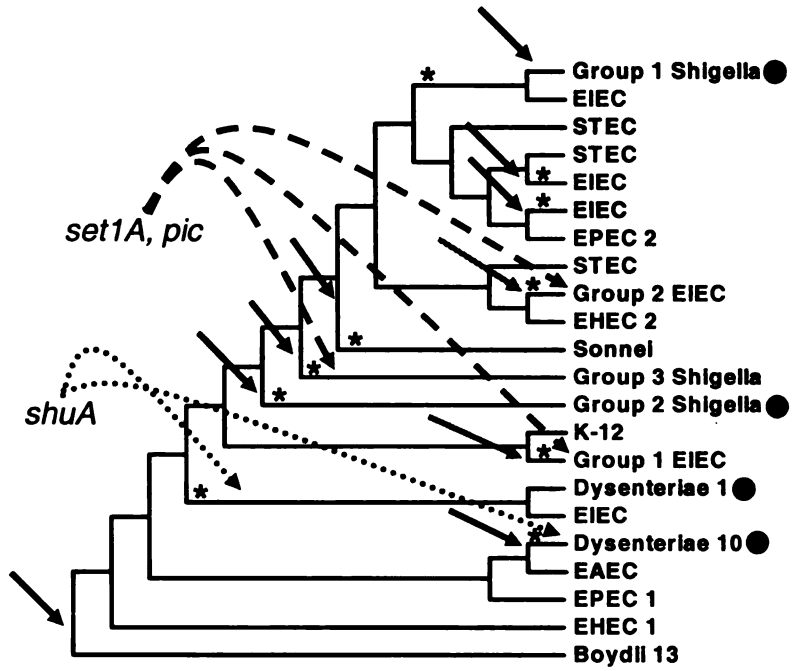

Figure 11. A phylogenetic perspective of gene acquitition and loss in invasive *E. coli*. The acquisition of pINV is indicated by an asterisk, *shuA* by arrows with dotted lines, *pic* and *set1A* by arrows with dashed lines, and *iutA* and *iucD* by arrows with solid lines. Lines with gray arrows indicate variable distribution of the indicated loci within the clonal group. The grey circles indicate variable loss of the *mtlD* locus among the Shigellae.

DISCUSSION

In this study, nucleotide sequencing of housekeeping loci was used to provide a phylogenetic framework for invasive *E. coli* and *Shigella* isolates. In agreement with previous studies (81, 87, 88), the findings presented here based on isolates and loci independent of the afore mentioned studies indicate that *Shigella* fall within the diversity of *E. coli* and should be reclassified as *Escherichia*. Within the phylogenetic framework, *Shigella* and the enteroinvasive *E. coli* pathovar have arisen independently at numerous times to form distinct phylogenetic groups. The serotypes included in each *Shigella* phylogenetic group are concordant with the results of Reeves and colleagues (88).

Multilocus sequencing and MLEE have proven to be robust approaches in determining phylogenetic relatedness. The goal of both methods is to distinguish many genotypes in which the variation accumulates slowly (28). In the case of MLEE, this is measured objectively by comparing the mobility of proteins in a gel against a known standard. As nucleotide sequencing has become less expensive and higher throughput, multilocus sequencing is now widely used to establish relationships as well as identify and type isolates involved in outbreaks of disease. In contrast to MLEE, multilocus sequencing can be applied directly to clinical material and there is no need to obtain reference isolates for comparision (28).

Multilocus sequencing has been used to investigate the genetic diversity within numerous populations of bacterial pathogens. A recent study by Adiri (2) examined the relationship of *E. coli* serotype O78 isolates based on six housekeeping loci. The results were able to show that the invasive isolates of this serotype clustered together regardless of the host organism. Examples of population genetic studies in other genera using

multilocus sequencing include measuring the diversity of *Neisseria gonorrhoeae* isolates using 18 loci (123), determining the clonality of *Staphylococcus aureus* by sequencing 7 housekeeping loci (33), and characterizing antibiotic resistant *Streptococcus pneumonia* isolates (106).

Other studies have sought to study the relationships of *E. coli*, *Salmonella,* and *Shigella* from an evolutionary perspective using alternative approaches. A study by Fukushima (36) used the nucleotide sequence of the B subunit of DNA gyrase (*gyrB*) as an alternative to 16S rRNA sequencing to establish a phylogeny. This approach proved useful for the differentiation of the closely related *Escherichia* isolates; however, only a minimal sample of *Shigella* isolates were included in the study. A PCR based primer – probe set method was used by Wang and coworkers (127) to first identify *Escherichia* and then differentiate the isolates based on the amplification of the *Shigella* virulence loci, *ipaH* and *set1A*. The *malB* locus used to detect *Escherichia* was unable to identify *S. boydii* and *S. dysenteriae*; however, the PCR assays were able to detect the virulence loci.

Mannitol. There are several metabolic traits in *Shigella* that appear to have been lost in parallel at multiple times in the divergence of invasive clones. It is suspected that some of these loss-of-function phenotypes will be a result of major deletions as found with the lysine decarboxylase regions of Flexneri (67). For example, Dysenteriae isolates often do not utilize mannitol (26). The mannitol operon contains three loci involved in the utilization of mannitol; *mtlA*, which encodes a mannitol permease, *mtlD*, encoding mannitol-1-phosphate dehydrogenase, and *mtlR*, which encodes a repressor. The results

presented here identified the loss of loci and insertions in this operon which offers a genetic explanation for the previously observed phenotype.

Interestingly, early studies in *S. flexneri* 2A identified the *arg – mlt* chromosomal region to be necessary for fluid accumulation in rabbit ileal loop assays. When this region was replaced by the homologous *E. coli* K-12 region, the *S. flexneri* recipient became Sereny-negative (103). It is suggested that this region is involved in the production of a Shiga-like enterotoxin (45). Another study showed the incorporation of the *E. coli* chromosomal region bounded by *xyl* and *rha* (which includes *mtl*) into a *S. flexneri* 2A background led to a loss of fatal infection in the starved guinea pig model (34). This construct maintained the ability to invade cultured mammalian cells and elicit an inflammatory response in rabbit ileal loop assays. Although this region does not hamper invasiveness, it may play a role in bacterial survival after entry into the host cell. Genes within this region encode the aerobactin binding protein and receptor (40) in *S. flexneri* 2A and the structural genes for Shiga toxin in *S. dysenteriae* type 1(104). These reports suggest the possibility of a selectively advantageous deletion or black hole in this region of the chromosome. A recent report by Talukder (114) identified a subgroup of atypical *S. flexneri* serotype 4 isolates that are also mannitol negative.

Acquisition of virulence loci. There are several groups of chromosomal genes that have been implicated in virulence of *Shigella* strains and have the characteristics of pathogenicity islands. These include 3 major islands, SHI-1, SHI-2, and SHI-3: SHI-1 encodes a ShET1 enterotoxin (*set1*), autotransporter protease (*sigA*), and mucinase (*pic*, formerly known as *she*) and occurs in Flexneri 2A strains. SHI-2, which encodes an iron acquisition system and several other proteins is inserted near the *selC* locus of *S. flexneri*

(73). Parts of the SHI-2 have been detected in other *Shigella* strains (124). SHI-3, discovered in a Boydii 5 strain, contains genes encoding the synthesis and transport of aerobactin and is present at the *pheU* tRNA locus in some *S. boydii* isolates but not in others (89).

The occurrence and distribution of the *Shigella* islands and virulence genes has been examined based on species isolates but has not been studied from an evolutionary perspective. A study by Purdy et al. used PCR assays to examine the distribution of SHI-2 and SHI-3 along with known integration sites and reported the results using the traditional species classification (89). By examining the results using a phylogenetic perspective, it is suspected that SHI-2 occurs in Reeves Group 3 in the *selC* site, SHI-3 occurs in Group 2 and some Group 3 strains in the *pheU* site, and neither island occurs in Group 1 strains. SHI-2 is also found in D1, and both SHI-2 and SHI-3 appear to be in Sonnei. The results presented here are not in complete agreement with the previous report. It appears that in addition to the SHI-2 and SHI-3 islands a third unrecognized island containing the *iucD* and *iutA* loci may be unique to the Group 1 Shigellae. The loci indicative of these islands were not amplified in the D1 isolates of this study which may be due to the natural variability of the isolates.

Two additional studies examined the distribution of the SHI-1 and SHI-4 islands using a species based approach. Al-Hasani et al. (5) used PCR to detect *sigA* and *pic* in enteropathogens. When placed in a phylogenetic perspective, the results of this study found *sigA* to have a wider presence among the shigellae, whereas, *pic* appears to be localized to the Group 3 *Shigella*. The molecular epidemiology of the SHI-4 island was investigated by Turner an colleagues (122). PCR amplification of three marker loci was

used to screen for the SRL island which appeared to be widespread among the clonal groups. Because these isolates were initially selected for multiple antibiotic resistances (122), it is possible that a bias may occur as only a portion of the population is then surveyed. Runyen-Janecky et al. (98) identified an iron acquisition locus, *sit,* and found the distribution to be widespread among the shigellae and EIEC. It is suggested that this locus may be located on a previously unidentified pathogenicity island (98).

A report by Wyckoff (133) showed that two EIEC serotypes (O136 and O143) hybridized to a probe for the heme binding locus, *shuA*. Results from the current study show that an additional serotype, O167 and Dysenteriae type 10 also harbor the *shuA* gene. There are few additional reports on the distribution of virulence elements in the EIEC groups; however, the results from this current study have begun to elucidate the molecular similarities between the EIEC and *Shigella*. It appears that the EIEC isolates stably maintain the large virulence plasmid conferring the invasive phenotype. The acquisition of chromosomal and pathogenicity island associated virulence loci is much more variable within the EIEC isolates. Because of this, it is more likely that these loci have been transferred recently in the evolutionary history of this pathovar by horizontal exchange.

A recent report by Talukder and colleagues (115) also provides evidence for the initial acquisition of the plasmid followed by later horizontal transfer events to obtain additional virulence loci. In a recently emerged Flexneri serotype, 1C, Talukder found that all of these isolates were positive for *Shigella* enterotoxin 2 (ShET-2 or *senA*) encoded by the large virulence plasmid while none were positive for *Shigella* enterotoxin 1 (ShET-1 or *set1*) which is encoded by a pathogenicity island (115). The results from

this study corroborate this finding as *senA* is present in all of the Group 3 *Shigella*. The distribution of the *set1A* locus suggests lateral transfer could be responsible for the inconsistent dispersal within the group.

The identification and sequencing of *set1A, senA, pic, iutA, iucD,* and *shuA* in this report expands to include additional *Shigella* serotypes and the EIEC pathovar to identify occurrences of virulence gene acquisition and patterns of distribution within the evolutionary framework provided by the housekeeping loci. This information allows for the development of a parsimonious and testable model for the acquisition of the mobile elements as demonstrated in Figure 11. The results of this study identify the acquisition of the large virulence plasmid at least ten times in the evolutionary history of *E. coli*. Early studies by Hale (46) and Sansonetti (101) determined that the plasmids were derived from a common ancestor; however they had evolved independently by accumulating mutations in restriction sites. It is also possible that the plasmid was present in a common *E. coli/Shigella* ancestor but could not be maintained in the lineages that gave rise to the other *E. coli* pathovars. At least two pathogenicity islands harboring an aerobactin operon have been acquired at least eight times. It is probable that this is a conservative estimate as the loci are variably distributed in additional invasive clones and these lineages may be incurring either loss or gain. The SHI-1 island appears to be a stably acquired element in the Group 3 *Shigella* whereas, it is intermittently gained among the EIEC. The chromosomal loci, *shuA* appears in two lineages and the gain is most likely attributable to horizontal transfer as homologs of this locus are found among pathogenic *E. coli* isolates (133). Gene loss has occurred in the case of mannitol dehydrogenase at four times in the evolutionary history of shigellae. Because this loss

occurs numerous times within an operon of housekeeping function, it seems likely that there is a selective advantage to inactivating the metabolic pathway.

ACKNOWLEDGEMENTS

I thank N. Strockbine from the Centers for Disease Control and Prevention, K. Talukder from the International Centre for Diarrhoeal Diseases Research, Bangladesh, and L. Trabulsi for providing bacterial isolates used in this study. D. Lacher, T. Large, and L. Ouellette were instrumental in providing sequence data for additional pathovars and reference isolates. This work will be submitted for publication with the incorporation of data from K. Hyma's analysis of unclassified *Shigella* serotypes. The nucleotide sequences of both housekeeping and virulence loci in these invasive isolates will be submitted to GenBank for deposition.

CHAPTER 3

PHENOTYPIC VIRULENCE CHARACTERISTICS OF *SHIGELLA* AND ENTEROINVASIVE *ESCHERICHIA COLI*

SUMMARY

The virulence attributes of 42 *Escherichia coli* and *Shigella* isolates were assessed using *in vitro* cell culture assays. Adherence, invasion, and intracellular multiplication were measured in HEp-2 and Henle 407 cell lines using gentamicin protection assays. Plaque assays were used to assess the ability of the bacteria to spread to neighboring eukaryotic cells. Bacterial adhesion and invasion were variable among the identified clonal lineages as well as between eukaryotic cell lines. Overall, the Group 3 *Shigella* had the highest average invasion in HEp-2 cells, whereas, invasion by the Group 2 *Shigella* isolates was slightly higher in Henle 407 cells. Statistical analysis of the data determined that there was not a significant difference in the invasiveness in HEp-2 cells between the three main *Shigella* groups or between the *Shigella* and EIEC. Intracellular multiplication was measured for a subset of the isolates over a 10 hour time course in Henle 407 cells. One isolate, SA100 displayed evidence of replication over the course of the assay; however, little or no intracellular multiplication occurred in the other bacterial isolates. Five isolates were able to form plaques in Henle 407 monolayers indicating the ability to spread to adjacent cells. Four of these five were from recent clinical cases of diarrheal disease in Bangladesh. Due to the variability of the results, it is not apparent from this study that invasive clonal groups are an accurate predictor of virulence phenotype.

INTRODUCTION

Shigellae and enteroinvasive *Escherichia coli* have a characteristic form of pathogenesis involving invasion of the mucosal epithelial cells of the large intestines. The molecular and cellular events underlying epithelial cell invasion by Shigellae have been intensively studied and reviewed (35, 44, 83, 100, 102). Briefly, invasion occurs via bacterium-directed phagocytosis with the major events as follows: contact of bacteria with the surface of the epithelial cell induces rearrangements of the cyto-skeleton, local membrane ruffling, and uptake of the bacteria (17). Inside the cell, the bacteria escape from the endosomal vacuole by lysing the membrane, enter the cytoplasm, and multiply there. The intracellular bacteria move through the cytoplasm by polymerizing actin filaments. This movement results in protrusions from an infected cell's membrane that contains bacterial cells at the tip, which can then be engulfed by adjacent cells. In this way, the invasive bacteria can multiply and spread from cell-to-cell without being exposed to the extracellular environment.

The components underlying the invasive phenotype are encoded on a large (~200 kb) pINV plasmid. The pINV plasmids vary in size and composition, but in general, they include an entry region, containing 35 genes organized into at least 4 transcriptional units (83). These include the secretory machinery, secreted proteins, molecular chaperones, and regulators encoded by *virB-ipgD*, *icsB-mxiE*, *mxiM-spa13*, and *spa47-spa40*. The entry region genes are homologous to the genes of the SPI-1 island of *Salmonella* (37). The pINV plasmid also carries genes for actin-based motility of shigellae inside the cell, a variety of plasmid antigens, and other suspected virulence-related proteins. Although most of the research has been conducted with *S. flexneri,* it is clear that many of the genes

on pINV are critical to cell invasion and are required for full virulence of enteroinvasive strains (48).

There are numerous characteristics of shigellae that suggest a differential ability to cause disease. Shigellae were originally characterized as a distinct genus from *Escherichia* due to a lack of biochemical traits. Even within the *Shigella* genus, further biochemical profiles are used to differentiate the isolates into the four recognized species. The species and serotypes causing the characteristic dysentery disease are also differentially distributed with regard to season (92), geographic location and socio-economic demography (120). Talukder and colleagues (115) have discovered that a recent Flexneri serotype, 1C, has increased in prevalence while serotype 1A has decreased in Bangladesh. It is possible that phage-mediated serotype conversion may have allowed for the evolution and recent spread of this new serotype. Although this new serotype is related to other Flexneri type 1 isolates, there are metabolic and plasmid differences that could impact virulence. Hsu (51) has shown that within the *Shigella* genus there are differences in the copy numbers of IS1 (insertion sequence 1) elements. Flexneri, Dysenteriae, and Sonnei have high numbers of this element in their genomes that could interrupt and inactivate the expression of housekeeping or virulence proteins. A recent study by Lan et al. (58) examined three loci on the 140 MDa pINV plasmid of *Shigella* and EIEC isolates and found that there are two distinct forms, pINVA and pINVB, based on nucleotide sequencing. This molecular evidence adds further support to the idea that not all invasive *E. coli* and *Shigella* harbor the same disease causing potential. Because infection with invasive pathogens is a multiple step process, it is probable that the population harbors isolates that are retarded at any stage of

pathogenesis: attachment, invasive ability, intracellular multiplication, spread to adjacent cells, survival in the host environment or survival in the natural environment.

Numerous models have been implemented to study invasion by bacterial pathogens. In cell culture assays, susceptible host cells are infected with bacteria for a specified period of time. Typically, gentamicin is then added to kill any extracellular bacteria. Because gentamicin does not cross the eukaryotic membrane, bacteria that are able to invade are protected from the antibiotic. Gentamicin is then removed and the host cells are lysed to determine the number of bacteria able to invade. This assay can be adapted to measure adherence as well as intracellular multiplication. Plaque assays described by Oaks (78) measure the virulence phenotype from the early stages of adhesion to the later stages of spread to adjacent cells. An additional animal model also exists for the measurement of invasiveness. Sereny tests are used to assess virulence by testing for the induction of keratoconjunctivitis in guinea pigs.

It is the goal of this study to examine differences in virulence characteristics; specifically, attachment, invasion, intracellular multiplication and spread, in a population of EIEC and *Shigella* isolates. Many of these isolates were used in the previous study to identify distinct clonal lineages of invasive bacteria. Previously described cell culture assays were used to investigate differences both within and between the identified clonal groups as well as potential differences in host eukaryotic cell lines.

MATERIALS AND METHODS

Bacterial isolates. Forty-two isolates representing each of the Reeves groups, EIEC, EPEC, EHEC, EAEC, and non-pathogenic *E. coli* (Table 9) were chosen to assess relative virulence by measuring invasive ability, intracellular multiplication, and spread to adjacent cells. Isolates were grown overnight from freezer stocks in 10 mL of Tryptic Soy broth at 37°C with shaking.

Microbial inhibitory concentration (MIC) of gentamicin. All bacterial isolates were assessed for sensitivity to gentamicin. 250 μl aliquots of tryptic soy broth containing gentamicin at concentrations ranging from 0 μg/ml to 70 μg/ml were inoculated with approximately 10 μl of 10^7 CFU/ml of bacteria. The cultures were incubated overnight at 37°C and examined for growth inhibition.

Eukaryotic cell lines. Monolayers of HEp-2 (ATCC CCL-23) and Henle 407 (ATCC CCL-6) cells were used for invasion assays. Only Henle 407 cells were used in the intracellular multiplication and plaque assays. HEp-2 cells were maintained in the laboratory in minimal essential media (MEM) supplemented with 5% fetal bovine serum. Henle 407 cells were maintained in Henle media supplemented with 10% fetal bovine serum as described by Reeves (93). Both cell lines were grown in a 5% CO_2 atmosphere at 37°C.

Adherence, invasion and intracellular multiplication assays. Tissue-culture invasion assays were performed using an adapted procedure from Donnenberg (22) as follows. All liquid media was pre-warmed in a 37°C water bath. HEp-2 and Henle 407 cells at a density of 10^5 cells/ml were added to 24-well microtiter plates (Costar 3526) and incubated overnight at 37°C in 5% CO_2. After the incubation, the cells were washed

Table 9. Isolates assayed for phenotypic virulence attributes. The subset of the isolates used in intracellular multiplication and plaque assay are indicated by an asterisk (*).

Clonal Group	TW Number	Strain	Serotype	Locale	Year	Source
Shigella 1	01154	3594-74	B4	USA (Colorado)	1974	CDC
	01503	225-75*	D3	India (Bombay)	1975	CDC
	01504	2415-49*	D3		1949	CDC
	08830	K-66*	F6	Bangladesh		Talukder
	08831	K-313	F6	Bangladesh		Talukder
	08835	K-730	B	Bangladesh		Talukder
	08836	K-2085*	B	Bangladesh		Talukder
	07585	2054-75	D4	USA (Massachusetts)	1975	CDC
Shigella 2	01175	965-58	B15	USA (Minnesota)	1958	CDC
	01155	3615-53*	B17	Vietnam (Hanoi)	1953	CDC
	02615	155-74	D2	USA (California)	1974	CDC
	01151	3408-67	B5	USA (Maryland)	1967	CDC
	01146	291-75*	B9	USA (California)	1975	CDC
	01162	5254-60	B11	Antilles	1960	CDC
	07547	5216-82	B17	Bulgaria	1963	CDC
	07550	513-84*	B15			CDC
Shigella 3	06299	2457T	F2A	No Data		CVD
	01143	2794-71	F5	USA (California)	1971	CDC
	01130	1170-74	F5	USA (Massachusetts)	1974	CDC
	08837	SA100*	F2a			Payne
	08828	K-482*	F1c	Bangladesh		Talukder
	08833	K-147*	F4	Bangladesh		Talukder
	07554	3390-91	B12	USA (Florida)	1991	CDC
	01144	2850-71	F3a	USA (New Jersey)	1971	CDC

Table 9 (continued).

Clonal Group	TW Number	Strain	Serotype	Locale	Year	Source
Shigella other	02637	5514-56*	D10	Rhodesia	1956	CDC
	01161	4822-66	SS	USA (Arizona)	1966	CDC
	01150	3233-85	SS	USA (Florida)	1985	CDC
	08839	12032*	B13			ATCC
	08881	3556-77	B13			CDC
	08891	3097-02	Unknown			CDC
EIEC 1	06117	LT-15*	O28:H-	Brazil	1983	Trabulsi
	06139	LT-41*	O136:H-	Bangladesh	1983	Trabulsi
	06186	LT-91	O164:H-	Japan	1981	Trabulsi
EIEC 2	01096	202-72*	O124:H30	No Data	1972	CDC
	06189	LT-94*	O-:H-	Brazil	No Data	Trabulsi
EIEC other	06162	LT-68	O144:H-	Brazil	1984	Trabulsi
	03203	1758-70	O28:H21	USA (Tennessee)	1970	CDC
	06177	LT-82	O167:H-	Brazil	1981	Trabulsi
Reference E. coli						
	06375	E2348/69	O127:H6	United Kingdom (Taunton)	1969	Donnenberg
	02302	EDL933	O157:H7	USA (Oregon)	1982	O'Brien
	08017	K-12				ATCC
	04393	O42	O44:H18	Peru	1983	Nataro

with PBS (pH 7.4) and covered in MEM without supplements or Henle 407 media without supplements for HEp-2 and Henle 407 cells respectively. The cells were infected with 10 μl of 7×10^7 CFU of bacteria from overnight cultures. The tissue culture plates were centrifuged at 1000 rpm for 10 minutes to allow for contact of the bacteria to the monolayers and incubated at 37°C in 5% CO_2 for 1 hour. The infected cells were washed three times with PBS and lysed in 0.1% Triton X-100 in PBS for 20 minutes. A dilution of the lysate was plated on Tryptic Soy Agar to determine levels of bacterial attachment. To measure invasion, the infected monolayers were washed three times in PBS and incubated for 1 hr in MEM or Henle 407 media containing 50 µg/ml of gentamicin. The monolayers of cells were then washed 4 times with PBS and lysed with 0.1% Triton X-100 in PBS for 20 minutes and plated on Tryptic Soy Agar to determine the number of surviving bacteria. All lysates were plated using an Autoplate 4000 (Spiral Biotech) and incubated overnight at 37°C. Colonies were counted using the Spiral Biotech Q-count. Additionally, some experiments were run in parallel for visualization using Giemsa stain. The procedure was performed as indicated above; however, the infected cells were fixed for staining after the final wash.

Intracellular multiplication assays were performed with Henle 407 cells using the protocol described above with minor modifications. After the 1 hour treatment with gentamicin, the infected cells were washed and lysed in 2 hour intervals over the course of 10 hours. Cell lysates were plated and quantified as described above. A subset of 16 invasive isolates was chosen for the time course assay.

Plaque assays. Plaque assays using Henle 407 cells were performed according to the protocols of Oaks et al. (78) and Hong and Payne (50). Cells were maintained at

37°C in 5% CO_2 in a humid environment and grown to confluency in 35-mm plates. Overnight cultures of bacteria were subcultured and grown at 37°C with shaking until reaching mid-log phase. The monolayers of cells were washed with PBS and covered in Henle 407 media without supplements. The cells were infected with dilutions of 10^2, 10^3, and 10^4 bacteria and the culture plates were centrifuged at 1000 rpm for 10 minutes followed by incubation at 37°C for 90 minutes in 5% CO_2. The infected cells were washed 4 times with PBS and an agarose overlay of Henle 407 media containing 10% fetal bovine serum, 20% glucose, 20 μg/ml of gentamicin, and 0.5% agarose was added to each well. The plates were incubated at 37°C in 5% CO_2 in a humid environment for 24 – 48 hrs. Plaques were visualized using a second agarose overlay with a final concentration of 0.01% neutral red. All 16 isolates tested for intracellular multiplication were tested in the plaque assays.

Experimental design. Bacterial adherence was determined as an average of two samples from a single well. Invasion studies were performed in replicate with duplicate samples from each well. Intracellular multiplication was measured with four replicate wells at each time point. Plaque formation was determined from the average of three wells inoculated with a serial dilution of the stock titer. Initial CFU/ml was determined by plating a dilution of the bacterial stock used to infect the eukaryotic cells.

Statistical analysis. All virulence assays included a standard, SA100 (F2A) (84), to which all measurements were calibrated. Statistical analyses were performed with Systat 5.05 (SPSS, Inc.).

RESULTS

Gentamicin resistance. All isolates were tested for gentamicin resistance in broth containing a final concentration of gentamicin ranging from 10 μg/ml to 70 μg/ml. Broth containing no antibiotic was used as a control. The results of overnight growth at 37°C indicated that all control isolates (no antibiotic treatment) were viable and 40 of the 42 isolates were sensitive in media containing a final concentration of antibiotic of 10 μg/ml. Two isolates, 2415-49 and 5514-56, were resistant to the antibiotic with 2415-49 growing at all concentrations tested (Figure 12) and 5514-56 having noticeable growth in media with gentamicin at a final concentration of 10 μg/ml. The assay was repeated to confirm the resistance of both isolates.

Bacterial adherence. In order for the bacteria to cause a successful infection, they must be able to adhere to eukaryotic cells for uptake to occur. Adherence was measured for all isolates in this study for both HEp-2 and Henle 407 cell lines. The number of adherent bacteria ranged from 1.12×10^6 CFU/ml (isolate 2415-49) to 6.70×10^4 CFU/ml (isolate 12032) with HEp-2 cells and 4.75×10^5 CFU/ml (isolate 5254-60) to 1.63×10^4 CFU/ml (isolate LT-94) with Henle 407 cells. Additionally, wells containing only cell culture media were used to measure the background adherence to the culture well surface. The measures of background adherence to the culture wells were surprisingly high ranging from 9.07×10^5 CFU/ml (2415-49) to 6.00×10^4 CFU/ml (12032) in HEp-2 cell assays and 8.71×10^5 CFU/ml (2415-49) to 1.72×10^4 CFU/ml (EDL-933) in Henle 407 cell assays.

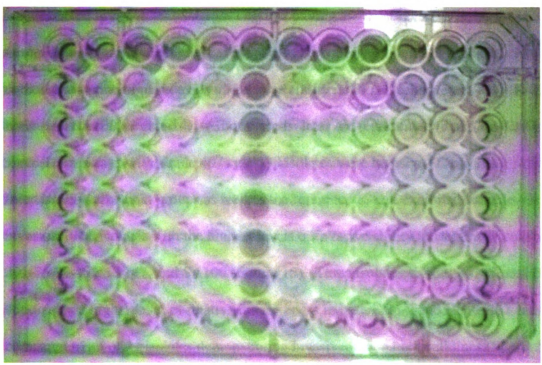

Figure 12. Image of gentamicin microbial inhibition assay. The 96 well trays are filled from top to bottom with increasing antibiotic concentrations (0 to 70 μg/ml) and left to right with bacterial isolates. This image shows the gentamicin resistant isolate (2415-49) in lane 6.

Eukaryotic cell invasion. Both HEp-2 and Henle 407 cell lines were used to measure the invasive ability of all isolates. SA100, a Flexneri 2A isolate was used as a standard and all measurements were standardized to this control for comparative purposes (Figure 13). In the case of the HEp-2 cells, 21 isolates were invasive (Table 10). The analysis of invasion with respect to clonal groups shows all groups of EIEC isolates being invasive and *Shigella* Group 1 having the most invasive isolates. The Group 3 *Shigella* has the highest average invasion of any group. The isolates demonstrating the highest levels of invasive ability in HEp-2 cells were K-482, K-2085, K-66, and SA100 (Figure 14). All of these isolates are representative of *Shigella* Groups 1 and 3. In assays with Henle 407 cells, 39 isolates were invasive (Table 10). Interestingly, all clonal groups as well as the reference *E. coli* exhibited some level of invasion. The isolates with the highest levels of invasive ability in Henle 407 cells were 291-75, K-482, K-147, and SA100 (Figure 15). Three isolates K-482, K-147, and 291-75, display a higher level of relative invasiveness in both HEp-2 and Henle 407 cells as compared to all other isolates demonstrating invasiveness in both cell lines (Figure 16).

Stastical comparison of HEp-2 invasion. In order to determine if the three main *Shigella* Reeves groups differ in the levels of invasiveness, the groups were compared by the Kruskal-Wallis non-parameter test for comparing means. The dependent variable in this analysis was the relative invasion measured in CFU/ml standardized to SA100, the positive invasion control strain. Each group of replicates was analyzed separately with the replicates being the average of the duplicate counts. This analysis determined that the *Shigella* groups do not differ significantly (Table 11) and therefore, they were pooled to be compared to the EIEC isolates. To examine if the *Shigella* isolates are more invasive

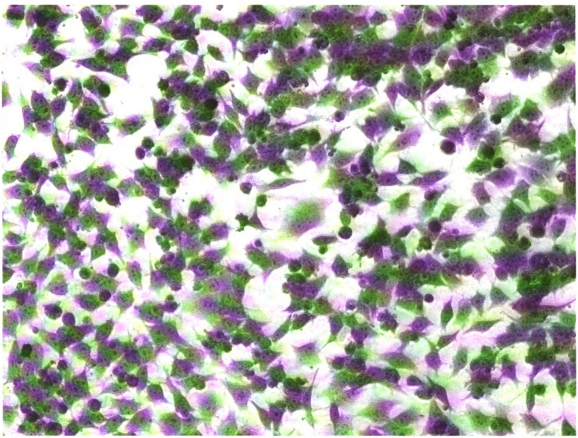

Figure 13. Photomicrograph of bacterial invasion by SA100 in Henle 407 cells. Giemsa staining is used to enhance visualization and differentiate the eukaryotic and bacterial cells.

Table 10. Summary of invasiveness as tested by gentamicin protection assays.

Clonal Group	Number of invasive isolates (total isolates)		Average invasion per invasive isolate (average invasion of group)	
	HEp-2	Henle 407	HEp-2	Henle 407
Shigella 1	7(8)	7(8)	67.02 (58.64)	22.07 (19.31)
Shigella 2	2(8)	7(8)	21.16 (5.29)	66.27 (57.99)
Shigella 3	4(8)	8(8)	205.26 (102.63)	57.72 (57.72)
Shigella Other	0(6)	5(6)	0 (0)	4.88 (4.06)
EIEC 1	3(3)	3(3)	31.45 (31.45)	11.55 (11.55)
EIEC 2	2(2)	2(2)	9.36 (9.36)	9.06 (9.06)
EIEC Other	2(3)	3(3)	22.41 (14.94)	7.34 (7.34)
Reference *E. coli*	1(4)	4(4)	0.14 (0.04)	6.11 (6.11)

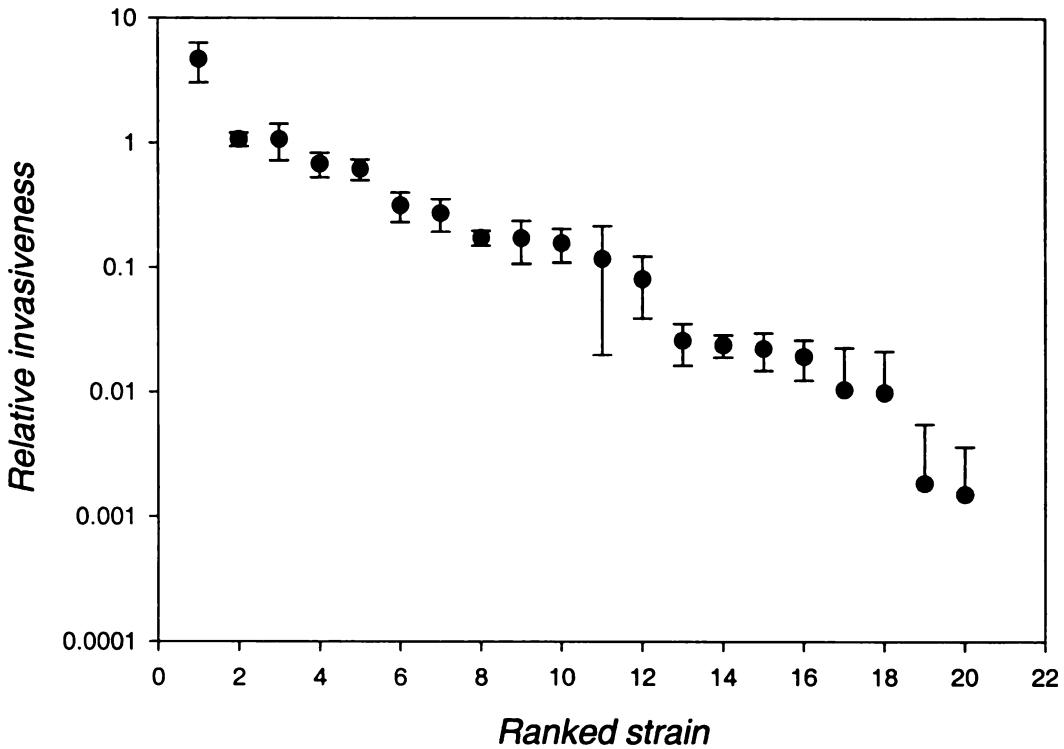

Figure 14. Relative invasiveness of *Shigella* and EIEC in HEp-2 cells. The strains are ranked from most invasive (K-482) to least invasive (225-75). SA100, the control isolate has a relative invasiveness equal to 1. Error bars indicate the standard error values calculated from the standard deviation divided by the square root of the sample size.

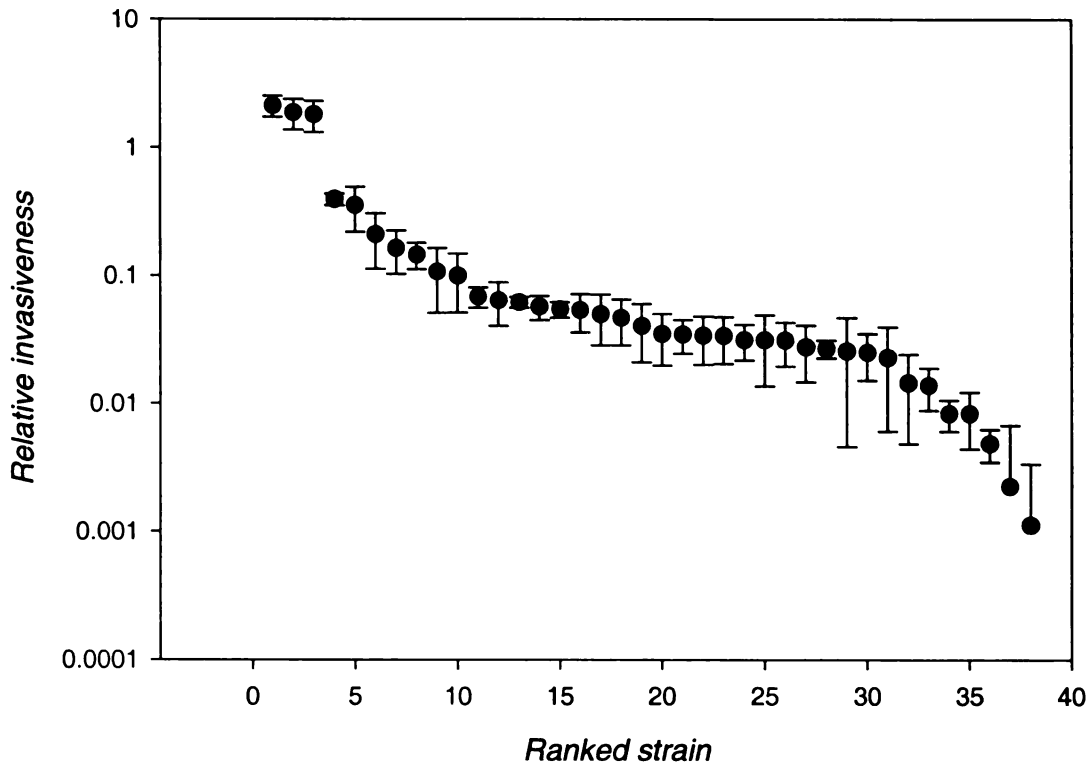

Figure 15. Relative invasiveness of *Shigella* and EIEC in Henle 407 cells. The strains are ranked from most invasive to least invasive. Three strains, K-147, 291-75, and K-482 have higher levels of invasiveness compared to the other strains tested. Error bars indicate the standard error values calculated from the standard deviation divided by the square root of the sample size.

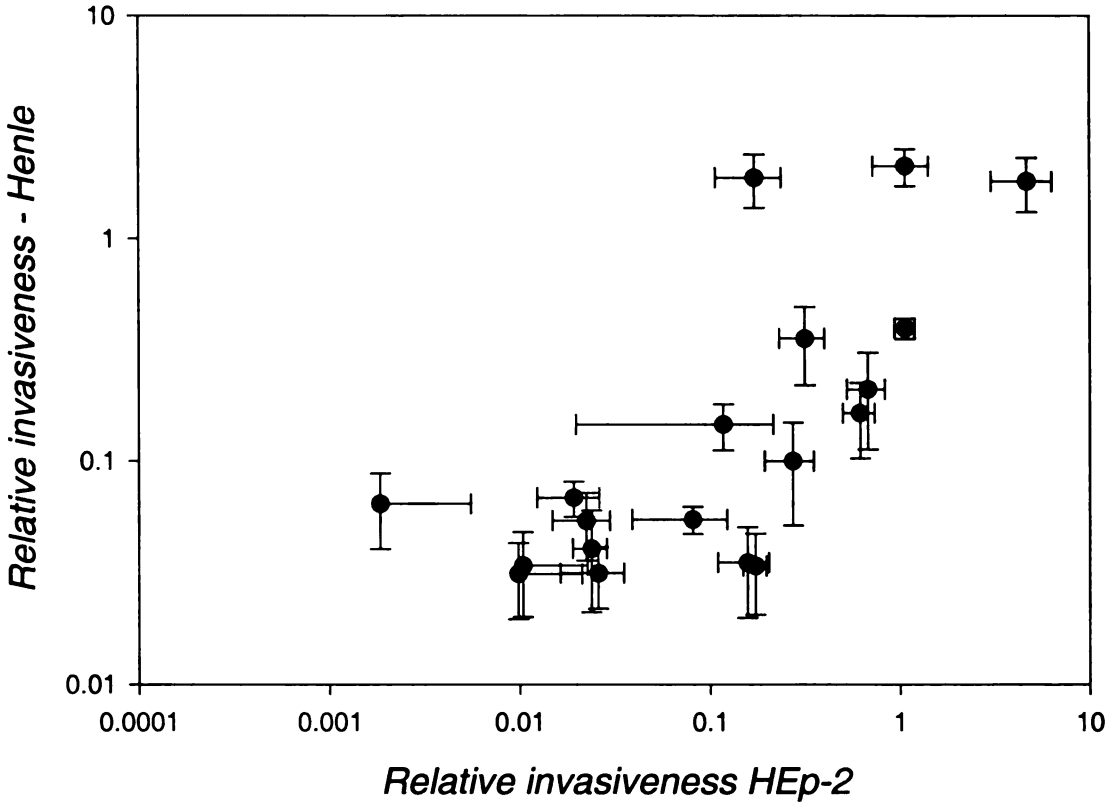

Figure 16. Correlation of invasiveness of *Shigella* and EIEC in HEp-2 and Henle 407 cells. Two isolates, K-482 and K-147, have relative invasiveness greater than 1 for both cell lines tested. Error bars indicate the standard error values calculated from the standard deviation divided by the square root of the sample size.

Table 11. Statistical comparison of HEp-2 invasion of the main Reeves groups. The analysis is based on the Kruskal-Wallis non-parameter test for comparing means.

Reeves Group	Replicate 1		Replicate 2	
	n	rank sum	*n*	rank sum
Shigella 1	8	122	8	123
Shigella 2	8	70	8	67
Shigella 3	7	84	7	86
	K-W 4.12, p = 0.127, 2df		K-W 4.80, p = 0.091, 2df	

than the EIEC isolates, a non-parametric test was used to compare the two means (Table 12). Again, there was no significant difference in the comparison of isolates.

Intracellular multiplication. Intracellular multiplication was measured in the Henle 407 cell line using a subset of the invasive isolates as well as the standard, SA100. The time course surveyed invasion followed by intracellular multiplication over the course of 10 hours. The results of this assay are summarized in Figure 17. Consistent with the invasion assay data, isolates K-2085, K-482, LT-41, K-147, and 291-75 have initial invasion levels greater than or equal to that of the SA100 standard. Over the time course, only SA100 is able to efficiently multiply with the peak growth occurring at 4 hours. Several of the isolates appear only to survive in the intracellular environment over the course of the assay, while others begin to show a decrease by the two hour time point.

Spread to adjacent cells. Another aspect of *Shigella* pathogenesis is measured by the ability to spread to adjacent cells. Plaque assays have been adapted to measure this phenomenon in infected cell culture assays. The sixteen isolates examined for the ability to multiply intracellularly were also examined for the ability to spread to neighboring Henle 407 cells. A clear plaque is formed when the cells lyse as can be seen in Figure 18. A summary of plaque formation results are provided in Table 13. Of the isolates tested, only five, including SA100, were able to form plaques indicating the spread to adjacent cells.

Table 12. Statistical comparison of HEp-2 invasion between *Shigella* and EIEC. The analysis is based on a non-parametric test for comparing two means.

	Replicate 1		Replicate 2	
	n	mean	*n*	mean
Shigella	23	348.5	23	363.5
EIEC	8	147.5	8	132.5
	Mann-Whitney U test is 72.5, p = 0.36		Mann-Whitney U test is 87.5, p = 0.83	

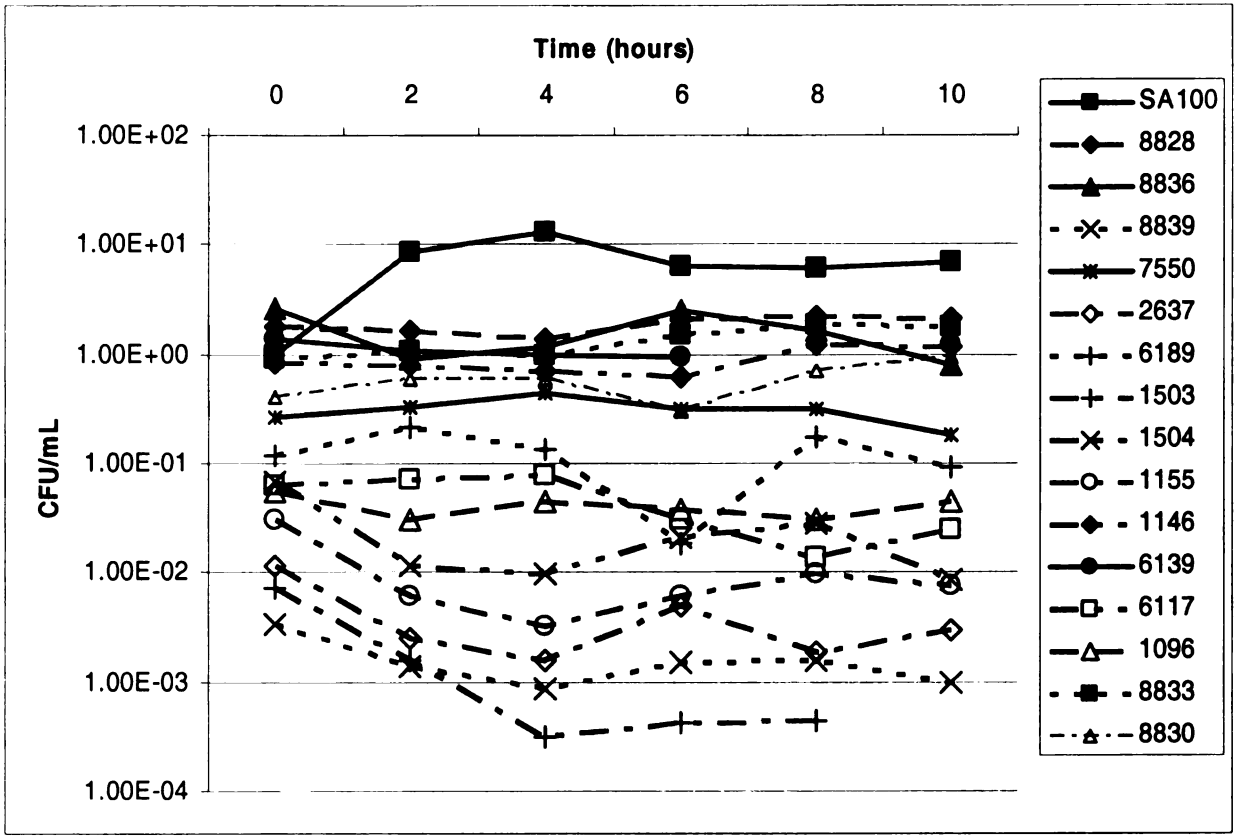

Figure 17. Plot of intracellular multiplication in Henle 407 cells over the course of 10 hours. The control isolate, SA100, shows peak intracellular growth at 4 hours.

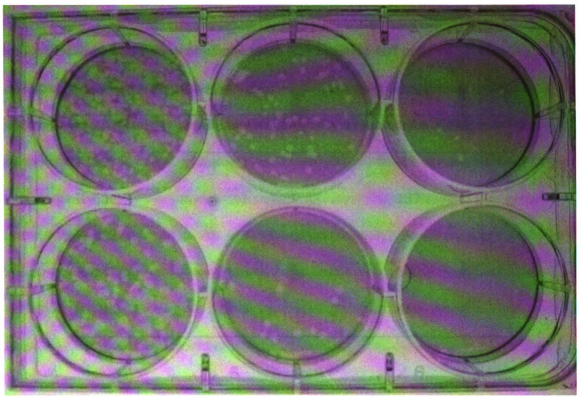

Figure 18. Plaque formation in Henle 407 cells. The top row of wells was infected with SA100 and the bottom row by K-482. An agar overlay containing 0.01% neutral red was added to visualize the plaques.

Table 13. Summary of plaque formation in Henle 407 cells.

Isolate	Clonal Group	Average number of plaques
K-2085	1	13
K-66	1	137
K-482	3	94
K-147	3	140
SA100	3	456

DISCUSSION

In vitro cell culture assays provide a method by which bacterial invasion can be readily studied in a laboratory setting. In addition, numerous stages of the invasion process can be monitored. This study used a population based sample to make inferences about the overall invasive phenotypes of *Shigella* and EIEC. The relationship of these invasive isolates was first determined by assignment to clonal groups based on the nucleotide sequence of housekeeping loci. Isolates were then screened for the presence of know virulence loci, some of which are known to be involved in invasion processes. The results from the phenotypic assays along with the known genotypes could provide insight into the necessary gene complement for optimal virulence.

Overall, the phenotypic assays demonstrated a wide range in virulence phenotype within and between clonal groups as well as between cell lines. The statistical analyses determined that there was not a significant difference in the invasiveness in HEp-2 cells between the three main *Shigella* groups or between the *Shigella* and EIEC. Previous cell culture studies with *Shigella* and other *E. coli* pathovars have noted differences in invasiveness depending on the type of cell line that was used. A study by Elsinghorst (27) using enterotoxigenic *E. coli* showed that these isolates would invade cells derived from various tissues; however, the isolates were most invasive for human ileocecum and colonic epithelial cells. In a study examining differences between EPEC and EIEC, Donnenberg (22) notes that these isolates are invasive in HeLa and Chinese hamster ovary cell lines. Sen (105) also reports variability in results from different cell lines. In comparison to HeLa cells, both Henle 407 and Hct 8 cells are better models for *Shigella*

invasion (105). Additionally, Henle 407 cells are Shiga toxin resistant and therefore useful for assays with *S. dysenteriae* (105).

Comparison of adherence and invasion results between laboratories is complicated by various means used to quantify these two virulence properties. In some cases, invasion is expressed by the numbers viable bacteria (11, 21, 22, 27, 95) and others by the number of infected eukaryotic cells (50, 86). When the numbers of viable bacteria are considered, subtle differences arise in the formulas used to calculate invasion. Donnenberg (21, 22) calculates invasion as the percent of original inoculum surviving gentamicin treatment. This percentage usually ranges from 0.5 to 25% (95). Robins-Browne (95) take adherence into account and calculates the proportion of cell-associated bacteria that survived antibiotic treatment. Rosa (97) proposes that invasion be measured as the percentage of intracellular bacteria divided by the number of extracellular bacteria plus the number of intracellular bacteria. Hong and Payne (50) along with Pope (86) calculate the percentage of eukaryotic cells infected by microscopic examination of 300 cells. A cell is considered infected when it contains 3 or more bacteria (50, 86).

Both intracellular multiplication and spread to neighboring cells can be measured in a more straightforward manner. In the case of intracellular multiplication, a peak of replicative growth can easily be determined from a plot of the data. In a study by Cersini et al. (13) the peak of growth occurs at the 3 hour time point. The samples from this study were measured at 2 and 4 hours with the peak of intracellular growth occurring at 4 hours. Plaque assays measure numerous aspects of virulence: attachment, internalization, escape from the phagosome, intracellular replication and spread to neighboring cells (78). The results from these assays can easily be quantified and used for comparative purposes.

Variation in invasive ability has been reported previously (11, 22, 27, 46). Elsinghorst (27) finds results of invasion studies to be variable on a daily basis; however, the controls are internally consistent. In an attempt to correct for the day to day variability, the results of this study were standardized to a control strain, SA100. Differences within the isolates themselves can contribute to the variability. Invasive strains that have mutations in or have lost the virulence plasmid typically can no longer invade (46). These mutants can be detected by a rough colony morphology (11).

A 1982 study by Bukholm (11) examines differences in invasive ability in *Shigella* species. In accordance with the results presented here, they report variability within species as well as within the serotype. The overall results demonstrated a small invasive potential for *S. dysenteriae* isolates with Sonnei showing the least amount of invasion. Bukholm (11) attributes some of this variation to differences between fresh isolates and stocks stored on agar. Loss of virulence in older cultures was also reported by Sansonetti (101). The fresh isolates obtained for this study from Talukder (113, 114) proved to be the most virulent as measured by these assays.

Because there is so much variability within each clonal group examined in this study, it is difficult to draw substantial conclusions regarding the impact of genetic complement on the virulence phenotype of a particular isolate. It is possible that this variation could be minimized by surveying strains recently isolated in disease cases or maintaining selective pressures for the inclusion of the large virulence plasmid in the laboratory.

ACKNOWLEDGEMENTS

The work presented in this chapter will be reviewed by S. Payne for her suggestions and comments before it is submitted for publication. I thank S. Payne and L. Wyckoff at the University of Texas at Austin and R. Binet at the Uniformed Services University for Health Sciences for their comments and suggested protocols. I also thank M. Saeed at Michigan State University for providing the HEp-2 cell line.

CHAPTER 4

METHODS TO COMPARE BACTERIAL GENOMES

ABSTRACT

Extensive genomic variation can exist within a species due to highly mobile elements such as pathogenicity islands, insertion sequences, and bacteriophage-mediated gene transfer. Because genome sequences are only available for only a limited number of representative bacterial isolates, methods are necessary to allow for genomic comparisons of additional isolates to their closely related reference isolate. This study examines two techniques that can allow for these comparisons to be made. Suppression subtractive hybridization (SSH) is a technique that has been employed in both eukaryotic and prokaryotic organisms to explore the differences in gene expression and genomic content. In the prokaryotic realm, this technique has the power to identify unique sequences that are present in one organism but absent in another. In contrast to SSH, paired end sequence mapping (PESM) a technique described in this report, allows for the comparison of more than two genomes and provides a location for the identified genomic fragment. PESM borrows the idea of scaffold building used in whole-genome shotgun sequencing. In this method, two ends of a fragmented, unknown genome are mapped to a sequenced reference genome in order to identify insertions and deletions in the genome. BLAST algorithms and database searches are then used to map and identify the resulting genomic fragments. Unique candidate genomic regions representing gene acquisition or loss were identified by SSH and PESM techniques. These regions were then screened by PCR methods and expanded to additional isolates to determine the extent of genomic change within the genus.

INTRODUCTION

The acquisition of new genes by horizontal transfer has played a major role in the adaptation and ecological specialization of bacterial lineages (61). It has been estimated, for example, that ~18% of the current genome of *Escherichia coli* K-12 represents foreign DNA acquired by horizontal transfers since the divergence of *E. coli* and *Salmonella enterica* (62). Gene acquisitions have also contributed to the variation in virulence among strains and closely related bacterial species (43, 96). In *E. coli* and *S. enterica*, blocks of virulence genes, called pathogenicity islands, have been acquired at different times, thus generating a variety of pathogens with distinct virulence genes and mechanisms of pathogenesis (41, 79, 80).

In the evolution of enteroinvasive *E. coli* and *Shigella*, gene acquisition has been important in two ways: first with the spread of the pINV plasmid that encodes invasive ability, and second with the presumed acquisition of a variety of mobile pathogenicity islands. In addition, there is growing evidence that gene loss has been important in adaptive radiation and the evolution of bacterial virulence. For example, Maurelli and coworkers (67) present evidence that the universal deletion of the lysine decarboxylase gene (*cadA*) has enhanced the virulence of *Shigella* species because cadaverine, a product of the reaction catalyzed by lysine decarboxylase, inhibits the activity of the *Shigella* enterotoxin. Maurelli and coworkers refer to such large, universal deletions that enhance virulence as "black holes" (67), the loss-of-function counterpart to pathogenicity islands.

Black hole formation is one example of pathogenicity-adaptive, or pathoadaptive, mutation (111). These genetic alterations represent a mechanism for enhancing bacterial virulence without horizontal transfer of specific virulence factors (111). Pathoadaptive

mutations include, for example, increases in bacterial virulence by random functional mutations in a commensal trait that are adaptive for a pathologic environment, such as that found for the FimH variants of uropathgenic *E. coli* (110).

Evidence for the formation of new black holes and novel islands will be investigated by developing a genomic method for finding major insertions and deletions. This method is based on the concept of paired-end sequencing and makes use of known genomic sequences. It is expected that the application of this method will provide insights into the genomic alterations and molecular adaptations that accompany the shift to intracellular invasion and multiplication.

An important concept in genome projects is called pairwise end sequencing (or paired end sequencing) in which nucleotide sequences are determined from both ends of random subclones derived from a DNA target. Overlapping end sequences are identified and grouped into contigs, and when a clone's paired ends fall in different contigs, the contigs can be connected together to form scaffolds (107). Here this idea is adapted, not for constructing scaffolds, but for discovering and mapping positions of major insertions and deletions (indels) in an unknown genome. This method will be referred to as paired end sequence mapping (PESM). PESM was used in this study to identify large insertions or deletions in the genome of a Dysenteriae type 1 isolate, 3823-69 by comparison of the fragmented Dysenteriae 1 genome to the published genomes of *E. coli* K-12 (9), EDL-933 (85) and *S. flexneri* 301 (54).

Another method that has been used to identify strain specific genomic regions is suppression subtractive hybridization (SSH), also referred to as genomic subtraction. SSH is a PCR-based technique that has been used in eukaryotic systems to identify

tissue-specific and differentially expressed genes (18). This method has also been applied to the study of prokaryotic systems. Due to the smaller and less complex nature of bacterial genomes, SSH can be used to identify unique genomic sequences among these organisms. The theory of the technique relies on selectively amplifying target fragments and suppressing non-target amplification. Two genomes are compared with one being referred to as the 'tester', or the genomic DNA of interest, and the 'driver', or the reference sample. The 'tester' and 'driver' DNA's are hybridized and the hybrid sequences are then removed leaving the 'tester' specific sequences.

In this study, SSH is used to compare a pathogenic O111:H21 *E. coli* clone, DEC6a, to the laboratory strain, K-12. These strains were chosen due to their close relationship determined by phylogenetic analysis as determined by Donnenberg (23) as well as the multilocus sequencing data presented earlier (Figure 3). In the clonal group of the EIEC, it is interesting that both K-12, a non-pathogen, and DEC6a, a causative agent of diarrheal disease, are so closely related to the invasive clones yet they lack the large invasion plasmid. Little is known about the virulence properties of the DEC6a pathogenic clone and there have been discrepancies in the literature as to whether this isolate belongs to the enteropathogenic *E. coli* (EPEC) or enteroaggregative *E. coli* (EAEC) pathovars (12, 132). In order to elucidate how this pathogen compares to other pathogenic and non-pathogenic *E. coli* isolates, a genomic approach was used to determine the genetic features that distinguish DEC6a from other *E. coli* isolates.

The purpose of this study is to identify genomic changes that are unique to Dysenteriae type 1 isolate, 3823-69, and an atypical EPEC isolate, DEC6a. Two methods are used to identify changes; one, a commercially available approach that allows for a

one way comparison of two isolates, and the other, a proposed technique that allows for multiple comparisons of an unknown isolate to the growing repertoire of completed genome sequences. By discovering the loss or acquisition of novel fragments using either approach, previously unidentified virulence factors can be elucidated that have allowed for the evolution of these two distinct pathogenic clones.

MATERIALS AND METHODS

PESM library construction. Because the library construction is a critical step in this method, the Lucigen Corporation (Middleton, Wisc.) was contracted to create a shotgun library of randomly sheared, end-repaired DNA from strain 3823-69 (TW02630), a *S. dysenteriae* D1. The library consists of at least 50,000 independent clones, which contain fractionated DNA size selected in the 8-12 kb range.

50 μg of high molecular weight 3823-69 genomic DNA was isolated and purified. Scientists at Lucigen randomly sheared 10 μg of the supplied genomic DNA (with a HydroShear instrument), end-repaired the sheared genomic DNA, and size selected the molecules by agarose gel electrophoresis to include 8-12 kb DNA and exclude other sizes. The size-selected DNA was then ligated to the gap-free cloning vector pSMART, which was then used to transform MC12 competent cells by electroporation. The pSMART vector does not use a promoter or indicator gene so there is no transcription either into or out of the insert DNA. This design reduces the cloning bias typical of conventional plasmid vectors. Scientists at Lucigen re-engineered the standard pSMART vector for the project to reduce the copy number and enhance cloning success of DNA in the 10 kb range (David Meade, President of Lucigen, personal communication). Plating of 50 μl of transformed cells yielded 416 colonies or 8.3×10^3 CFU/ml. An aliquot of 50 μl of transformed cells from self-ligated vector gave 30 colonies, representing 7.2% background empty vector.

The library supplied by Lucigen was plated in 25 μl aliquots onto TY^+ agar plates containing ampicillin at a final concentration of 100 μg/ml and incubated overnight at 37°C. Single colonies were selected from the plates and grown overnight at 37°C with

shaking in 4 ml of Terrific Broth containing ampicillin (100 μg/ml). One ml of the overnight culture was used for a freezer stock of each isolated clone. The remaining overnight culture was used for plasmid DNA preparation using either the QIAprep 8 Miniprep Kit or the QIAprep Spin Miniprep Kit with the procedures including the recommended steps for low-copy plasmids. A total of 652 single isolates were prepared with this method.

All plasmid DNA preparations were electrophoresed on a 0.8% agarose gel at 90V with a 1 kb ladder size marker to select the clones containing the largest inserts. Of the 652 clones, 136 were determined to have insert sizes equal to or greater than 8 kb. These clones were then digested overnight at 37°C with 20U of either *Eco*RI or *Eco*RV enzyme to determine a more precise size estimate of the insert. The digests were electrophoresed as described above with the addition of a λ*Hin*dIII ladder. Size estimates of the inserts were determined using the DNA ProScan software.

PESM analysis. A set of Perl scripts called PENDMAP ("pee-end-map") was developed for the purpose of analyzing data from the following type of experiment (65). Nucleotide sequences are determined for both ends of randomly cloned fragments (inserts) from an unknown genome, that is, a genome that has not been completely mapped and sequenced. The cloned fragments are size-selected to have a narrow distribution of length (average length *L* in bp; lower limit *L1*; upper limit, *L2*), for example in the 8-12 kb range. A number (*n*) of random fragments are chosen, the paired ends sequenced, and the end sequences are mapped to locations in a reference genome (a closely related, completely sequenced genome). Ends of length *k1* and *k2* are compared separately to the reference genome sequence by the BLAST algorithm. A threshold value

(z) of percent similarity is selected to determine an end match. (If the threshold value is set too low, ends can match to many genomic locations.)

For each set of paired ends there are four possible outcomes (Figure 19) as follows: Match (M), both ends have single map locations within a chromosomal distance (d) of $L1 < d < L2$; Partial match (P), one end maps to a single location, the other end does not match or matches to a location that gives a chromosomal distance between the ends of $<L1$, or $>L2$; No match (N), neither end matches to the reference genome. There is also the possibility of multiple matches, in which one or both ends map to more than one genomic location because of past gene duplications, the presence of multiple copies of genes, or mobile elements in the genome. Multiple matches are initially uninformative for major indel mapping purposes.

The interpretation of the paired end sequencing and mapping to a reference genome is illustrated in Figure 19. Matches are assumed to mark regions of the genome that are conserved. Partial matches can detect small insertions and deletions ($< L$) by deviations in chromosomal distances outside the distribution of fragment lengths, that is $d < L1$ or $d > L2$. Perhaps the most informative outcome is the partial matches caused by single end matches. This is shown in Figure 19 around a large insertion and deletion. These large insertions or deletions (length $>> L$) will be detected by a concentration of single end matches on the reference genome around the alteration. There will also be paired ends that do not match at either end because they lie within the major indel; these "no matches" are not initially informative about indel location.

The reference genomes used in this PESM analysis are: *E. coli* K-12 (9), EDL933 (85), and Sf301 (54).

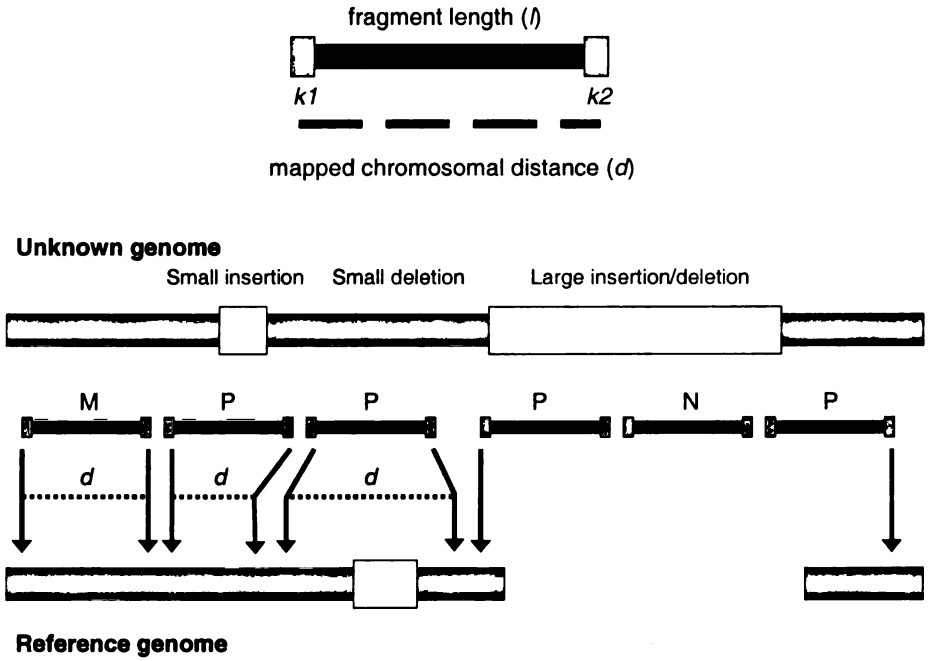

Figure 19. Diagram of Paired End Sequence Mapping (PESM). This technique can be used to detect large insertions or deletions in prokaryotic genomes. The top of the diagram shows a typical genomic fragment with ends *k1* and *k2*. The bottom illustrates the possible outcomes when comparing the unknown genome to a reference genome.

PCR analysis of PESM results. Clone 249 identified a potential black hole in the genome of 3823-69. This region was examined using a series of PCR assays with primers designed from the aligned genomic sequences of *E. coli* K-12 and EDL-933 (Table 14) to detect genes left intact and identify genes that may have been lost in Dysenteriae type 1. Additional *Shigella* and EIEC isolates were examined to determine the extent of gene loss among the invasive clonal groups. These isolates include: 1007-74 (D1), 2770-51 (B14), 3470-56 (D7), 5514-56 (D10), 4822-66 (SS), 2747-71 (F2A), LT-94 (O-:H-), and 202-72 (O124:H30) (Table 4 provides additional information for these isolates).

Long PCR of the *hca* region in *E. coli* and *Shigella* isolates. PCR primers were designed within the conserved flanking regions of the *hca* locus from the completed genomes of *E. coli* K-12 (9), EDL-933 (85), Sakai (49), CFT073 (130), and *S. flexneri* 2A strain 301 (54). Primers used were hca-F4 (5' - TTT CAT GGC ACG GGC AAC AGA ACC - 3') and hca-R7 (5' - ATG AAA CAG TGG GCG CAA GAG ATG G - 3').

Using Epicentre MasterAmp™ Extra-Long PCR cut, a PCR reaction was done using the nine MasterAmp Extra Long PCR 2x PreMixes and the Extra-Long DNA Polymerase Mix. 100 ng of a *S. dysenteriae* type 1 strain 3823-69, *E. coli* strains Sakai (O157:H7), and E2346/69 (O127:H6) were amplified using the following conditions. Denature at 98°C for 3 min, during which time a hot start was done adding the polymerase, followed by 28 cycles of 98°C for 30 sec, 63°C for 1 min, and 72°C for 17 min. A final step of 72°C for 30 min was used for completion of any partially extended product. Positive control, furnished by Epicentre is of a 20 kb region of lambda DNA. The negative control did not contain any template DNA. Strain 3823-69 amplified best

with premix 5, Sakai with premixes 5 & 6, and E2348/69 with premixes 4 and 6 with fainter bands with other premixes.

SSH molecular manipulations. *E. coli* K-12 (MG1688) and DEC6a (5338-66) were grown overnight in 100 ml of LB broth at 37°C. DNA was isolated from the cells using phenol-chloroform extraction. Genomic DNA from both strains was digested using *RsaI*. Subtractive hybridization was preformed using the Clontech PCR-Select Bacterial Genomic Subtraction kit (Clontech Laboratories Inc., Palo Alto, CA) according to the manufacturer's instructions. Briefly, the tester DNA (DEC6a) is divided into two aliquots with each being ligated to a different adaptor. Two hybridizations are performed. In the first, an excess of the driver DNA (K-12) is added to each of the two adaptor-ligated aliquots. The samples are denatured and then allowed to anneal. In the second hybridization, the products from the first hybridization are mixed and denatured excess driver DNA is added. The mixture of molecules is then subjected to PCR amplification which amplifies the DEC6a specific sequences.

A library of the subtracted fragments was constructed using the TA Cloning kit (Invitrogen Corporation, Carlsbad, CA). The clones containing the inserts were selected using kanamycin (50 μg/ml) and X-gal markers. The clones were then purified using the UltraClean Mini Plasmid Prep Kit (MoBio Laboratories, Inc., Solana Beach, CA) and sequenced using the universal forward (T7) and reverse (M13) primers on a Beckman CEQ2000 (Beckman Coulter Inc., Fullerton, CA) automated sequencer.

SSH analysis. The vector and adaptor sequences were trimmed from each sequence and a contig was constructed for each clone using Lasergene software (DNASTAR, Inc., Madison, WI). The concatenated sequences were screened using the

National Center for Biotechnology Information (NCBI) Basic BLAST, Unfinished Genomes BLAST, and ORF Finder databases to identify homologous genes and proteins.

PCR screening of SSH results. Eleven O111:H12 and O111:H21 isolates and EAEC isolate N49 (also known as O42) were grown overnight in 100 mL of LB broth at 37°C. DNA was isolated from the strains using the Puregene DNA Isolation Kit (Gentra Systems, Minneapolis, MN). Primers were designed for two of the genes (*virK* and *wbdM*) identified from the database searches and are as follows: virK_1 5' – GGGTATTGTCCGTTCCGAT – 3'; virK_2 5' – ACAACGATACCGTCTCCCG – 3'; wbdM p1 5' – CTTACTTGTGGTGGAGCCGA – 3'; and wbdM p2 5' – GGACGTTCACACGCCATAGC – 3'. Primers for the *astA* gene were previously reported by Monterio-Neto (72). Boehringer-Manheim Taq DNA Polymerase was used to amplify the products in the O111:H12 and O111:H21 isolates under the following conditions: 94° for 1 min, with 35 cycles of 94° for 1 min, 50° – 53° for 2 min, 72° for 3 min. DEC6a and K-12 were used as positive and negative controls respectively. All products were electrophoresed on a 0.8% agarose gel.

Table 14. Primer sequences, positions and amplicon sizes for loci identified by PESM clone 249.

Locus	Primer	Primer sequence	Size of amplicon
suhB	suhB140	5' – CCGAAGCGGTGATTATCGAC – 3'	652 bp
	suhB791	5' – GCGTCGCTTAACTCGTCAC – 3'	
csiE	csiE940	5' – TCCTGCGCTATCATCAACTCACAC – 3'	911 bp
	csiE1104	5' – TTCGCGTAACTGCTGCTCAATCT – 3'	
hcaT	hcaT350	5' – TGGCGAATACGTGGCAAAAGCAGT – 3'	657 bp
	hcaT1006	5' – CCATCGCGACGGCAGAGTAAACC – 3'	
hcaA1	hcaA1260	5' – ACCGGGCCATGCGTGTGAGTT – 3'	958 bp
	hcaA11217	5' – TCGTCGCGGCGCTTTTCCTG – 3'	
hcaD	hcaD35	5' – GGCAAGCGGCGGCAATGG – 3'	919 bp
	hcaD953	5' – CACGGCGGCGGCAGTAGC – 3'	
yphB	yphB185	5' – TTGTCTGGCAGGGGCGTGAGTATC – 3'	540 bp
	yphB724	5' – CAAACGCAGGGTCGGAAACAAAGA – 3'	
yphC	yphC144	5' – CGGGATTTGCGGAAGCGATGTC – 3'	877 bp
	yphC1020	5' – CAGCGAGAAGCGATGGGTAATGG – 3'	
yphE	yphE137	5' – GCGCGGGCAAATCGACTCTCAT – 3'	1221 bp
	yphE1357	5' – CGGCAGCCAGCTCACGGACAATA – 3'	
yphF	yphF141	5' – GCGTCAGGGCGTTCAGGATGC – 3'	573 bp
	yphF713	5' – GCTTTTACCGCGCCGAGTGTC – 3'	
yphG	yphG35	5' – GTTCAATACACTGCCACAAATCTT – 3'	3263 bp
	yphG3297	5' – CAATTCAGCGCGAGCAGACT – 3'	
glyA	glyA284	5' – CGCACTCCGGCTCCCAGGCTAACT – 3'	961 bp
	glyA1244	5' – ACCGGGTAACGTGCGCAGATGTCG – 3'	

RESULTS

PESM. The ends of the 136 clones with insert size greater or equal to 8 kb were sequenced using the Beckman CEQ DNA Analyzer. The sequencing was of relatively good quality with reads of greater than 400 bases. Of the 136 paired ends, 51 sequences matched vector (pSMART) sequences in BLAST searches and were not useful for the analysis.

With the program PENDMAP, there were 47 inserts in which both ends mapped to the *E. coli* K-12, EDL-933, and/or Sf301 genomes and fell between 5 and 20 kb apart. The mapped distances are summarized in Table 15. There were 26 clones in which only one end mapped to a genome or the mapped distance was much greater than 20 kb. The paired ends of 12 clones had no match to either genome and potentially represent sequence that is unique to the *S. dysenteriae* type 1 strain. Among the 47 conserved regions, there are 19 clones that map to regions of similar length (within 1 kb) in all of the reference genomes. A two-way comparison of genomes identified 30 clones mapping to regions of similar length in the K-12 and EDL-933 genomes, 22 clones between the K-12 and Sf301 genomes, and 22 clones between the EDL-933 and Sf301 genomes. The paired ends of 14 clones have map distances greater than 12 kb in at least one genome. The size of the actual insert estimated on agarose gels is < 12 kb so that each of these regions are candidates for deletions in *Shigella dysenteriae* of one to several kb. Interestingly, paired end mapping of clone 245 differs in distance of 10 to 12 kb from the Sf301 to the K-12 and EDL-933 genomes. The region identified by this clone includes *rfa* genes involved in the LPS core biosynthesis.

Table 15. PESM fragments with both ends (*k1* and *k2*) matching the reference genomes of *E. coli* K-12, EDL-933, or Sf301.

Clone	Region	K-12 distance (kb)	EDL-933 distance (kb)	Sf301 distance (kb)
37	*Z0609 – Z0615*		20.159	
48	*yaiC – proC*	1.941	1.941	1.941
50	*artJ – art M*	1.69	1.69	1.691
51	*yagN – yagR*	5.795		
69	*ygaA – ascF*	10.206	10.002	8.531
73	*ygcF – pyrG*	4.919	7.51	5.656
80	*cysI – ygcE*	14.986	14.986	
88	*recQ – udp*	11.31	11.333	11.311
92	*yjjW – serB*	11.183	10.14	10.13
98	*Z0609 – Z0615*		20.159	
121	*b1754 – ansA*	15.806	16.597	18.514
146	*ipaH9.8 – yphD*			4.759
154	*ybgL – sdhC*	10.401	10.4	9.325
155	*fepE – fepB*	6.256	6.256	7.146
164	*yfhK – pinH*			10.851
165	*rrsC – ilvG_1 (ilvG)*	9.735	10.398	11.531
176	*adiY – SF4285*			19.3215
179	*lysP – yeiL*	8.951	8.9	7.146
195	*rrlC – ilvM*	8.761	8.763	11.791
198	*ybgJ – ybgD*		9.204	
214	*yi22_1 – hemB*	8.086		5.464
242	*ybjN – Z1087*		8.311	27.67
245	*rfaF – rfaQ (kdtA)*	13.169	11.914	1.811
249	*b2532 (Z3799) – yphG*	20.255	20.253	20.001
277	*ybgH – sdhA*	16.589	16.597	14.868

Table 15 (continued).

Clone	Region	K-12 distance (kb)	EDL-933 distance (kb)	Sf301 distance (kb)
301	*nusA – yhbX*	5.592	5.592	3.941
302	*slp – yhiD*	10.996	10.996	2.105
309	*alaS – ygaD*	4.966	2.648	5.142
320	*rfaQ (waaQ) – yicF*	14.112	14.11	14.754
376	*yjfN – sgaE*	8.369	8.369	8.369
382	*yajO – ampG*	16.384	16.259	22.573
390	*ptsO – yhcC*	6.078	6.085	5.687
400	*yqjE – yhaI*	3.584	3.581	3.995
401	*yhaL – tdcE*	7.382	7.381	
420	*b2809 (Z4126) – recB*	13.085	13.085	12.984
433	*ydiA – btuD*	6.113	6.113	6.114
435	*yjgR – fecA*	28.116		
438	*dapB – caiB*	10.741	10.773	10.741
443	*yraK – yhbV*	12.026	11.426	12.104
444	*yjcC – yjcE*	6.158	6.157	6.308
471	*ycfO – mfd*	9.869	9.888	9.956
550	*b2373 (Z3637) – b2380 (Z3645)*	10.112	10.114	9.890
562	*ytfJ – aspA*			23.228
589	*arsB – chuA*		8.197	
610	*mazG – chpR*	1.445	1.445	
615	*rpoN – gltB*	14.426	14.468	14.466
652	*hyfR – purM*	10.313	10.336	10.411

The best candidate for a "black hole" deletion of the order of 5-10 kb is detected by the paired end mapping for clone 249. The paired ends from this cloned insert matched the *suhB* and *yphG* genes respectively, which covers an approximately 18- 20 kb region in the reference genomes region containing 17 ORFs (Figure 20). This region includes the *hca* cluster of five catabolic genes arranged as a putative operon (*hcaA1A2CBD*) and two additional genes transcribed in the opposite direction that encode a potential permease (*hcaT*) and a regulator (*hcaR*) (19). The products of these genes are involved in the dioxygenolytic pathway for initial catabolism of 3-phenylpropionic acid in *E. coli* K-12 (19).

The *hca* region was examined using a series of PCR assays with primers designed from genomic sequences to detect genes left intact and identify genes that may have been lost. This screen identified the loss of at least 7 genes within this region. Additionally, eight *Shigella* and EIEC isolates were assayed for gene loss in this region. Four isolates were missing at least one locus with 1007-74, a *S. dysenteriae* type 1 isolate having the same pattern of loss as 3823-69 (Table 16).

The *hca* region was examined in *E. coli* strains Sakai (O157:H7), E2348/69 (O127:H6), and the Dysenteriae 1 library strain (3823-69) using long PCR to confirm the deletion (Figure 21). The expected product sizes were determined from either completed or unfinished genome sequences and are as follows: *E. coli* K-12 and Sakai, 16,488 bp; E2348/69, 11,469 bp; and *S. dysenteriae*, 6,270 bp.

A report of the genome sequence of *S. flexneri* 301 identified the *hcaD* locus as a pseudogene with inactivation caused by a mutational stop codon (54). Additionally, a

search of the unfinished genome of *S. dysenteriae* M131649 using coliBASE (http://colibase.bham.ac.uk/) identified a similar deletion of ORFs in this strain.

SSH. Nucleotide sequences were obtained for 114 of the 120 clones that were screened from the subtraction library. In 19 of the clones, the forward and reverse sequences were non-overlapping and were considered to be two separate contigs for the remaining analyses. Database searches of the clones resulted in 119 matches with the NCBI databases with only 8 clones having no reported matching sequence (Figure 22). Of the known database matches, 21 of the cloned fragments showed homology with the previously published *E. coli* K-12 genome sequence (9). These genes were not explored further as they are most likely remnants that were not removed by the technique. From the 119 database matches, 38 of the known matches were homologous to the published sequence of an *E. coli* O157:H7 genome (85).

Extrachromosomal elements including bacteriophages, plasmids and insertion elements accounted for 47% of the identified differences between K-12 and DEC6a. From the database and literature searches, three loci were chosen to investigate the distribution among O111 serotypes and *E. coli* that express the aggregative phenotype. The enteroaggregative protein (Eap) is encoded by *virK* on the pAA2 plasmid of EAEC (15) and showed homology to clone 15. EAST-1, a heat-stable enterotoxin of EAEC encoded by the *astA* locus had been previously used as a probe in virulence studies of O111:H12 *E. coli* strains (72). The *wbdM* locus is a putative glycosyl transferase that is specific to the O111 serogroup (126). PCR screening of the eleven additional O111 isolates resulted in 4 of the isolates, including DEC6a being positive for *virK*, *astA*, and

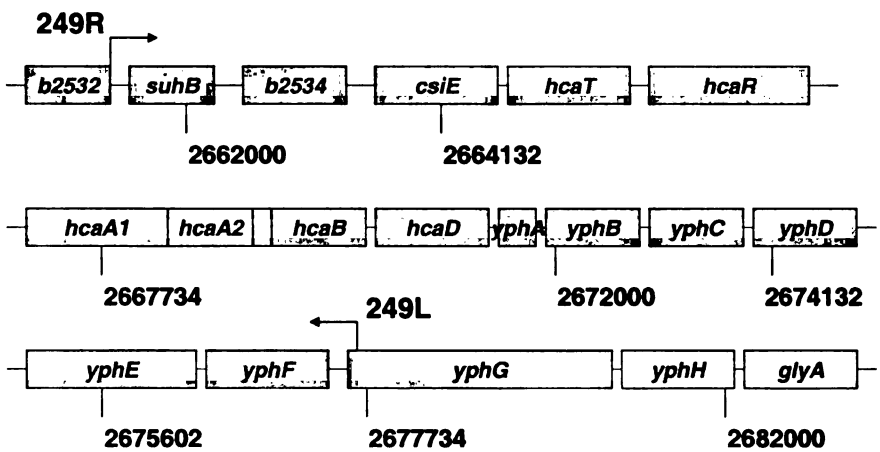

Figure 20. Diagram of the *E. coli* K-12 *hca* genomic region. This region was identified by PESM clone 249 as a potential deletion in the Dysenteriae 1 genome. The numbers indicate the position within the K-12 genome. The ends and direction of the Dysenteriae genome fragment are indicated by 249R and 249L.

Table 16. PCR screening results of the *hca* region in *Shigella* and EIEC isolates. A larger than expected amplicon (indicated by *) was found for the *hcaD* locus in isolate B14.

Isolate	Serotype	Locus										
		suhB	*csiE*	*hcaT*	*hcaA1*	*hcaD*	*yphB*	*yphC*	*yphE*	*yphF*	*yphG*	*glyA*
3823-69	D1	+	+	+	-	-	-	-	-	-	-	+
1007-74	D1	+	+	+	-	-	-	-	-	-	-	+
5514-56	D10	+	-	+	+	+	+	-	+	-	-	+
2770-51	B14	+	+	+	+	+*	+	+	+	+	+	+
4822-66	SS	+	+	+	+	+	+	+	+	+	+	+
2747-71	F2A	+	+	+	+	+	+	+	+	+	+	+
3470-56	D7	+	+	+	+	+	+	+	+	+	-	+
LT-94	O-:H-	+	+	+	+	+	+	+	+	+	+	+
202-72	O124:H30	+	-	+	+	+	+	+	+	+	-	+

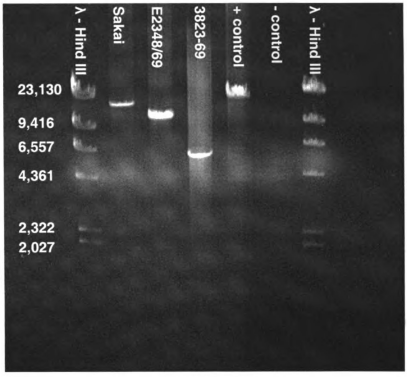

Figure 21. Long PCR confirmation of the "black hole" in the *hca* genomic region. The lanes are as follows: 1 and 7) λ Hind III size marker, 2) Sakai (O157:H7), 3) E2348/69 (O127:H6), 4) 3823-69 (Dysenteriae 1), 5) + control provided by Epicentre, and 6) – control. The expected product sizes are based on both finished and unfinished genome sequences. Both E2348/69 and the unfinished Dysenteriae genomes indicate a loss of open reading frames within this region.

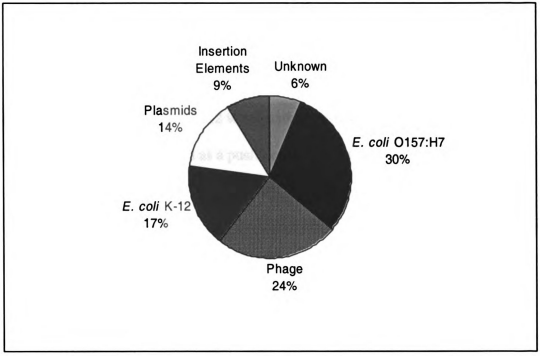

Figure 22. Summary database matches for the subtracted Dec 6a genomic fragments. Almost half of the genomic fragments matched with mobile elements known to be responsible for lateral gene transfer.

wbdM (Table 17). Eap was present in 50% of the O111 isolates while *wbdM* locus was present in all isolates except AC-C10 and AC-C6.

Table 17. PCR screening results for *virK*, *astA*, and *wbdM*. These loci were identified by the SSH method. Dec 6a and K-12 were used as positive and negative controls respectively and N49 was used as a positive contol for the *virK* locus.

Isolate	Serotype	*virK*	*astA*	*wbdM*
TR131	O111ab:H12	-	+	+
TR175	O111ab:H21	+	+	+
TR7	O11ab:H12	+	+	+
2277-67	O111:H12	-	-	+
448-71	O111:H21	-	+	+
AC-C10	O111:H12	+	-	-
AC-C6	O111:H12	+	-	-
C750-59	O111:H12	-	-	+
C586-63	O111:H21	-	+	+
LT-177	O111:H12	-	+	+
247-69	O111:H12	+	+	+
DEC6a	O111:H21	+	+	+
K-12		-	-	-
N49	O44:H18	+	Not tested	Not tested

DISCUSSION

Gene acquisition, loss, and rearrangment are major contributing factors in the evolution of virulence. A recent report on the sequence of the *S. flexneri* 2A, 2457T isolate indicates the presence of 37 islands, 372 psuedogenes and 6.7% of the genome consisting of insertion sequences (129). As with the genome sequences of the two *E. coli* O157:H7 isolates (49, 85), there are distinct differences between the sequenced Flexneri genomes (54, 129). Based on these findings, one can conclude that the genome sequence for a given isolate is not necessarily representative of all isolates within the species or even within a particular serotype. Lan and Reeves (60) propose that a true "species" genome can only be obtained by having the sequence of all the DNA that is important for a species. For example, to have a "species" genome for *E. coli*, one would have to know the nucleotide sequence for at least one representative of each distinct pathovar.

In order to find the important nucleotide regions that may be responsible for a distinct phenotype, alternative approaches to genome sequencing can be employed. This project reports the development of a method to screen pathogenic *E. coli* and *Shigella* genomes for major insertions and deletions. By analyzing the distribution of major indels in a diversity of strains, it can be determined which are strain specific, which are group specific, and which have occurred in parallel among multiple groups. It is the last class of indels that are the most likely candidates marking adaptive changes resulting from natural selection.

A paired end sequencing mapping approach was used to compare the genomes of a Dysenteriae type 1 isolate to the published genomes of *E. coli* K-12 (9), EDL-933 (85), and Sf301 (54). This method proved useful in identifying indels in the unknown

Dysenteriae 1 genome, but also in the comparison of the reference genomes to one another. A potential pathogenicity adaptive mutation (111) was discovered to be a presumed deletion in the *hca* operon. PCR screening of loci within this region indicates the loss of at least 7 ORF's in *S. dysenteriae* type 1 isolates. The results were confirmed independently with a different Dysenteriae 1 isolate using the same primer sets for amplification (R. Binet, personal communication). Additionally, the unfinished genome of *S. dysenteriae* isolate M131649 shows a large gap at this region when aligned to other *Shigella* and *Escherichia* sequences. It is quite possible that there is a selective advantage to inactivating the genes or gene products of the *hca* encoded pathway.

The *hca* pathway is involved in the degradation of 3-phenylpropionic acid (19). The bacterial mediated degradation of this aromatic compound occurs in the intestinal tract (20). A study measuring acid resistance and acid sensitivity found that propionic acid was more inhibitory to *S. flexneri* than other acids (118). It would be interesting to measure the acid resistance of the Dysenteriae type 1 isolate and compare with the results of Tetteh (118) to determine if loss of this operon may play a role in the survival of the acidic environment.

Additional studies are necessary to confirm or refute the identified indels. The long range PCR approach showing the difference in amplicon sizes between the known and unknown genomes provides greater evidence in support of a true insertion or deletion. Another method to confirm the PESM findings is by hybrizidation techniques. A future goal of this approach is to survey the distribution of genes of the major indels in related strains of the same phylogenetic group and across representative strains of different groups. Again, long range PCR or hybridizations could be used for these

analyses. Cases where major deletions have occurred in parallel in divergent clones are the most likely candidates for black hole formation (16).

The previously described technique of suppression subtractive hybridization was used to identify the genomic content that was unique to DEC 6a, a close relative to the non-pathogenic laboratory strain K-12. This procedure yielded fragments that were unique to DEC6a as well as some remnants of the K-12 genome. Out of 120 clones, 94% were able to be identified using NCBI databases. Many of these matches were with extrachromosomal elements that can be readily exchanged between bacteria. Matches with both pathogenic and non-pathogenic *E. coli* accounted for 46% of the identified fragments of the DEC6a genome. A smaller fraction of sequences obtained using this method provided no known homologues and could be unique to this group of pathogenic clones. These unidentified open reading frames could be potential virulence determinants and warrant further investigation. With additional database depositions and genome sequences for comparative genomic studies, it is likely that a better characterization and understanding of this isolate will come sooner rather than later.

In addition to the methods described here, there are numerous other comparative genomic techniques that have been employed. A genome based approach for comparing closely related isolates was put forth by Ohnishi et al. (82). This approach uses 549 overlapping primer pairs to cover the entire Sakai genome. The results from the comparison of O157 isolates having diverse restriction profiles identified conserved O157 specific regions. The genome scanning approach also identified diversity between the isolates created by bacteriophage (82). This technique proves to be useful in

differentiating O157 isolates; however, it can not be applied to other bacterial systems without the development of additional species specific primer pairs.

It is the ultimate goal of comparative genomics to reveal similarities and differences between bacterial organisms. One readily available and high throughput approach is the hybrization of an unknown genome to DNA or oligonucleotide arrays constructed from a known genome. This approach fails in identifying genes present in the unknown strain but absent from the standard on the array (10). Unlike hybridization to microarrays, the SSH approach is conducive to finding novel elements that may encode new pathogenicity islands or other virulence factors. Agron and colleagues (4) showed that SSH could identify 95% of the unique open reading frames of two *Helicobacter pylori* isolates whose genomes were entirely sequenced. Before the completion of a genome sequence, SSH was used to compare the unknown *Salmonella typhimurium* genome to the known *E. coli* K-12 genome. These results indicated that the genomes were similar in size and content with 170 kb of sequence being present in *S. typhimurium* but absent from *E. coli* (10). Recently, a putative 42 kb pathogenicity island was identified in *S. flexneri* 2A by Walker (125). This comparison was made to an *E. coli* K-12 isolate with the majority of the subtracted differences being relatively small in size (125). SSH has also been used by Radnedge to infer evolutionary relationships among the biovars of *Yersinia pestis* (90). This study identified six difference regions which were then examined for the presence or absence in a diverse collection of isolates. The findings of the Radnedge study (90) along with the results of an IS100 fingerprint study by Motin (74) were combined by Whittam (131) to provide an overview of the

genotypic classification and evolutionary events of the *Y. pestis* strains that were common to both studies.

SSH is an attractive technique for identifying novel genomic sequences, however, it will not be able to provide information on point mutations or interrupted genes. In a comparison of two known genomes, this method provided numerous false positive results (10). Even though the results from a SSH experiment can be used to survey additional isolates, the hybridization methods will only permit a one-way subtraction between two genomes allowing for the identification of insertions of large DNA regions. A second, SSH experiment in which the tester and driver were switched would be necessary for the identification of deletions. In an attempt to remedy these downfalls, paired end sequence mapping was developed to allow for comparison of multiple genomes, detect both chromosomal acquisitions and losses simultaneously, and provide a chromosomal frame of reference to pinpoint location. Another appealing aspect of the PESM method is the ease with which additional genomes can be added for *in silico* comparison.

ACKNOWLEDGEMENTS

Portions of the work presented in this chapter will be published to introduce the PESM technique. Further work will be done to verify additional indels before the chapter can be published in its entirety.

I thank R. Mangold and T. Whittam for writing the computing modules for the PESM analysis. I also thank D. Mead at Lucigen for library construction. A. Plovanich-Jones and D. Lacher have graciously taken on the PESM project to verify the indels and explore other vectors for library construction. I thank C. Tarr for helping with the library construction and providing suggestions regarding the SSH project.

CHAPTER 5

SUMMARY AND SYNTHESIS

The overall purpose of the research presented in this dissertation is to examine the evolution of invasiveness in *Escherichia coli* and *Shigella*. Each project within this work serves to add to the existing body of knowledge regarding these pathogens by providing insight into the genotypic and phenotypic characteristics that distinguish them from other *E. coli* as well as from each other. In addition, the genetic methodologies described in this work can be expanded to examine the relationships and genomic content of other prokaryotic organisms.

The multilocus sequencing approach based on housekeeping loci first developed a phylogenetic framework that provided the basis for many of the hypotheses that stimulated additional experiments and comparisons. Overall, the results of the multilocus sequencing demonstrated that *Shigella* fall within the diversity of *Escherichia*. This is in support of previous work by Ochman (81), Pupo (87), and Reeves (88). The dendogram clearly shows that the invasive clones, either enteroinvasive *E. coli* or *Shigella*, have arisen independently at multiple times within the evolutionary history of *Escherichia*. With the knowledge of the relationships, hypotheses were made regarding the acquisition of virulence loci, loss of housekeeping loci, and variation in phenotypic virulence attributes.

Numerous virulence loci have been identified and characterized in *Shigella* (1, 5, 7, 31, 64, 67, 73, 77, 89, 91, 124). By using the information obtained from the phylogenetic analysis, a survey of the population was performed with the goal of identifying patterns of acquisition within and between the clonal lineages. A set of six loci was examined in both *Shigella* and EIEC. When the distribution of virulence loci is examined with respect to the framework, numerous patterns are apparent. The loss of a

functional mannitol operon appears to have occurred independently at four different points in the evolution of these pathogens. The pINV plasmid conferring the distinctive phenotype has been acquired many times to give rise to the invasive clones. The distribution of the areobactin loci indicative of either the SHI-2 or SHI-3 pathogenicity islands is rather predictable within the shigellae; however, the distribution is a bit sparser in the EIEC lineages.

It can be hypothesized that the EIEC isolates are intermediates between the *E. coli* and *Shigella.* Support for this idea comes from individual EIEC isolates that lie immediately outside of a recognized *Shigella* group (as is the case for Group 1 and isolate LT-68). EIEC isolates harbor the large invasion plasmid, however this appears to be only the first step in the evolutionary pathway that leads to *Shigella.* Further gene acquisition of virulence loci, such as *set1A* or *pic*, leads to another step. Loss of certain biochemical traits are also necessary for this stepwise evolution to occur.

The combined results of the distinct clonal groups with attained virulence loci along with biochemical and epidemiological reports suggest a basis for variation in the phenotypic virulence properties of these pathogens. The goal of the second project was to assay numerous invasive and reference isolates to determine phenotypic variation within and between clonal groups. Additional hypotheses regarding the impact of genetic composition on virulence was also examined. This population study based on the clonal framework showed extensive variation both within and between groups. This is the first known report of a survey of *Shigella* invasive phenotypes with respect to a phylogenetic approach.

The impact of inactivation of the mannitol operon was assessed using two Dysenteriae type 3 isolates. One isolate, 225-75 had an insertional element within the *mtlD* locus. The second isolate, 2415-49 had an intact and functional mannitol operon. Phenotypic virulence assays did not identify either mannitol phenotype as exhibiting a pathoadaptive mutation or selective advantage. It is interesting to note, however, that 2415-49 was gentamicin resistant thereby perhaps inflating the adherence measurement.

The third project used genome based comparisons to identify genome additions or deletions as compared to a reference strain. A new approach, paired end sequence mapping (PESM) was introduced and demonstrated to be useful in identifying a potential "black hole" in the genome of a *S. dysenteriae* type 1 isolate. Another technique, suppression subtractive hybridization (SSH) was used to identify unique chromosomal elements in an atypical EPEC isolate. Both approaches provided the results necessary to begin investigating the presence or absence of loci using PCR based screening methods. Additionally, closely related isolates could be screened to determine the distribution of the indel within the context of a phylogeny.

Future considerations. The multilocus sequencing approach using 15 loci has proved useful in elucidating the evolutionary relationships of enteroinvasive as well as other pathovars of *E. coli*. A study by Hyma (52) has shown that a subset of these loci will also discriminate between the Reeves Groups (88). An additional report by Tarr et al. (116) has demonstrated a high throughput approach termed Multilocus Virulence Gene Profiling (MVGP) to screen for numerous virulence loci simultaneously. Both of these approaches maintain sensitivity while decreasing time and cost.

Phenotypic virulence assays can provide useful information regarding the many stages in the invasion process. Due to the age and unknown storage methods of many of the isolates used in these studies, it is conceivable to believe that freshly isolated strains or those that have been selected to maintain the virulence plasmid would yield somewhat different results. The data presented for these assays is informative and offers suggestion for future assays. Additionally, the data obtained in these experiments can be expressed and analyzed by various methods.

The paired end sequence mapping technique has preliminarily shown to be a useful technique for genome comparison. Additional confirmation must be made before it could be considered a new technique with wide-spread application. Exploration for alternatives in library construction is on-going. A graphical comparison of genomes and paired end fragments would greatly enhance the PENDMAP program.

REFERENCES

1. **Adhikari, P., G. Allison, B. Whittle, and N. K. Verma.** 1999. Serotype 1a O-antigen modification: molecular characterization of the genes involved and their novel organization in the *Shigella flexneri* chromosome. J Bacteriol **181:**4711-8.

2. **Adiri, R. S., U. Gophna, and E. Z. Ron.** 2003. Multilocus sequence typing (MLST) of *Escherichia coli* O78 strains. FEMS Microbiol Lett **222:**199-203.

3. **Adler, B., C. Sasakawa, T. Tobe, S. Makino, K. Komatsu, and M. Yoshikawa.** 1989. A dual transcriptional activation system for the 230 kb plasmid genes coding for virulence-associated antigens of *Shigella flexneri*. Mol Microbiol **3:**627-35.

4. **Agron, P. G., M. Macht, L. Radnedge, E. W. Skowronski, W. Miller, and G. L. Andersen.** 2002. Use of subtractive hybridization for comprehensive surveys of prokaryotic genome differences. FEMS Microbiol Lett **211:**175-82.

5. **Al-Hasani, K., B. Adler, K. Rajakumar, and H. Sakellaris.** 2001. Distribution and structural variation of the *she* pathogenicity island in enteric bacterial pathogens. J Med Microbiol **50:**780-6.

6. **Al-Hasani, K., I. R. Henderson, H. Sakellaris, K. Rajakumar, T. Grant, J. P. Nataro, R. Robins-Browne, and B. Adler.** 2000. The *sigA* gene which is borne on the *she* pathogenicity island of *Shigella flexneri* 2a encodes an exported cytopathic protease involved in intestinal fluid accumulation. Infect Immun **68:**2457-63.

7. **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.

8. **Andrade, A., J. A. Giron, J. M. Amhaz, L. R. Trabulsi, and M. B. Martinez.** 2002. Expression and characterization of flagella in nonmotile enteroinvasive *Escherichia coli* isolated from diarrhea cases. Infect Immun **70:**5882-6.

9. **Blattner, F. R., G. Plunkett, 3rd, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao.** 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277:**1453-74.

10. **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. A. Vorobiev, G. L. Andersen, and E. D. Sverdlov.** 1999. Identification and localization of differences between *Escherichia coli* and *Salmonella typhimurium* genomes by suppressive subtractive hybridization. Mol Gen Genet **262:**721-9.

11. **Bukholm, G., and J. Lassen.** 1982. Bacterial adhesiveness and invasiveness in cell culture monolayer. 2. In vitro invasiveness of 45 strains belonging to the family Enterobacteriaceae. Acta Pathol Microbiol Immunol Scand [B] **90:**409-13.

12. **Campos, L. C., T. S. Whittam, T. A. Gomes, J. R. Andrade, and L. R. Trabulsi.** 1994. *Escherichia coli* serogroup O111 includes several clones of diarrheagenic strains with different virulence properties. Infect Immun **62:**3282-8.

13. **Cersini, A., A. M. Salvia, and M. L. Bernardini.** 1998. Intracellular multiplication and virulence of *Shigella flexneri* auxotrophic mutants. Infect Immun **66:**549-57.

14. **Colonna, B., M. Casalino, P. A. Fradiani, C. Zagaglia, S. Naitza, L. Leoni, G. Prosseda, A. Coppo, P. Ghelardini, and M. Nicoletti.** 1995. H-NS regulation of virulence gene expression in enteroinvasive *Escherichia coli* harboring the virulence plasmid integrated into the host chromosome. J Bacteriol **177:**4703-12.

15. **Czeczulin, J. R., T. S. Whittam, I. R. Henderson, F. Navarro-Garcia, and J. P. Nataro.** 1999. Phylogenetic analysis of enteroaggregative and diffusely adherent *Escherichia coli*. Infect Immun **67:**2692-9.

16. **Day, W. A., Jr., R. E. Fernandez, and A. T. Maurelli.** 2001. Pathoadaptive mutations that enhance virulence: genetic organization of the *cadA* regions of *Shigella* spp. Infect Immun **69:**7471-80.

17. **Dehio, C., M. C. Prevost, and P. J. Sansonetti.** 1995. Invasion of epithelial cells by *Shigella flexneri* induces tyrosine phosphorylation of cortactin by a pp60c-*src*-mediated signalling pathway. Embo J **14:**2471-82.

18. **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 for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc Natl Acad Sci U S A **93:**6025-30.

19. **Diaz, E., A. Ferrandez, and J. L. Garcia.** 1998. Characterization of the *hca* cluster encoding the dioxygenolytic pathway for initial catabolism of 3-phenylpropionic acid in *Escherichia coli* K- 12. J Bacteriol **180:**2915-23.

20. **Diaz, E., A. Ferrandez, M. A. Prieto, and J. L. Garcia.** 2001. Biodegradation of aromatic compounds by *Escherichia coli*. Microbiol Mol Biol Rev **65:**523-69.

21. **Donnenberg, M. S., A. Donohue-Rolfe, and G. T. Keusch.** 1990. A comparison of HEp-2 cell invasion by enteropathogenic and enteroinvasive *Escherichia coli*. FEMS Microbiol Lett **57:**83-6.

22. **Donnenberg, M. S., A. Donohue-Rolfe, and G. T. Keusch.** 1989. Epithelial cell invasion: an overlooked property of enteropathogenic *Escherichia coli* (EPEC) associated with the EPEC adherence factor. J Infect Dis **160:**452-9.

23. **Donnenberg, M. S., and T. S. Whittam.** 2001. Pathogenesis and evolution of virulence in enteropathogenic and enterohemorrhagic *Escherichia coli*. J Clin Invest **107:**539-48.

24. **DuPont, H. L., M. M. Levine, R. B. Hornick, and S. B. Formal.** 1989. Inoculum size in shigellosis and implications for expected mode of transmission. J Infect Dis **159:**1126-8.

25. **Echeverria, P., O. Sethabutr, O. Serichantalergs, U. Lexomboon, and K. Tamura.** 1992. *Shigella* and enteroinvasive *Escherichia coli* infections in households of children with dysentery in Bangkok. J Infect Dis **165:**144-7.

26. **Edwards, P. R., and W. H. Ewing.** 1986. Edwards and Ewing's Identification of Enterobacteriaceae, 4th ed. Elsevier, New York.

27. **Elsinghorst, E. A., and D. J. Kopecko.** 1992. Molecular cloning of epithelial cell invasion determinants from enterotoxigenic *Escherichia coli*. Infect Immun **60:**2409-17.

28. **Enright, M. C., and B. G. Spratt.** 1999. Multilocus sequence typing. Trends Microbiol **7:**482-7.

29. **Ewing, W. H., and J. L. Gravatti.** 1947. *Shigella* types encountered in the Mediterranean area. J Bacteriol **53:**191-195.

30. **Falconi, M., B. Colonna, G. Prosseda, G. Micheli, and C. O. Gualerzi.** 1998. Thermoregulation of *Shigella* and *Escherichia coli* EIEC pathogenicity. A temperature-dependent structural transition of DNA modulates accessibility of *virF* promoter to transcriptional repressor H-NS. Embo J **17:**7033-43.

31. **Fasano, A., F. R. Noriega, D. R. Maneval, Jr., S. Chanasongcram, R. Russell, S. Guandalini, and M. M. Levine.** 1995. *Shigella* enterotoxin 1: an enterotoxin of *Shigella flexneri* 2a active in rabbit small intestine *in vivo* and *in vitro*. J Clin Invest **95:**2853-61.

32. **Faundez, G., G. Figueroa, M. Troncoso, and F. C. Cabello.** 1988. Characterization of enteroinvasive *Escherichia coli* strains isolated from children with diarrhea in Chile. J Clin Microbiol **26:**928-32.

33. **Feil, E. J., J. E. Cooper, H. Grundmann, D. A. Robinson, M. C. Enright, T. Berendt, S. J. Peacock, J. M. Smith, M. Murphy, B. G. Spratt, C. E. Moore, and N. P. Day.** 2003. How clonal is *Staphylococcus aureus*? J Bacteriol **185:**3307-16.

34. **Formal, S. B., E. H. LaBrec, T. H. Kent, and S. Falkow.** 1965. Abortive intestinal infection with an *Escherichia coli - Shigella flexneri* hybrid strain. J Bacteriol **89:**1374-1382.

35. **Frischknecht, F., and M. Way.** 2001. Surfing pathogens and the lessons learned for actin polymerization. Trends Cell Biol **11:**30-38.

36. **Fukushima, M., K. Kakinuma, and R. Kawaguchi.** 2002. Phylogenetic Analysis of *Salmonella*, *Shigella*, and *Escherichia coli* strains on the basis of the *gyrB* gene sequence. J Clin Microbiol **40:**2779-85.

37. **Galan, J. E.** 1996. Molecular genetic bases of *Salmonella* entry into host cells. Mol Microbiol **20:**263-71.

38. **Germani, Y., and P. J. Sansonetti.** 2001. The Genus *Shigella*. *In* M. Dworkin (ed.), The Prokaryotes., First electronic edition. ed. Springer-Verlag, New York.

39. **Gordillo, M. E., G. R. Reeve, J. Pappas, J. J. Mathewson, H. L. DuPont, and B. E. Murray.** 1992. Molecular characterization of strains of enteroinvasive *Escherichia coli* O143, including isolates from a large outbreak in Houston, Texas. J Clin Microbiol **30:**889-93.

40. **Griffiths, E., P. Stevenson, T. L. Hale, and S. B. Formal.** 1985. Synthesis of aerobactin and a 76,000-dalton iron-regulated outer membrane protein by *Escherichia coli* K-12-*Shigella flexneri* hybrids and by enteroinvasive strains of *Escherichia coli*. Infect Immun **49:**67-71.

41. **Groisman, E. A., and H. Ochman.** 1993. Cognate gene clusters govern invasion of host epithelial cells by *Salmonella typhimurium* and *Shigella flexneri*. Embo J **12:**3779-87.

42. **Guerrant, R. L., T. S. Steiner, A. A. Lima, and D. A. Bobak.** 1999. How intestinal bacteria cause disease. J Infect Dis **179 Suppl 2:**S331-7.

43. **Hacker, J., and J. B. Kaper.** 1999. The concept of pathogenicity islands, p. 1-11. *In* J. Hacker and J. B. Kaper (ed.), Pathogenicity islands and other moblie virulence elements. American Society for Microbiology, Washington D. C.

44. **Hale, T. L.** 1991. Genetic basis of virulence in *Shigella* species. Microbiol Rev **55:**206-24.

45. **Hale, T. L., and S. B. Formal.** 1987. Pathogenesis of *Shigella* infections. Pathol Immunopathol Res **6:**117-27.

46. **Hale, T. L., P. J. Sansonetti, P. A. Schad, S. Austin, and S. B. Formal.** 1983. Characterization of virulence plasmids and plasmid-associated outer membrane proteins in *Shigella flexneri*, *Shigella sonnei*, and *Escherichia coli*. Infect Immun **40:**340-50.

47. **Harris, J. R., J. Mariano, J. G. Wells, B. J. Payne, H. D. Donnell, and M. L. Cohen.** 1985. Person-to-person transmission in an outbreak of enteroinvasive *Escherichia coli*. Am J Epidemiol **122:**245-52.

48. **Harris, J. R., I. K. Wachsmuth, B. R. Davis, and M. L. Cohen.** 1982. High-molecular-weight plasmid correlates with *Escherichia coli* enteroinvasiveness. Infect Immun **37:**1295-8.

49. **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.

50. **Hong, M., and S. M. Payne.** 1997. Effect of mutations in *Shigella flexneri* chromosomal and plasmid- encoded lipopolysaccharide genes on invasion and serum resistance. Mol Microbiol **24:**779-91.

51. **Hsu, W. B., and J. H. Chen.** 2003. The IS1 elements in *Shigella boydii*: horizontal transfer, vertical inactivation and target duplication. FEMS Microbiol Lett **222:**289-95.

52. **Hyma, K. E.** Unpublished data.

53. **Ingersoll, M., E. A. Groisman, and A. Zychlinsky.** 2002. Pathogenicity islands of *Shigella*. Curr Top Microbiol Immunol **264:**49-65.

54. **Jin, Q., Z. Yuan, J. Xu, Y. Wang, Y. Shen, W. Lu, J. Wang, H. Liu, J. Yang, F. Yang, X. Zhang, J. Zhang, G. Yang, H. Wu, D. Qu, J. Dong, L. Sun, Y. Xue, A. Zhao, Y. Gao, J. Zhu, B. Kan, K. Ding, S. Chen, H. Cheng, Z. Yao, B. He, R. Chen, D. Ma, B. Qiang, Y. Wen, Y. Hou, and J. Yu.** 2002. Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. Nucleic Acids Res **30:**4432-41.

55. **Karaolis, D. K., R. Lan, and P. R. Reeves.** 1994. Sequence variation in *Shigella sonnei* (Sonnei), a pathogenic clone of *Escherichia coli*, over four continents and 41 years. J Clin Microbiol **32:**796-802.

56. **Kotloff, K. L., J. P. Winickoff, B. Ivanoff, J. D. Clemens, D. L. Swerdlow, P. J. Sansonetti, G. K. Adak, and M. M. Levine.** 1999. Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. Bull World Health Organ **77:**651-66.

57. **Kumar, S., K. Tamura, I. Jakobsen, and M. Nei.** 2000. MEGA2: Molecular Evolutionary Genetics Analysis Program., 2.0 ed. Pennsylvania State University, University Park, Pennsylvania.

58. **Lan, R., B. Lumb, D. Ryan, and P. R. Reeves.** 2001. Molecular evolution of large virulence plasmid in *Shigella* clones and enteroinvasive *Escherichia coli.* Infect Immun **69:**6303-9.

59. **Lan, R., and P. Reeves.** 2002. *Escherichia coli* in disguise: molecular origins of *Shigella*. Microbes Infect **4:**1125.

60. **Lan, R., and P. R. Reeves.** 2000. Intraspecies variation in bacterial genomes: the need for a species genome concept. Trends Microbiol **8:**396-401.

61. **Lawrence, J. G.** 1999. Gene transfer, speciation, and the evolution of bacterial genomes. Curr Opin Microbiol **2:**519-23.

62. **Lawrence, J. G., and H. Ochman.** 1998. Molecular archaeology of the *Escherichia coli* genome. Proc Natl Acad Sci U S A **95:**9413-7.

63. **Levine, M. M., C. Ferreccio, V. Prado, M. Cayazzo, P. Abrego, J. Martinez, L. Maggi, M. M. Baldini, W. Martin, D. Maneval, and et al.** 1993. Epidemiologic studies of *Escherichia coli* diarrheal infections in a low socioeconomic level peri-urban community in Santiago, Chile. Am J Epidemiol **138:**849-69.

64. **Luck, S. N., S. A. Turner, K. Rajakumar, H. Sakellaris, and B. Adler.** 2001. Ferric dicitrate transport system (Fec) of *Shigella flexneri* 2a YSH6000 is encoded on a novel pathogenicity island carrying multiple antibiotic resistance genes. Infect Immun **69:**6012-21.

65. **Mangold, R. F.** Unpublished data.

66. **Martinez, M. B., T. S. Whittan, E. A. McGraw, J. Rodrigues, and L. R. Trabulsi.** 1999. Clonal relationship among invasive and non-invasive strains of enteroinvasive *Escherichia coli* serogroups. FEMS Microbiol Lett **172:**145-51.

67. **Maurelli, A. T., R. E. Fernandez, C. A. Bloch, C. K. Rode, and A. Fasano.** 1998. "Black holes" and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*. Proc Natl Acad Sci U S A **95:**3943-8.

68. **Maurelli, A. T., and P. J. Sansonetti.** 1988. Identification of a chromosomal gene controlling temperature-regulated expression of *Shigella* virulence. Proc Natl Acad Sci U S A **85:**2820-4.

69. **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:**852-6.

70. **Mead, P. S., L. Slutsker, V. Dietz, L. F. McCaig, J. S. Bresee, C. Shapiro, P. M. Griffin, and R. V. Tauxe.** 1999. Food-related illness and death in the United States. Emerg Infect Dis **5:**607-25.

71. **Mills, M., and S. M. Payne.** 1995. Genetics and regulation of heme iron transport in *Shigella dysenteriae* and detection of an analogous system in *Escherichia coli* O157:H7. J Bacteriol **177:**3004-9.

72. **Monteiro-Neto, V., L. C. Campos, A. J. Ferreira, T. A. Gomes, and L. R. Trabulsi.** 1997. Virulence properties of *Escherichia coli* O111:H12 strains. FEMS Microbiol Lett **146:**123-8.

73. **Moss, J. E., T. J. Cardozo, A. Zychlinsky, and E. A. Groisman.** 1999. The *selC*-associated SHI-2 pathogenicity island of *Shigella flexneri*. Mol Microbiol **33:**74-83.

74. **Motin, V. L., A. M. Georgescu, J. M. Elliott, P. Hu, P. L. Worsham, L. L. Ott, T. R. Slezak, B. A. Sokhansanj, W. M. Regala, R. R. Brubaker, and E. Garcia.** 2002. Genetic variability of *Yersinia pestis* isolates as predicted by PCR-based IS100 genotyping and analysis of structural genes encoding glycerol-3-phosphate dehydrogenase (*glpD*). J Bacteriol **184:**1019-27.

75. **Nataro, J. P., J. Seriwatana, A. Fasano, D. R. Maneval, L. D. Guers, F. Noriega, F. Dubovsky, M. M. Levine, and J. G. Morris, Jr.** 1995. Identification and cloning of a novel plasmid-encoded enterotoxin of enteroinvasive *Escherichia coli* and *Shigella* strains. Infect Immun **63:**4721-8.

76. **Nei, M., and T. Gojobori.** 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol **3:**418-26.

77. **Noriega, F. R., F. M. Liao, S. B. Formal, A. Fasano, and M. M. Levine.** 1995. Prevalence of *Shigella* enterotoxin 1 among *Shigella* clinical isolates of diverse serotypes. J Infect Dis **172:**1408-10.

78. **Oaks, E. V., M. E. Wingfield, and S. B. Formal.** 1985. Plaque formation by virulent *Shigella flexneri*. Infect Immun **48:**124-9.

79. **Ochman, H., and E. A. Groisman.** 1996. Distribution of pathogenicity islands in *Salmonella* spp. Infect Immun **64:**5410-2.

80. **Ochman, H., and E. A. Groisman.** 1994. The origin and evolution of species differences in *Escherichia coli* and *Salmonella typhimurium*. Exs **69:**479-93.

81. **Ochman, H., T. S. Whittam, D. A. Caugant, and R. K. Selander.** 1983. Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*. J Gen Microbiol **129:**2715-26.

82. **Ohnishi, M., J. Terajima, K. Kurokawa, K. Nakayama, T. Murata, K. Tamura, Y. Ogura, H. Watanabe, and T. Hayashi.** 2002. Genomic diversity of enterohemorrhagic *Escherichia coli* O157 revealed by whole genome PCR scanning. Proc Natl Acad Sci U S A **99:**17043-8.

83. **Parsot, C., and P. J. Sansonetti.** 1999. The virulence plasmid of Shigellae: an archipelago of pathogenicity islands?, p. 151-165. *In* J. B. Kaper and J. Hacker (ed.), Pathogenicity islands and other mobile virulence elements. ASM Press, Washington, D.C.

84. **Payne, S. M., D. W. Niesel, S. S. Peixotto, and K. M. Lawlor.** 1983. Expression of hydroxamate and phenolate siderophores by *Shigella flexneri*. J Bacteriol **155:**949-55.

85. **Perna, N. T., G. Plunkett, 3rd, 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:**529-33.

86. **Pope, L. M., K. E. Reed, and S. M. Payne.** 1995. Increased protein secretion and adherence to HeLa cells by *Shigella* spp. following growth in the presence of bile salts. Infect Immun **63:**3642-8.

87. **Pupo, G. M., D. K. Karaolis, R. Lan, and P. R. Reeves.** 1997. Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and mdh sequence studies. Infect Immun **65:**2685-92.

88. **Pupo, G. M., R. Lan, and P. R. Reeves.** 2000. Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. Proc Natl Acad Sci U S A **97:**10567-72.

89. **Purdy, G. E., and S. M. Payne.** 2001. The SHI-3 iron transport island of *Shigella boydii* 0-1392 carries the genes for aerobactin synthesis and transport. J Bacteriol **183:**4176-82.

90. **Radnedge, L., P. G. Agron, P. L. Worsham, and G. L. Andersen.** 2002. Genome plasticity in *Yersinia pestis*. Microbiology **148:**1687-98.

91. **Rajakumar, K., C. Sasakawa, and B. Adler.** 1997. Use of a novel approach, termed island probing, identifies the *Shigella flexneri she* pathogenicity island which encodes a homolog of the immunoglobulin A protease-like family of proteins. Infect Immun **65:**4606-14.

92. **Ram, S., S. Khurana, S. B. Khurana, S. Sharma, and D. V. Vadehra.** 1990. Seasonal fluctuations in the occurrence of enteroinvasive *Escherichia coli* diarrhoea. Indian J Med Res **91:**258-62.

93. **Reeves, S. A., A. G. Torres, and S. M. Payne.** 2000. TonB is required for intracellular growth and virulence of *Shigella dysenteriae*. Infect Immun **68:**6329-36.

94. **Reid, S. D., C. J. Herbelin, A. C. Bumbaugh, R. K. Selander, and T. S. Whittam.** 2000. Parallel evolution of virulence in pathogenic *Escherichia coli*. Nature **406:**64-7.

95. **Robins-Browne, R. M., and V. Bennett-Wood.** 1992. Quantitative assessment of the ability of *Escherichia coli* to invade cultured animal cells. Microb Pathog **12:**159-64.

96. **Rode, C. K., L. J. Melkerson-Watson, A. T. Johnson, and C. A. Bloch.** 1999. Type-specific contributions to chromosome size differences in *Escherichia coli*. Infect Immun **67:**230-6.

97. **Rosa, A. C., M. A. Vieira, A. Tibana, T. A. Gomes, and J. R. Andrade.** 2001. Interactions of *Escherichia coli* strains of non-EPEC serogroups that carry eae and lack the EAF and *stx* gene sequences with undifferentiated and differentiated intestinal human Caco-2 cells. FEMS Microbiol Lett **200:**117-22.

98. **Runyen-Janecky, L. J., S. A. Reeves, E. G. Gonzales, and S. M. Payne.** 2003. Contribution of the *Shigella flexneri* Sit, Iuc, and Feo iron acquisition systems to iron acquisition in vitro and in cultured cells. Infect Immun **71:**1919-28.

99. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol **4:**406-25.

100. **Sansonetti, P. J.** 2001. Rupture, invasion and inflammatory destruction of the intestinal barrier by *Shigella*, making sense of prokaryote-eukaryote cross-talks. FEMS Microbiol Rev **25:**3-14.

101. **Sansonetti, P. J., H. d'Hauteville, C. Ecobichon, and C. Pourcel.** 1983. Molecular comparison of virulence plasmids in *Shigella* and enteroinvasive *Escherichia coli*. Ann Microbiol (Paris) **134A:**295-318.

102. **Sansonetti, P. J., and C. Egile.** 1998. Molecular bases of epithelial cell invasion by *Shigella flexneri*. Antonie Van Leeuwenhoek **74:**191-7.

103. **Sansonetti, P. J., T. L. Hale, G. J. Dammin, C. Kapfer, H. H. Collins, Jr., and S. B. Formal.** 1983. Alterations in the pathogenicity of *Escherichia coli* K-12 after transfer of plasmid and chromosomal genes from *Shigella flexneri*. Infect Immun **39:**1392-402.

104. **Sekizaki, T., S. Harayama, G. M. Brazil, and K. N. Timmis.** 1987. Localization of *stx*, a determinant essential for high-level production of shiga toxin by *Shigella dysenteriae* serotype 1, near *pyrF* and generation of stx transposon mutants. Infect Immun **55:**2208-14.

105. **Sen, A., M. A. Leon, and S. Palchaudhuri.** 1990. Comparative study of attachment to and invasion of epithelial cell lines by *Shigella dysenteriae*. Infect Immun **58:**2401-3.

106. **Shi, Z. Y., M. C. Enright, P. Wilkinson, D. Griffiths, and B. G. Spratt.** 1998. Identification of three major clones of multiply antibiotic-resistant *Streptococcus pneumoniae* in Taiwanese hospitals by multilocus sequence typing. J Clin Microbiol **36:**3514-9.

107. **Siegel, A. F., G. van den Engh, L. Hood, B. Trask, and J. C. Roach.** 2000. Modeling the feasibility of whole genome shotgun sequencing using a pairwise end strategy. Genomics **68:**237-46.

108. **Silva, R. M., M. R. Toledo, and L. R. Trabulsi.** 1980. Biochemical and cultural characteristics of invasive *Escherichia coli*. J Clin Microbiol **11:**441-4.

109. **Snyder, J. D., J. G. Wells, J. Yashuk, N. Puhr, and P. A. Blake.** 1984. Outbreak of invasive *Escherichia coli* gastroenteritis on a cruise ship. Am J Trop Med Hyg **33:**281-4.

110. **Sokurenko, E. V., V. Chesnokova, D. E. Dykhuizen, I. Ofek, X. R. Wu, K. A. Krogfelt, C. Struve, M. A. Schembri, and D. L. Hasty.** 1998. Pathogenic adaptation of *Escherichia coli* by natural variation of the FimH adhesin. Proc Natl Acad Sci U S A **95:**8922-6.

111. **Sokurenko, E. V., D. L. Hasty, and D. E. Dykhuizen.** 1999. Pathoadaptive mutations: gene loss and variation in bacterial pathogens. Trends Microbiol **7:**191-5.

112. **Stephens, C., and W. Murray.** 2001. Pathogen evolution: How good bacteria go bad. Curr Biol **11:**R53-6.

113. **Talukder, K. A., D. K. Dutta, A. Safa, M. Ansaruzzaman, F. Hassan, K. Alam, K. M. Islam, N. I. Carlin, G. B. Nair, and D. A. Sack.** 2001. Altering trends in the dominance of *Shigella flexneri* serotypes and emergence of serologically atypical *S. flexneri* strains in Dhaka, Bangladesh. J Clin Microbiol **39:**3757-9.

114. **Talukder, K. A., M. A. Islam, D. K. Dutta, F. Hassan, A. Safa, G. B. Nair, and D. A. Sack.** 2002. Phenotypic and genotypic characterization of serologically atypical strains of *Shigella flexneri* type 4 isolated in Dhaka, Bangladesh. J Clin Microbiol **40:**2490-7.

115. **Talukder, K. A., Z. Islam, M. A. Islam, D. K. Dutta, A. Safa, M. Ansaruzzaman, A. S. Faruque, S. N. Shahed, G. B. Nair, and D. A. Sack.** 2003. Phenotypic and genotypic characterization of provisional serotype *Shigella flexneri* 1c and clonal relationships with 1a and 1b strains isolated in Bangladesh. J Clin Microbiol **41:**110-7.

116. **Tarr, C. L., T. M. Large, C. L. Moeller, D. W. Lacher, P. I. Tarr, D. W. Acheson, and T. S. Whittam.** 2002. Molecular characterization of a serotype O121:H19 clone, a distinct Shiga toxin-producing clone of pathogenic *Escherichia coli*. Infect Immun **70:**6853-9.

117. **Taylor, D. N., P. Echeverria, O. Sethabutr, C. Pitarangsi, U. Leksomboon, N. R. Blacklow, B. Rowe, R. Gross, and J. Cross.** 1988. Clinical and microbiologic features of *Shigella* and enteroinvasive *Escherichia coli* infections detected by DNA hybridization. J Clin Microbiol **26:**1362-6.

118. **Tetteh, G. L., and L. R. Beuchat.** 2001. Sensitivity of acid-adapted and acid-shocked *Shigella flexneri* to reduced pH achieved with acetic, lactic, and propionic acids. J Food Prot **64:**975-81.

119. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res **25:**4876-82.

120. **Toledo, M. R. F., and L. R. Trabulsi.** 1990. Frequency of enteroinvasive *Escherichia coli* in children with diarrhea and healthy controls, in Sao Paulo, SP, Brazil. Rev Microbiol Sao Paulo **21:**1-4.

121. **Tulloch, E. F., Jr., K. J. Ryan, S. B. Formal, and F. A. Franklin.** 1973. Invasive enteropathic *Escherichia coli* dysentery. An outbreak in 28 adults. Ann Intern Med **79:**13-7.

122. **Turner, S. A., S. N. Luck, H. Sakellaris, K. Rajakumar, and B. Adler.** 2003. Molecular Epidemiology of the SRL Pathogenicity Island. Antimicrob Agents Chemother **47:**727-34.

123. **Viscidi, R. P., and J. C. Demma.** 2003. Genetic diversity of *Neisseria gonorrhoeae* housekeeping genes. J Clin Microbiol **41:**197-204.

124. **Vokes, S. A., S. A. Reeves, A. G. Torres, and S. M. Payne.** 1999. The aerobactin iron transport system genes in *Shigella flexneri* are present within a pathogenicity island. Mol Microbiol **33:**63-73.

125. **Walker, J., and N. Verma.** 2002. Identification of a putative pathogenicity island in *Shigella flexneri* using subtractive hybridisation of the *S. flexneri* and *Escherichia coli* genomes. FEMS Microbiol Lett **213:**257.

126. **Wang, L., H. Curd, W. Qu, and P. R. Reeves.** 1998. Sequencing of *Escherichia coli* O111 O-antigen gene cluster and identification of O111-specific genes. J Clin Microbiol **36:**3182-7.

127. **Wang, R. F., W. W. Cao, and C. E. Cerniglia.** 1997. Phylogenetic analysis and identification of *Shigella* spp. by molecular probes. Mol Cell Probes **11:**427-32.

128. **Wanger, A. R., B. E. Murray, P. Echeverria, J. J. Mathewson, and H. L. DuPont.** 1988. Enteroinvasive *Escherichia coli* in travelers with diarrhea. J Infect Dis **158:**640-2.

129. **Wei, J., M. B. Goldberg, V. Burland, M. M. Venkatesan, W. Deng, G. Fournier, G. F. Mayhew, I. G. Plunkett, D. J. Rose, A. Darling, B. Mau, N. T. Perna, S. M. Payne, L. J. Runyen-Janecky, S. Zhou, D. C. Schwartz, and F. R. Blattner.** 2003. Complete Genome Sequence and Comparative Genomics of *Shigella flexneri* Serotype 2a Strain 2457T. Infect Immun **71:**2775-2786.

130. **Welch, R. A., V. Burland, G. Plunkett, 3rd, P. Redford, P. Roesch, D. Rasko, E. L. Buckles, S. R. Liou, A. Boutin, J. Hackett, D. Stroud, G. F. Mayhew, D. J. Rose, S. Zhou, D. C. Schwartz, N. T. Perna, H. L. Mobley, M. S. Donnenberg, and F. R. Blattner.** 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc Natl Acad Sci U S A **99:**17020-4.

131. **Whittam, T. S., and A. C. Bumbaugh.** 2002. Inferences from whole-genome sequences of bacterial pathogens. Curr Opin Genet Dev **12:**719-25.

132. **Whittam, T. S., M. L. Wolfe, I. K. Wachsmuth, F. Orskov, I. Orskov, and R. A. Wilson.** 1993. Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. Infect Immun **61:**1619-29.

133. **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:**1139-52.