UNIVERSIDAD SAN FRANCISCO DE QUITO

Colegio de Postgrados

Geographical Distribution of Diarrheagenic *Escherichia coli* in Remote Communities in Northwestern Ecuador

María Daniela Castillo Jaramillo

Gabriel Trueba, PhD., Director de Tesis

Tesis de grado como requisito para el título de Magister en Microbiología

Quito, octubre de 2014

UNIVERSIDAD SAN FRANCISCO DE QUITO

Colegio de Postgrados

HOJA DE APROBACIÓN DE TESIS

Geographical Distribution of Diarrheagenic *Escherichia coli* in Remote Communities in Northwestern Ecuador

María Daniela Castillo Jaramillo

Gabriel Trueba, Ph.D. ----------
Director de la Maestría en Microbiología
y Director de Tesis

Verónica Barragan, MSc. ----------
Miembro del Comité de Tesis

Venancio Arahana, Ph.D. ----------
Miembro del Comité de Tesis

Víctor Viteri Breedy, PhD.
Decano del Colegio de Postgrados _____

Quito, octubre de 2014

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído la Política de Propiedad Intelectual de la Universidad San Francisco de Quito y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo de investigación quedan sujetos a lo dispuesto en la Política.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo de investigación en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Firma:

Nombre: María Daniela Castillo Jaramillo

C. I.: 1104357296

Fecha: Quito, octubre de 2014

DEDICATORIA

Dedico esta tesis a Dios y a mi familia, en especial a mis padres que siempre fueron mi soporte y apoyo emocional.

RESUMEN

La diarrea es una de las principales causas de morbilidad y mortalidad en niños a nivel mundial, donde *Escherichia coli* juega un papel importante, especialmente en los países en desarrollo. Nosotros utilizamos “Multilocus sequence typing” (MLST) para describir la relación genética de *E. coli* diarreogénicas aisladas de 498 muestras de heces humanas recolectadas de 22 comunidades vecinas en el noroeste de Ecuador entre Julio de 2012 y Julio de 2013. De un total de 39 cepas de *E. coli* pertenecientes a diferentes patotipos encontramos 19 *Sequence Types* (STs). La mayor diversidad de STs fue encontrado en la principal comunidad, Borbón. Estos resultados confirman estudios previos que muestran evidencia que un pueblo, que era el centro de actividades sociales y comerciales, puede ser también la fuente de patógenos entéricos.

ABSTRACT

Diarrhea is one of the main causes of morbidity and mortality among children in the world with *Escherichia coli* playing an important role, especially in developing countries. We used Multilocus sequence typing (MLST) to describe the genetic relatedness of diarrheagenic *E. coli* isolated from 498 human fecal samples collected from 22 neighboring communities in northwestern Ecuador between July 2012 and July 2013. From a total of 39 *E. coli* strains belonging to different pathotypes we found 19 Sequence Types (STs). The largest diversity of ST was found in the main community, Borbón. These results confirm previous studies showing evidence that a village, which was a center for social and commercial activities, may be also the source of enteric pathogens.

CONTENT

FIGURES AND TABLES

ANNEXES

PART I

GENERAL INTRODUCTION

1. GENERAL INTRODUCTION

1.1Infectious diarrhea

Escherichia coli is a gut commensal bacterium of warm-blooded animals. It typically colonizes the gastrointestinal tract of human infants within a few hours after birth (1). Usually, *E. coli* and its human host coexist in good health acting as a beneficial member of the human microbiome in both digestion and defense against opportunistic pathogens (2), and rarely cause disease except in immunocompromised hosts or where the normal gastrointestinal barriers are breached (1). However, there are some strains that have developed the ability to cause damage even in healthy individuals (3).

Horizontal gen transfer (HGT) has allowed *E. coli* the acquisition of pathogenic characteristics, which enable it to cause disease in healthy hosts (4). In fact, virulence factors acquired for these *E. coli* strains are frequently encoded in mobile genetic elements (MGE), which could move into different strains or remain “locked” into the genome (2). The most successful combinations of virulence factors have persisted to become specific ‘pathotypes’. Three general clinical syndromes can result from infection from one of these pathotypes: enteric/diarrheal disease, urinary tract infections (UTIs) and sepsis/meningitis (1, 5). In fact, *E. coli* is responsible for up to 90% of all human urinary tract infections, and they are a frequent cause of septicemia, gastro-intestinal and other infections (2).

Among the intestinal pathogens, there are well-described categories of *E. coli*: enteropathogenic *E. coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC), enterotoxigenic

E. coli (ETEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli (*EIEC), diffusely adherent *E. coli* (DAEC) (1) and Shigellae.

Despite the fact that *Escherichia coli* and Shigellae have been historically separated into different genera within the Enterobacteriaceae family, it has been suggested to include *Shigella* sp. as part of *E. coli.* In fact, numerous studies (6, 7, 8) have shown that *Shigella* sp. and *E. coli* are taxonomically similar at the species level; suggesting they should be considered a single species. For instance, EIEC is more similar to Shigellae than to typical *E. coli,* and some EIEC strains have essentially all the properties of *Shigella* strains (8). Therefore we refer to all *Shigella* strains as forms of *E. coli*.

Escherichia coli is one of the most important causes of childhood diarrhea (9). It is responsible for 30% of all the deaths caused by diarrhea in children under the age of five in developing countries (2). According to the World Health Organization, diarrhea is one of the principal causes of morbidity and mortality among children in the world, and although mortality rates have declined in the past several decades, it still causes up to 760,000 childhood deaths every year (10).

In the developing countries where there is inadequate clean water and poor sanitation, ETEC is an important cause of diarrhea (11, 12, 13). Among all the etiologic agents of diarrhea, diarrheagenic *Escherichia coli* (DEC) play an important role in low-income countries (14). For instance, two studies conducted in a rural area in Bangladesh (15) and in Egypt (16) showed that ETEC was the most common pathogen found in children. Therefore, serious efforts need to be made in order to improve the awareness of the importance of diarrheagenic *E. coli* mainly in people

living in the developing world.

1.2 Bacterial Molecular Genotyping

The ability to differentiate accurately among related strains (of infectious agents) is vital for epidemiological surveillance and public health decisions, but there are no totally satisfactory methods to achieve this goal (17, 18).

Identifying different types of organisms within a microbial species is called typing, and for a long time, conventional typing schemes, such as bacteriophage typing (based on resistance to a standard set of phages), serotyping (based on differences in antigens), biotyping (according to the metabolic capabilities of the cell) (19) or biochemical typing, have been used to understand the epidemiology of infections caused by strains of clinically relevance (17). Although bacterial phenotyping is useful for specific purposes, this usually does not reflect the genetic relationships of the microorganism (17, 20). In fact, in outbreak investigations a typing method must have the discriminatory power to distinguish all epidemiologically unrelated isolates from the same species (21).

Determining the genetic relatedness of isolates of pathogenic microorganisms is fundamental to molecular epidemiological studies in order to identify sources of pathogens (22). Therefore, conventional phenotyping methods have been largely replaced by genotyping methods because they show lower discriminatory power than molecular methods (20).

Molecular epidemiology employs genetic techniques to characterize or identify clones of infectious agents (23). Different techniques have been developed to compare the

genotypes of microbial species. Some genotyping methods such as gel-based fingerprinting techniques (e.g., ribotyping, MLEE and PFGE) rely on comparison of banding patterns generated by gel electrophoresis (22). PCR-ribotyping is based on polymorphism of the intergenic spacer region located between the 16S and the 23S rRNA genes (24). Multilocus enzyme electrophoresis (MLEE) detects allelic variation of slowly evolving metabolic genes, on the basis of the differing electrophoretic mobility of enzymes (19). Pulsed-field Gel Electrophoresis (PFGE) is based on the digestion of chromosomal DNA with a restriction endonuclease that cleaves infrequently and produces only a few high-molecular-weight fragments (25). Even several studies have shown that techniques such as PFGE resolve isolates that are indistinguishable by MLEE (18) and by ribotipying (26, 27), all these methods can suffer from inter-laboratory variability introducing subjectivity and greater potential for error (18, 22).

To overcome these problems scientists have started to use nucleotide sequence data; for instance, variation at multiple housekeeping loci has become increasingly popular for strain characterization (22, 23). Finally, whole genome sequence (WGS), the ultimate method for bacterial typing, has the potential to resolve single base differences between two genomes provinding high-resolution in genomic epidemiology; however, it cannot completely replace other typing systems due to its cost (21). Thus, it is important to choose carefully the typing method because some typing methods do not work with certain microorganisms or work at different temporal or geographic scales (21).

Multilocus sequence typing (MLST)

Multilocus sequence typing is based on the principles of phenotypic multilocus enzyme electrophoresis (MLEE) (21), and employs nucleotide sequences of internal fragments of seven genes for characterizing isolates (28, 29). MLST is a technique based on the sequence variation of housekeeping genes, which are under stabilizing selection (22, 23), and the variation within these genes is nearly neutral (19). The sequences of each fragment (alleles) are compared with all the previously identified sequences in the MLST database and are assigned allele numbers at each of the seven loci, and every novel sequence is assigned a new allele number by the curator and is entered in the allele database (28). The combination of the seven allele numbers determine the allelic profile which unambiguously defines the sequence type (ST), of each isolate.

Most bacterial species have variation within the seven loci providing many alleles per locus and making it possible to distinguish billions of distinct allelic profiles (18); therefore, it is unlikely that two unrelated isolates have the same allelic profile (30).
This technique provides a number of advantages over other typing approaches. For instance, it is relatively cheap, and easy to perform; therefore, does not require access to specialized reagents or training (28, 31). The fact that it uses sequence data, allows detecting changes at the DNA level not apparent by phenotypic approaches (31). In fact, sequence data are unambiguous in contrast to most tying which involve comparing DNA fragment sizes on gels (22). Moreover, the digital format of MLST data can be held on a central database; therefore, it has facilitated

the establishment of global accessible databases for a variety of organisms that can be queried through a web service rapidly contributing to our understanding of the clonal distribution of infectious disease agents (22, 29). Additionally, it allows the access to specific information as sequencing protocols and primer sequences on the MLS website for each species, allowing data to be produced rapidly (28). Furthermore, the data generated is fully portable among laboratories and can be shared easily throughout the world via the Internet (31, 32). Finally, analysis of multiple loci can provide better evolutionary pictures than the one that can be obtained by single-locus analysis (19).

MLST applications

MLST was first developed for *Neisseria meningitidis* in 1998 to overcome the poor reproducibility between laboratories of older molecular typing schemes (33). Since then, it has been mostly used in molecular epidemiological studies of several bacteria. MLST has been applied for the study of pathogenic bacterial species including *Streptococcus pyogenes* (29, 30*), Staphylococcus aureus, Escherichia coli* (34, 35), *Streptococcus pneumoniae* and *Salmonella* serotypes (35). Therefore, MLST has been applied to a number of different pathogenic and nonpathogenic bacteria as a tool for epidemiological surveillance as well as to study the population structure and evolutionary biology of the species (31, 35).

Unfortunately, the variability of housekeeping genes among different bacteria makes it impossible to develop MLST schemes for some bacteria (36). For example, despite the fact that MLST provides a satisfactory level of discrimination among *Salmonella* isolates, within a particular serovar of *Salmonella* (35), and in Methicillin-Resistant

Staphylococcus aureus (MRSA) typing, MLST may not be suitable for distinguishing related strains due to their low genetic variability (37). As a result, PFGE still remains the gold standard for surveillance and outbreak investigations of these bacteria (35, 37).

In addition, MLST does not provide sufficient discrimination for all typing purposes, including resolving differences among variants of single-clone (recently emerged pathogens), low diversity, asexual pathogens such as *Bacillus anthracis* and *Yersinia pestis* (36). Hence, it is important to have some idea of the population genetics of the pathogen before using as a typing tool (38).

In Enterobactereacea, MLST scheme has been used for the characterization of nosocomial isolates for *Klebsiella pneumoniae* (39). Furthermore, it has been recommended for population biology studies and epidemiological tracking of *Plesiomonas shigelloides* strains which are considered as an emerging pathogen responsible for intestinal diseases and extraintestinal infections in humans and other animals (40). Moreover, there are several sequence-based studies involving the use of MLST scheme to characterize clones or phylogenetic subgroups within *E. coli* (34, 41, 42, 43). However, it would appear that MLST is useful for epidemiologic studies of *E. coli* O78 strains while it is of little value for studies in O157:H7 (41, 42). In fact, the genes selected for the analysis of *E. coli* O157:H7 (a recently emerged pathogen) showed little diversity to be useful as an epidemiological tool (41). Thus, the genes selected for a typing study may have the sufficient variation to be useful in epidemiological investigation.

Therefore, there are three distinct *E. coli* MLST schemes that can be used in order to

establish the relationship among isolates (44). For instance, T. Whittam' s scheme uses up to 15 genes (http://www.shigatox.net); the Pasteur institute scheme, uses 8 genes (http://www.pasteur.fr/recherche/genopole/PF8/mlst/EColi.html), and the Mark Achtman scheme, set of 7 housekeeping genes (http://mlst.warwick.ac.uk). Although probably the most widely scheme used is the Achtman' s scheme.

In this study, for analyzing the distribution of different pathotypes of *Escherichia coli* diarrheagenic, we used the Achtman' s scheme. The seven loci used for *Escherichia coli* were internal fragments of the following genes: adenylate kinase (*adk*), fumarate hydratase (*fumC*), DNA gyrase (*gyrB*), isocitrate/isopropylmalate dehydrogenase (*icd*), malate dehydrogenase (*mdh*), adenylosuccinate dehydrogenase (*purA*), ATP/GTP binding motif (*recA*) (45). Consequently, determining the usefulness of MLST as a molecular epidemiological tool to study the distribution of diarrheagenic *E. coli* in remote communities in northwestern Ecuador could help us to determine not only the relatedness of a group of bacterial isolates, but also the interaction of genetic and environmental factors which may be playing a role in pathogen dissemination of diseases.

REFERENCES

1. Kaper JB, Nataro JP, Mobley HL. 2004. Pathogenic *Escherichia coli*. Nat. Rev. Microbiol. 2(2), 123-140.

2. Kaas RS, Friis C, Ussery DW, Aarestrup FM. 2012. Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. BMC genomics. 13(1): 577.

3. Vieira N, Bates SJ, Solberg OD, Ponce K, Howsmon R, Cevallos W, Trueba G, Riley Lee, Eisenberg JN. 2007. High prevalence of enteroinvasive *Escherichia coli* isolated in a remote region of northern coastal Ecuador. Am J Trop Med Hyg. *76* (3): 528-533.

4. Bayas Rea, R. D. L. A. 2011. Temporal Changes in Prevalence of *Escherichia coli* Pathotypes in Remote Communities of Ecuador.

5. Hommais F, Pereira S, Acquaviva C, Escobar-Páramo P, Denamur E. 2005. Single-nucleotide polymorphism phylotyping of *Escherichia coli.* Appl Environ Microbiol. 71(8): 4784-4792.

6. Ochman H, Whittam TS, Caugant DA, Selander RK. 1983. Enzyme polymorphism and genetic population structure in *Escherichia coli* and Shigella. J Gen Microbiol. 129(9): 2715-2726.

7. Lan R, Alles MC, Donohoe K, Martinez MB, Reeves PR. 2004. Molecular evolutionary relationships of enteroinvasive *Escherichia coli* and *Shigella* spp. Infect Immun. 72(9): 5080-5088.

8. Pupo GM, Karaolis DK, Lan R, Reeves PR. 1997. Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and mdh sequence studies. Infect Immun. 65(7): 2685-2692.

9. Tornieporth NG, John J, Salgado K, de Jesus P, Latham, E, Melo MC, Gunzburg ST, Riley LW. 1995. Differentiation of pathogenic *Escherichia coli* strains in Brazilian children by PCR. J Clin Microbiol. 33(5): 1371-1374.

10. WHO. 2000. Global Water Supply and Sanitation Assessment. World Health Organization. Geneva

11. Svennerholm AM, Holmgren J, Sack DA. 1989. Development of oral vaccines against enterotoxinogenic< i> *Escherichia coli*</i> diarrhea. Vaccine. 7(3): 196-198.

12. Sack DA, Kaminsky DC, Sack RB, Wamola I.A, Orskov F, Orskov I, Slack RC, Arthur RR, Kapikian AZ. 1977. Enterotoxigenic *Escherichia coli diarrhea* of travelers: a prospective study of American Peace Corps volunteers. Johns Hopkins Med J. 141(2): 63-70.

13. Shah N, DuPont HL, Ramsey DJ. 2009. Global etiology of travelers' diarrhea: systematic review from 1973 to the present. Am J Trop Med Hyg. 80(4): 609-614.

14. Lanata CF, Fischer-Walker CL, Olascoaga AC, Torres CX, Aryee MJ, Black RE for the Child Health Epidemiology Reference Group of the World Health Organization and UNICEF. 2013. Global causes of diarrheal disease mortality in children <5 years of age: a systematic review. PLoS One 9 (8): e72788.

15. Black RE, Merson RE, Rahman AS, Yunus M, Alim AR, Huq I, Yolken RH, Curlin GT. 1980. A 2-year study of bacterial, viral and parasitic agents associated with diarrhea in rural Bangladesh. J. Infect. Dis. 142:660–664.

16. Rao MR, Abu-Elyazeed R, Savarino SJ, Naficy AB, Wierzba TF, Abdel-Messih I, Shaheen H, Frenck RW, Svennerholm AM, Clemens JD. 2003. High disease burden of diarrhea due to enterotoxigenic Escherichia coli among rural Egyptian infants and young children. J. Clin. Microbiol. 41:4862-4864.

17. Van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK, Fussing V, Green J, Feli E, Gerner-Smidt P, Brisse S, Struelens M. 2007. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin Microbiol Infect. 13(Suppl3): 1-46.

18. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl Acad. Sci. USA. 95(6): 3140-3145.

19. Margos G, Gatewood AG, Aanensen DM, Hanincová K, Terekhova D, Vollmer SA, Cornet M, Piesman J, Donaghy M, Bormane A, Merrilee AH, Feil EJ, Fish, D, Casjens S, Wormser GP, Schwartz I, Kurtenbach K. 2008. MLST of housekeeping genes captures geographic population structure and suggests a European origin of *Borrelia burgdorferi.* Proc Natl Acad Sci 105(25): 8730-8735.

20. Tenover FC, Arbeit RD, Goering RV. 1997. How to select and interpret molecular strain typing methods for epidemiological studies of bacterial infections: a review for healthcare epidemiologists. Infect Control Hosp Epidemiol. 426-439.

21. Sabat AJ, Budimir A, Nashev D, Sa-Leao R, van Dijl JM, Laurent F, Grundmann H, Friedrich AW, ESCMID Study Group of Epidemiological Markers (ESGEM). 2013. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Euro Surveill. 18(4): 20380.
22. Tartof SY, Solberg OD, Manges AR, Riley LW. 2005. Analysis of an uropathogenic *Escherichia coli* clonal group by multilocus sequence typing. J Clin Microbiol. 43(12): 5860-5864.

23. Tartof SY, Solberg OD, Riley LW. 2007. Genotypic analyses of uropathogenic *Escherichia coli* based on fimH single nucleotide polymorphisms (SNPs). J Med Microbiol. 56(10): 1363-1369.

24. Bidet P, Barbut F, Lalande V, Burghoffer B, Petit JC. 1999. Development of a new PCR- ribotyping method for *Clostridium difficile* based on ribosomal RNA gene sequencing. FEMS Microbiol Lett. 175(2): 261-266.

25. Bidet P, Lalande V, Salauze B, Burghoffer B, Avesani V, Delmée M, Rossier A, Barbut F, Petit JC. 2000. Comparison of PCR-ribotyping, arbitrarily primed PCR, and pulsed-field gel electrophoresis for typing *Clostridium difficile*. J Clin Microbiol. 38(7): 2484-2487.

26. Prevost G, Jaulhac B, Piemont Y. 1992. DNA fingerprinting by pulsed-field gel electrophoresis is more effective than ribotyping in distinguishing among methicillin-resistant *Staphylococcus aureus* isolates. J Clin Microbiol. 30(4): 967-973.

27. Tynkkynen S, Satokari R, Saarela M, Mattila-Sandholm T, Saxelin M. 1999. Comparison of Ribotyping, Randomly Amplified Polymorphic DNA Analysis, and Pulsed-Field Gel Electrophoresis in Typing of *Lactobacillus rhamnosus* and *L. casei* Strains. Appl Environ Microbiol. 65(9): 3908-3914.

28. Aanensen DM, Spratt BG. 2005. The multilocus sequence-typing network: mlst. net. Nucleic Acids res. (Nucleic Acids Res). *33*(Suppl 2): W728-W733.

29. Chan MS, Maiden MC, Spratt BG. 2001. Database-driven multi locus sequence typing (MLST) of bacterial pathogens. Bioinformatics. 17(11): 1077-1083.

30. Enright MC, Spratt BG. 1999. Multilocus sequence typing. Trends in Microbiol. 7(12): 482-487.

31. Pavón ABI, Maiden MC. 2009. Multilocus sequence typing. Methods Mol Biol. 551: 129-14.

32. Clarke SC, Diggle MA, Edwards GFS. 2001. Semiautomation of multilocus sequence typing for the characterization of clinical isolates of *Neisseria meningitidis*. J Clin Microbiol. 39(9): 3066-3071.

33. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Pontén TS, Ussery DW, Aarestrup FM, Lund O. 2012. Multilocus sequence typing of total genome sequenced bacteria. J Clin Microbiol. 50:1355–1361.

34. Poirel L, Lagrutta E, Taylor P, Pham J, Nordmann P. 2010. Emergence of metallo-β-lactamase NDM-1-producing multidrug-resistant *Escherichia coli* in Australia. Antimicrob Agents Chemother. 54 (11): 4914-4916.

35. Harbottle H, White DG, McDermott PF, Walker RD, Zhao S. 2006. Comparison of multilocus sequence typing, pulsed-field gel electrophoresis, and antimicrobial susceptibility typing for characterization of *Salmonella enterica* serotype Newport isolates. J Clin Microbiol. 44(7): 2449-2457.

36. Maiden MC, van Rensburg MJJ, Bray JE, Earle SG, Ford SA, Jolley KA, McCarthy ND. 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. Nat Rev Microbiol. 11: 728-736.

37. Trindade PA, McCulloch JA, Oliveira GA, Mamizuka EM. 2003. Molecular techniques for MRSA typing: current issues and perspectives. Brazilian J Infect Dis. 7(1): 32-43.

38. Peacock SJ, De Silva GDI, Justice A, Cowland A, Moore CE, Winearls CG, Day NPJ. 2002. Comparison of multilocus sequence typing and pulsed-field gel electrophoresis as tools for typing Staphylococcus aureus isolates in a microepidemiological setting. J Clin Microbiol. 40(10): 3764-3770.

39. Diancourt L, Passet V, Verhoef J, Grimont PA, Brisse S. 2005. Multilocus sequence typing of Klebsiella pneumoniae nosocomial isolates. J Clin Microbiol. 43(8): 4178-4182.

40. Salerno A, Delétoile A, Lefevre M, Ciznar I, Krovacek K, Grimont P, Brisse S. 2007. Recombining population structure of *Plesiomonas shigelloides* (Enterobacteriaceae) revealed by multilocus sequence typing. J Bacteriol. 189(21): 7808-7818.

41. Noller AC, McEllistrem MC, Stine OC, Morris Jr, JG, Boxrud DJ, Dixon B, Harrison LH. 2003. Multilocus sequence typing reveals a lack of diversity among *Escherichia coli* O157: H7 isolates that are distinct by pulsed-field gel electrophoresis. J Clin Microbiol. 41(2): 675-679.

42. Adiri RS, Gophna U, Ron EZ. 2003. Multilocus sequence typing (MLST) of *Escherichia coli* O78 strains. FEMS Microbiol Lett. 222(2): 199-203.

43. Jaureguy F, Landraud L, Passet V, Diancourt L, Frapy E, Guigon G, Carbonnelle E, Lortholaty O, Clermont O, Denamur E, Picard B, Nassif X, Brisse S. 2008. Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. BMC Genomics. 9(1): 560.

44. Kaas RS, Friis C, Ussery DW, Aarestrup FM. 2012. Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. BMC Genomics. 13(1): 577.
45. Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MC, Ochman H, Achtman M. 2006. Sex and virulence in *Escherichia coli:* an evolutionary perspective. Mol.Microbiol. 60(5): 1136-1151.

PART II

SCIENTIFIC PAPER

Geographical Distribution of Diarrheagenic *E. coli* in Remote Communities in Northwestern Ecuador

AUTHORS

María Daniela Castillo, [a] Gabriel Trueba, [a] Thamara Andrade, [a] Maria Eugenia Mejía, [a] Veronica Barragán, [a] William Cevallos, [b], Joseph N. S. Eisenberg [c]#

Microbiology Institute, Universidad San Francisco de Quito, Quito, Ecuador [a]; Centro de Biomedicina, Universidad Central del Ecuador, Quito, Ecuador [b]; Department of Epidemiology, University of Michigan, Ann Arbor Michigan, United States [c]

Key Words: Diarrheagenic E. coli, Pathotypes, MLST, clones, clonal complex

#Address correspondence Joseph Eisenberg, University of Michigan, USA, jnse@umich.com

INTRODUCTION

Diarrhea is one of the main causes of morbidity and mortality among children in the world (1). Despite diarrhea mortality has been globally reduced, it is still considered as the second leading cause of death due to infections (1). It causes around 1.7 billion cases every year, and kills approximately 760,000 children under five years old (2). The global median incidence of all types of diarrhea in the developing world is high in children aged less than one year (3). The major etiologic agents of diarrhea are rotavirus (5) and diarrheagenic *Escherichia coli* (DEC) with DEC being the most common agent in children especially in low-income countries (6).

Diarrheagenic *E. coli* are classified in: enteropathogenic (EPEC), enterotoxigenic (ETEC), enteroinvasive (EIEC), enterohemorrhagic (EHEC), enteroaggregative (EAEC), diffusely adherent (DAEC) (7), and Shigellae (8,9). ETEC is the most common cause of *E. coli*-mediated human diarrhea worldwide (10), and it is the most significant cause of diarrhea in developing countries and among travelers (11, 12, 13). In fact, about 50 million children under five years old are at any time colonized with ETEC (3). However, despite it is difficult to estimate the mortality due to ETEC infections; with adequate treatment it should be very low (<1%) (14). In 2006, Ecuador reported that diarrhea was one of the most common causes of hospital visits (4) and 4% of the deaths in children under five years old were caused by diarrhea (15).

In this report we used MLST to document the transmission of 39 diarrheagenic *E. coli* among 22 neighboring communities in a remote region of northwestern coastal Ecuador.

MATERIALS AND METHODS

Study region. The study was carried out in the 22 communities located in northwestern coastal Ecuador in the Canton Eloy Alfaro, Esmeraldas province.

Some of these communities are located along three rivers: the Cayapas, the Santiago and the Onzole, which drain into the main town, Borbón (pop. ~ 8,000 inhabitants). Borbón is considered the commercial and social center of the region. In this community, a random sample of 200 households participated. In the remaining 21 communities, all households were chosen and 98% were enrolled in the study after they consented to participate. Institutional review boards at the Universidad San Francisco de Quito and University of Michigan approved all protocols.

Study design. From July 2012 to July 2013, eight 15-day case-control studies were conducted in the 22 communities. During each visit, fecal specimens were collected for every case of diarrhea in the community; for every case found, 3 asymptomatic control specimens were randomly selected (1 household control and 2 community controls). A case was defined as an individual that had 3 or more loose stools in a 24-hour period and a control was defined as an individual with no signs of diarrhea in the previous 6 days.

Pathogen detection. Stool samples were collected by field staff from cases and controls and tested for the presence of *E. coli* and Shigellae. Fecal samples were plated directly onto Salmonella/Shigella (SS) (BD, New Jersey, USA) agar or MacConkey agar (MKL) (BD, New Jersey, USA); lactose positive colonies were

further cultured in Chromocult® Coliforms Agar (Merck, Darmstadt, Germany). All lactose-negative isolates that were identified by an API-20E test kit (BioMèrieux, Marcy l'Etoile, France) as either *E. coli* or Shigellae, and a random sample of 5 lactose-positive isolates were pooled, suspended in 300µL of sterile distilled water, and boiled for 10 min to release the DNA. The resulting supernatant was used for PCR testing.

Identification of *E. coli* pathotypes was performed with primers designed to identify the heat-labile toxin gene (*eltB*) and heat-stable toxin gene (*estA*) of ETEC, the bundle-forming pilus gene (*bfp*) of EPEC, and the invasion plasmid antigen gene (*ipahH*) of EIEC or Shigellae (Table 1) (16). The PCR amplification protocols were published previously (16).

Multilocus sequence typing (MLST). Bacterial colonies belonging to different pathotypes were submitted to DNA extraction using Dnazol® Reagent (Invitrogen Carlsbad, CA, USA) following the manufacturer's protocol. Fragments of seven genes were amplified and sequenced from all the isolates using the primers in Table 2. The amplification reaction conditions for all the primers were published previously (17), but we introduced the following modifications: The reaction conditions for all the primers were as follows: 2 min at 95°, 30 cycles of 1 min at 95°, annealing at 57° for 1 min, 2 min at 72° followed by 5 min at 72°. The PCR reaction contained 50ng of chromosomal DNA; 0.6µL of 10µM of each primer; 0.2mM of dNPTs; 1X of PCR buffer; 1.25mM of $MgCl_2$; 0.025U of Taq polymerase (Promega, Madison, Wisconsin, U.S.A.), and water to 30µL. The PCR products were sequenced in Functional

Biosciences (http://functionalbio.com/web) and uploaded to the MLST for *E. coli* (http://mlst.warwick.ac.uk) in order to assign them an allelic profile.

RESULTS

Pathotypes. Between July 2012 and July 2013, a total of 445 *E. coli* isolates were analyzed from 498 feces samples collected in the study region. We isolated 39 diarrheagenic *E. coli* (8.76%) from cases and controls in 15 of the 22 studied communities (Table 3). Twenty-one (4.7%) of these isolates were identified as ETEC (12 ETEC-LT, 8 ETEC-ST, 1 ETEC-ST/LT); 11 (2.47%) isolates were identified as EIEC; 5 (1.12%) were Shigellae, and 2 (0.45%) were EPEC. All pathotypes were found in 15 of the 22 communities enrolled in this study, and a higher number of pathotypes was found in Borbón than in the rest of communities (Table 4).

Multilocus sequence typing. In total, 19 MLST profiles (STs) were identified (Table 5). The largest diversity of STs was found in Borbón (9 different STs); 6 of these STs (ST2332, ST4, ST10, ST155, ST279, ST6) were also found in other communities. (Table 5; Figure 1). We also found the STs 328,152, 630 that were present in some communities but they were absent in Borbón (Table 5; figure 2). The most frequent STs were the ST6 and ST4, which comprised the 12,8% (5 of 39) and 10,3% (4 of 39) of the whole population respectively. Thus, while the number of individual STs was low in most of the communities, the greatest proportion of STs was accounted in the main community (Borbón).

We found 2 novel sequence types that were named ST4267 and ST4268 which were isolated in Borbón (July 2012) and in the San Agustin community (March 2013) respectively.

DISCUSSION

In this study the largest number and the greatest diversity of pathotypes and STs were found in Borbón, the economical and social center of the region. Most of the STs found in Borbón (ST2332, ST4, ST10, ST155, ST279, ST6) were also found in smaller and more remote communities (San Francisco, Loma Linda, San Miguel, Guadual, Herradura, Timbire, San Agustin, Colon Eloy, Punta Piedra) (Figure 1). By selecting strains isolated within 12 months we were able to capture the dissemination of *E. coli* pathotypes from the most populated (and more connected) community to the least populated (and more remote) communities.

These results confirm previous studies, which suggested that the commercial and social center of the region (Borbón) was also the source of enteric pathogens for the smaller and more remote communities in the region (8, 18, 19, 20, 21). This phenomenon may result from road access, as Borbón is the main village with larger number of visitants from other cities in Ecuador and the center of social and commercial activities in the region (19, 21). In fact, some studies have shown higher rates of diarrheagenic *E. coli* (9,19), and other intestinal pathogens in Borbón than in the rest of communities in the study region (19). Consequently, in more remote communities, individuals may experience decreased risk due to a lower rate of contact with individuals from outside the region (20, 22).

Finally, it is accepted that strains with identical sequence types (ST) may not be clonal (23, 24) because MLST analyze only 7 housekeeping genes of the whole bacterial genome (25). As result, in this study we found 4 cases where bacteria belonging to distinct pathotypes shared the same ST (ST328 was found in EPEC and

ETEC; ST6 and ST630 were found in EIEC and Shigellae; ST4 was found in ETEC-ST, ETEC-LT, ETEC-ST/LT). We were unable to determine whether these ST coincidences were due to random event or represent different stages of evolutionary process occurring locally, i.e. ETECs that acquire (or lose) genes or EIEC evolving to Shigellae (26, 27, 28).

MLST has already been used to characterize different pathogenic bacteria (29, 30, 31, 32). For instance MLST is useful for epidemiologic studies of extended spectrum beta-lactamase BL-producing *E. coli* strains (33), and *E. coli* O78 strains (34) while it is of little value for studies in O157:H7 (35). This observation may be explained by the fact that *E. coli* O157:H7 (a recently emerged pathogen) is a highly clonal serotype (35). MLST may require different or in some cases, additional loci for different bacteria (33). However, our study shows that MLST could serve as a molecular tool to study dissemination of diarrheagenic *E. coli* among communities.

Financial Support: This project was supported with funds from The Institute of Microbioly of the San Francisco University and by National Institute of Allergy and Infectious Diseases Grants #RO1AI050038 and #K01AI103544.

Acknowledgments: We thank the *Ecología, Desarrollo, Salud y Sociedad* (EcoDess) field team in Esmeraldas for their invaluable contribution in collecting the data, and Diana Muñoz for providing the E. coli strains. We also acknowledge Deysi Parrales for being an essential part of the Microbiology team.

Disclaimer: The authors declare no conflict of interest.

Authors' addresses: María Daniela Castillo, Gabriel Trueba, Thamara Andrade, Maria Eugenia Mejía, Veronica Barragán, Microbiology Institute, Universidad San Francisco de Quito, Quito, Ecuador, William Cevallos Centro de Biomedicina, Universidad Central del Ecuador, Quito, Ecuador. Lixin Zhan and Joseph Eisenberg, School of Public Health, University of Michigan, Michigan, United States.

Author's contributions: María Daniela Castillo conducted the molecular tests of the stool samples and wrote the first manuscript. Gabriel Trueba supervised the research project and reviewed the manuscript. Maria Eugenia Mejía conducted microbiological tests of the stool samples. Thamara Andrade conducted part of the pathotype identification. Verónica Barragán conducted the data analyses and reviewed the manuscript. William Cevallos directed the survey and recruitment in Borbón. Joseph Eisenberg developed the overall study design, revised the manuscript and advised on the analyses. All authors read and approved the final manuscript.

REFERENCES

1. Walker, CLF, Aryee MJ, Boschi-Pinto C, Black RE. 2012. Estimating diarrhea mortality among young children in low and middle income countries. PloS one **7**(1): e29151.

2. World Health Organization, 2013. Fact sheet on Diarrhoeal disease Nº330. Available at: http://www.who.int

3. Wennerås C, Erling V. 2004. Prevalence of enterotoxigenic *Escherichia coli*-associated diarrhoea and carrier state in the developing world. J Health Popul Nutr 370-382.

4. Health Systems Profile Ecuador. Monitoring And Analysis Health Systems Change/Reform. 2008. Third Edition

5. Parashar UD, Gibson CJ, Bresse JS, Glass RI. 2006. Rotavirus and severe childhood diarrhea. Emerg Infect Dis **12**(2): 304.

6. Lanata CF, Fischer-Walker CL, Olascoaga AC, Torres CX, Aryee MJ, Black RE, et al. 2013. Global causes of diarrheal disease mortality in children <5 years of age: a systematic review. PLoS One **8**:e72788.
7. Nataro JP, Kaper JB. 1998. Diarrheagenic *Escherichia coli*. Clin Microbiol Rev **11**(1): 142-201.

8. Vieira N, Bates SJ, Solberg OD, Ponce K, Howsmon R, Cevallos W, Trueba G, Riley L, Eisenberg JN. 2007. High prevalence of enteroinvasive Escherichia coli isolated in a remote region of northern coastal Ecuador. Am J Trop Med Hyg **76**(3): 528-533.

9. Lan R, Alles MC, Donohoe K, Martinez MB, Reeves PR. 2004. Molecular evolutionary relationships of enteroinvasive *Escherichia coli* and *Shigella* spp. Infect and immun **72**(9): 5080-5088.

10. Turner SM, Chaudhuri RR, Jiang ZD, DuPont H, Gyles C, Penn CW., Pallen MJ, Henderson IR. 2006. Phylogenetic comparisons reveal multiple acquisitions of the toxin genes by enterotoxigenic *Escherichia coli* strains of different evolutionary lineages. J Clin Microbiol **44**(12): 4528-4536.

11. Svennerholm AM, Holmgren J, Sack DA. 1989. Development of oral vaccines against enterotoxinogenic< i> *Escherichia col*i</i> diarrhea. Vaccine **7**(3): 196-198.

12. Sack DA, Kaminsky DC, Sack RB, Wamola IA, Orskov F, Orskov I., Slack RC, Arthur RR, Kapikian, AZ. 1977. Enterotoxigenic *Escherichia coli* diarrhea of

travelers: a prospective study of American Peace Corps volunteers. Johns Hopkins Med J **141**(2): 63-70.

13. Shah N, DuPont HL, Ramsey DJ. 2009. Global etiology of travelers' diarrhea: systematic review from 1973 to the present. Am J Trop Med Hyg **80**(4): 609-614.

14. Qadri F, Svennerholm AM, Faruque ASG, Sack RB. 2005. Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. Clin Microbiol Rev **18**(3): 465-483.

15. WHO. "Ecuador: Health Profile." World Health Organization. 2014.15/10/2014. http://www.who.int/countries/ecu/en/.

16. Tornieporth NG, John J, Salgado K, de Jesus P, Latham, E, Melo MC, Gunzburg ST, Riley LW. 1995. Differentiation of pathogenic *Escherichia coli* strains in Brazilian children by PCR. J Clin Microbiol **33**(5): 1371-1374.

17. Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MC, Ochman H, Achtman M. 2006. Sex and virulence in *Escherichia coli:* an evolutionary perspective. Mol Microbiol **60**(5): 1136-1151.

18. Bhavnani D, Goldstick JE, Cevallos W, Trueba G, Eisenberg JN. 2012. Synergistic effects between rotavirus and coinfecting pathogens on diarrheal disease: evidence from a community-based study in northwestern Ecuador. Am J Epidemiol **176**(5): 387-395.

19. Eisenberg JN, Cevallos W, Ponce K, Levy K, Bates SJ, Scott JC, Hubbard A, Viera N, Endara P, Espinel M, Trueba G, Riley LW, Trostle, J. 2006. Environmental change and infectious disease: how new roads affect the transmission of diarrheal pathogens in rural Ecuador. Proc Natl Acad Sci **103**(51): 19460-19465.

20. Zelner JL, Trostle J, Goldstick JE, Cevallos W, House JS, Eisenberg JN. 2012. Social connectedness and disease transmission: social organization, cohesion, village context, and infection risk in rural Ecuador. Am J Public Health **102**(12): 2233-2239.

21. Bates SJ, Trostle J, Cevallos WT, Hubbard A, Eisenberg JN. 2007. Relating diarrheal disease to social networks and the geographic configuration of communities in rural Ecuador. Am J Epidemiol **166**(9): 1088-1095.

22. Eisenberg JN, Goldstick J, Cevallos W, Trueba G, Levy K, Scott J, Percha B, Segovia R, Ponce K, Hubbard A, Marrs C, Forxman B, Smith D, Trostle J. 2012. In-roads to the spread of antibiotic resistance: regional patterns of

microbial transmission in northern coastal Ecuador. J R Soc Interface **9**(70): 1029-1039.

23. Weissman SJ, Johnson JR, Tchesnokova V, Billig M, Dykhuizen D, Riddell K, Rogers P, Oin X, Butler-Wu S, Cookson BT, Fang FC, Scholes D, Chattopadhyay S, Sokurenko E. 2012. High-resolution two-locus clonal typing of extraintestinal pathogenic *Escherichia coli*. Appl Environ Microbiol **78**(5): 1353-1360.

24. Van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, Fry NK, Fussing V, Green J, Feli E, Gerner-Smidt P, Brisse S, Struelens M. 2007. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin Microbiol Infect **13**(s3): 1-46.

25. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl Acad. Sci. USA **95**(6): 3140-3145.

26. Ochman H, Whittam TS, Caugant DA, Selander RK. 1983. Enzyme polymorphism and genetic population structure in *Escherichia coli* and Shigella. J Gen Microbiol **129**(9): 2715-2726.

27. Lan R, Alles MC, Donohoe K, Martinez MB, Reeves PR. 2004. Molecular evolutionary relationships of enteroinvasive *Escherichia coli* and *Shigella* spp. Infect Immun **72**(9): 5080-5088.

28. Pupo GM, Karaolis DK, Lan R, Reeves PR. 1997. Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and mdh sequence studies. Infect Immun **65**(7): 2685-2692.

29. Diancourt L, Passet V, Verhoef J, Grimont PA, Brisse S. 2005. Multilocus sequence typing of Klebsiella pneumoniae nosocomial isolates. J Clin Microbiol **43**(8): 4178-4182.

30. Salerno A, Delétoile A, Lefevre M, Ciznar I, Krovacek K, Grimont P, Brisse S. 2007. Recombining population structure of *Plesiomonas shigelloides* (Enterobacteriaceae) revealed by multilocus sequence typing. J Bacteriol **189**(21): 7808-7818.

31. Peacock SJ, De Silva GDI, Justice A, Cowland A, Moore CE, Winearls CG, Day NPJ. 2002. Comparison of multilocus sequence typing and pulsed-field

gel electrophoresis as tools for typing Staphylococcus aureus isolates in a microepidemiological setting. J Clin Microbiol **40**(10): 3764-3770.

32. Chan MS, Maiden MC, Spratt BG. 2001. Database-driven multi locus sequence typing (MLST) of bacterial pathogens. Bioinformatics **17**(11): 1077-1083.

33. Nemoy LL, Kotetishvili M, Tigno J, Keefer-Norris A, Harris AD, Perencevich EN, Johnson JA, Torpey D, Sulakvelidze A, Morris JG, Stine OC. 2005. Multilocus sequence typing versus pulsed-field gel electrophoresis for characterization of extended-spectrum beta-lactamase-producing *Escherichia coli* isolates. J Clin Microbiol **43**(4): 1776-1781.

34. Adiri RS, Gophna U, Ron EZ. 2003. Multilocus sequence typing (MLST) of *Escherichia coli* O78 strains. FEMS Microbiol Lett **222**(2): 199-203.

35. Noller AC, McEllistrem MC, Stine OC, Morris Jr, JG, Boxrud DJ, Dixon B, Harrison LH. 2003. Multilocus sequence typing reveals a lack of diversity among *Escherichia coli* O157:H7 isolates that are distinct by pulsed-field gel electrophoresis. J Clin Microbiol **41**(2): 675-679.

PART III

TABLES AND FIGURES

TABLE 1. Pathotypes sequence primers

Gene	Primer sequence (a)	Product length:
bfp gene (EPEC)	F 5'CAATGGTGCTTGCGCTTGCT3'	324
	R 5'GCCGCTTTATCCAACCTGGT3'	
lt gene (ETEC)	F 5'GCGACAAATTATACCGTGCT3'	708
	R 5'CCGAATTCTGTTATATATGT3'	
sta gene (ETEC)	5'CTGTATTGTCTTTTTCACCT3'	182
	5'GCACCCGGTACAAGCAGGAT3'	
ipaH gene (EIEC, Shigellae)	5'GCTGGAAAAACTCAGTGCCT3'	424
	5'CCAGTCCGTAAATTCATTCT3'	

(a). F, forward; R, reverse

Reference: Tornieporth, N., *et al*. 1995

TABLE 2. PCR primers for multilocus sequence analysis

Gene (function)	Primer sequence (a)	Amplicon size (bp)
adk (adenylate kinase)	F 5'-ATTCTGCTTGGCGCTCCGGG-3'	583
	R 5'-CCGTCAACTTTCGCGTATTT-3'	
fumC (fumarate hydratase)	F 5'-TCACAGGTCGCCAGCGCTTC-3'	806
	R 5'-GTACGCAGCGAAAAAGATTC-3'	
gyrB (DNA gyrase)	F 5'-TCGGCGACACGGATGACGGC-3'	911
	R 5'-ATCAGGCCTTCACGCGCATC-3'	
icd (isocitrate/isopropylmalate dehydrogenase)	F 5'-ATGGAAAGTAAAGTAGTTGTTCCGGCACA-3'	878
	R 5'-GGACGCAGCAGGATCTGTT-3'	
mdh (malate dehydrogenase)	F 5'-ATGAAAGTCGCAGTCCTCGGCGCTGCTGGCGG-3'	932
	R 5'-TTAACGAACTCCTGCCCCAGAGCGATATCTTTCTT-3'	
purA (adenylosuccinate dehydrogenase)	F 5'-CGCGCTGATGAAAGAGATGA-3'	816
	R 5'-CATACGGTAAGCCACGCAGA-3'	
recA (ATP/GTP binding motif)	F 5'-CGCATTCGCTTTACCCTGACC-3'	780
	R 5'-TCGTCGAAATCTACGGACCGGA-3'	

(a). F, forward; R, reverse

Primers are maintained at http://mlst.warwick.ac.uk

TABLE 3. Percentage of pathotypes (cases and controls)

Pathotype	Number of *E. coli* isolates (n=445)	Percentage (%)
ETEC-ST	8	1.8
ETEC-LT	12	2.7
ETEC-ST/LT	1	0.22
EIEC	11	2.47
Shigellae	5	1.12
EPEC	2	0.45

TABLE 4. Pathotypes of *E. coli* founded per community

Community	Number of diarrheagenic *E. coli* isolates
BORBÓN	13
SAN AGUSTIN	4
COLON ELOY	3
SAN MIGUEL	3
GUADUAL	2
TRINIDAD	2
LOMA LINDA	2
PUNTA DE PIEDRA	2
TELEMBI	2
EL PROGRESO	1
ZANCUDO	1
VALDEZ	1
TIMBIRE	1
HERRADURA	1
SAN FRANCISCO	1

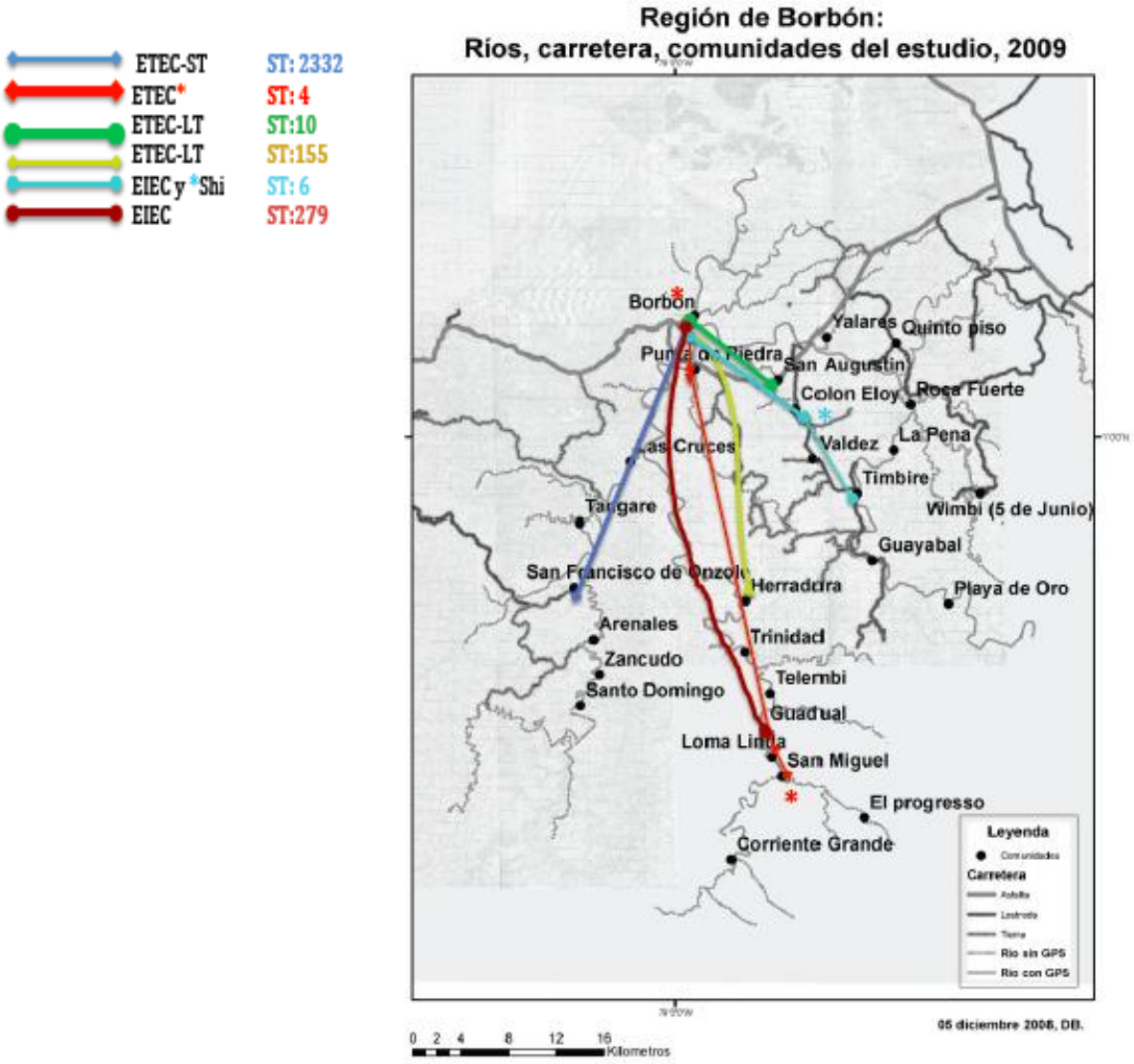
TABLE 5. Sequence type (ST) and ST complex

	Pathotype	Community	Month of Isolation/year	ST	Allelic Profile							ST Complex
					adk	*fumC*	*gyrB*	*icd*	*mdh*	*purA*	*recA*	
1	ETEC-ST	BORBON	07/12	**4267**	6	6	4	16	24	8	7	****
2	ETEC-ST	BORBON	08/12	**4267**	6	6	4	16	24	8	7	****
3	ETEC-ST	COLON ELOY	03/13	**69**	21	35	27	6	5	5	4	st69 Cplx
4	ETEC-ST	VALDEZ	06/13	**2525**	10	11	5	10	12	1	2	****
5	ETEC-ST	BORBON	08/12	**2332**	274	4	96	1	24	8	6	****
6	ETEC-ST	SAN FRANCISCO	01/13	**2332**	274	4	96	1	24	8	6	****
7	ETEC-ST	GUADUAL	10/12	**4**	6	5	4	8	8	8	2	st10 Cplx
8	ETEC-ST	PUNTA DE PIEDRA	01/13	**4**	6	5	4	8	8	8	2	st10 Clpx
9	ETEC-ST/LT	SAN MIGUEL	10/12	**4**	6	5	4	8	8	8	2	st10 Cplx
10	ETEC-LT	BORBON	08/12	**4**	6	5	4	8	8	8	2	st10 Clpx
11	ETEC-LT	BORBON	07/12	**10**	10	11	4	8	8	8	2	st10 Clpx
12	ETEC-LT	SAN AGUSTIN	03/13	**10**	10	11	4	8	8	8	2	st10 Clpx
13	ETEC-LT	SAN AGUSTIN	03/13	**10**	10	11	4	8	8	8	2	st10 Clpx
14	ETEC-LT	BORBON	07/12	**155**	6	4	14	16	24	8	14	st155 Cplx
15	ETEC-LT	HERRADURA	07/13	**155**	6	4	14	16	24	8	14	st155 Cplx
16	ETEC-LT	LOMA LINDA	10/12	**641**	9	6	33	131	24	8	7	st86 Cplx
17	ETEC-LT	LOMA LINDA	10/12	**2758**	10	27	5	8	8	8	2	****
18	ETEC-LT	BORBON	08/12	**3857**	9	6	162	131	24	8	7	****
19	ETEC-LT	TELEMBI	07/13	**2602**	6	95	3	274	9	8	2	****
20	ETEC-LT	SAN AGUSTIN	03/13	**328**	9	23	81	18	11	8	6	st278 Cplx
21	ETEC-LT	TRINIDAD	06/13	**328**	9	23	81	18	11	8	6	st278 Cplx
22	EPEC	EL PROGRESO	10/12	**328**	9	23	81	18	11	8	6	st278 Cplx
23	EPEC	SAN AGUSTIN	03/13	**4268**	142	43	13	455	348	14	137	****
24	EIEC	PUNTA DE PIEDRA	01/13	**99**	6	4	22	18	9	26	7	****

25	EIEC	BORBON	08/12	**279**	6	25	7	75	1	2	2	st280 Cplx
26	EIEC	BORBON	08/12	**279**	6	25	7	75	1	2	2	st280 Cplx
27	EIEC	GUADUAL	10/12	**279**	6	25	7	75	1	2	2	st280 Cplx
28	EIEC	TELEMBI	07/13	**152**	11	63	7	1	14	7	7	st152 Cplx
29	EIEC	TRINIDAD	07/13	**152**	11	63	7	1	14	7	7	st152 Cplx
30	EIEC	TIMBIRE	06/13	**6**	8	7	1	1	10	8	6	****
31	EIEC	BORBON	07/12	**6**	8	7	1	1	10	8	6	****
32	EIEC	BORBON	08/12	**6**	8	7	1	1	10	8	6	****
33	EIEC	COLON ELOY	03/13	**6**	8	7	1	1	10	8	6	****
34	Shigellae	COLON ELOY	03/13	**6**	8	7	1	1	10	8	6	****
35	EIEC	SAN MIGUEL	10/12	**630**	6	61	6	11	6	95	7	****
36	Shigellae	SAN MIGUEL	10/12	**630**	6	61	6	11	6	95	7	****
37	Shigellae	ZANCUDO	11/12	**630**	6	61	6	11	6	95	7	****
38	Shigellae	BORBON	07/12	**245**	6	61	6	11	13	3	50	st245 Cplx
39	Shigellae	BORBON	08/12	**245**	6	61	6	11	13	3	50	st245 Cplx

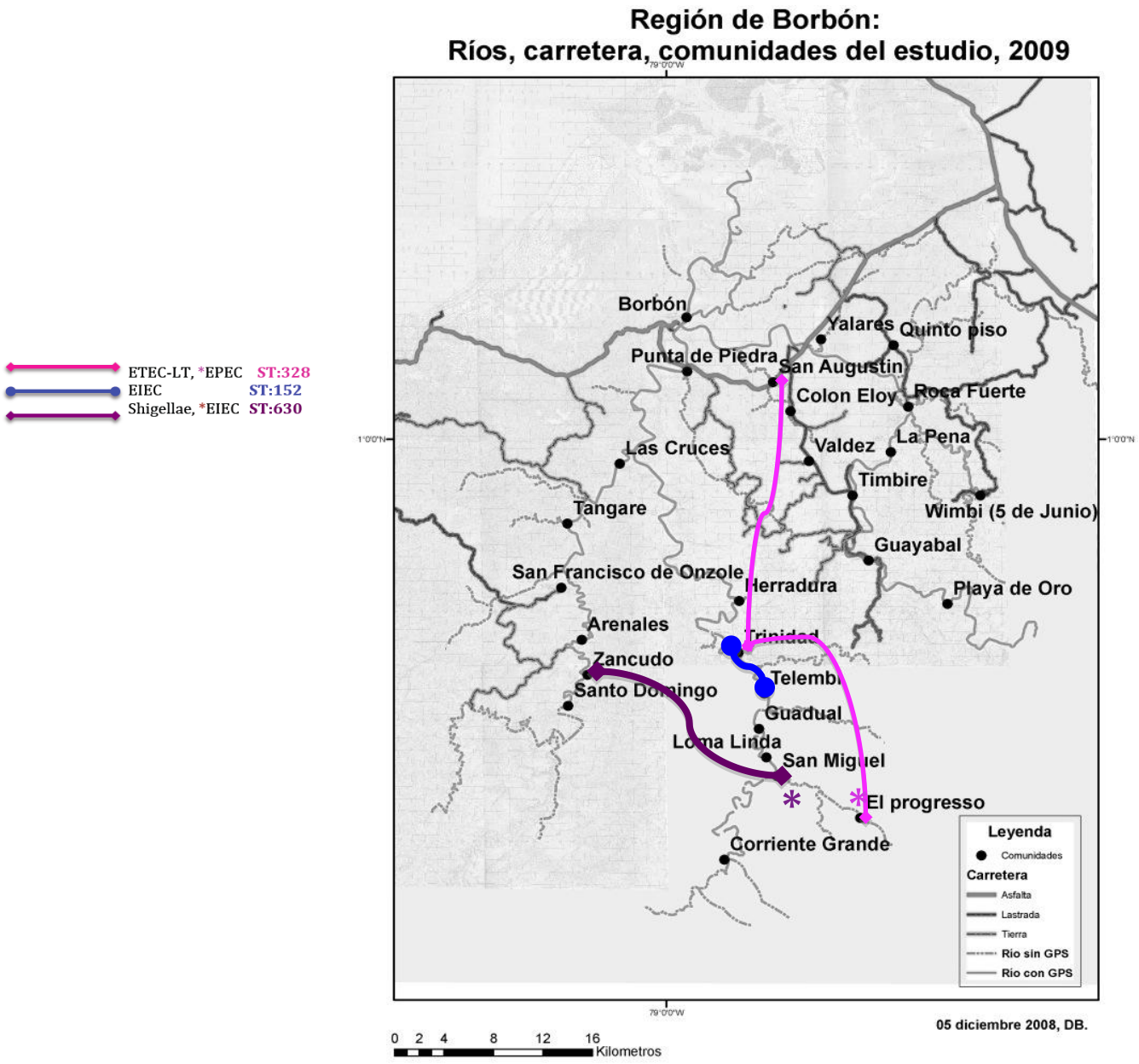
**** NO DATA

FIGURE 1. Distribution of STs on the map of the study region where most of the STs were found in Borbón



(*) ST6 found in Colon Eloy belongs to a Shigellae strain. The (*) ST4 founded in San Miguel and Borbón belong to an ETEC-ST/LT and ETEC-LT, respectively.

Figure 2. Distribution of STs on the map of the study region with the STs that were not found in Borbón



(*) ST328 found in El Progreso belongs to an EPEC strain. (*) One of the two ST4 found in San Miguel, belongs to an EIEC strain.

PART IV

ANNEXES

Appendix 1. Variable nucleotide position among the intern fragments of the 7 housekeeping genes. Numbers in the top row represent the position of the polymorphic base. Dots (.) represent nucleotide identity among all 19 STs

Sequence Type (ST)	*adk* gene																										
	40	145	201	202	205	295	298	316	322	331	334	340	346	358	388	400	424	430	442	454	460	481	493	496	505	514	516	517	526	535
ST4267	T	C	G	T	T	C	C	T	T	A	C	C	T	T	G	C	A	T	C	A	G	C	T	A	C	C	C	C	G	T
ST10	.	.	.	.	.	.	.	C	C	C	T	.	.	.	.	.	.	.	T	G	A	.	A	.	.	.	.	.	A	.
ST155	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST245	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST2332	.	.	.	.	.	T	T	C	C	G	G	T	C	C	.	T	G	C	T	.	A	T	.	G	T	T	.	A	T	C
ST3857	.	.	A	.	.	.	.	C	C	C	T	.	C	.	.	.	.	.	T	G	A	.	A	.	.	.	.	.	.	.
ST279	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST641	.	.	A	.	.	.	.	C	C	C	T	.	C	.	.	.	.	.	T	G	A	.	A	.	.	.	.	.	.	.
ST2758	.	.	.	.	.	.	.	C	C	C	T	.	.	.	.	.	.	.	T	G	A	.	A	.	.	.	.	.	A	.
ST328	.	.	A	.	.	.	.	C	C	C	T	.	C	.	.	.	.	.	T	G	A	.	A	.	.	.	.	.	.	.
ST630	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST99	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST4268	C	.	.	C	C	.	.	.	C	.	.	.	C	.	.	.	.	.	.	G	A	.	.	.	.	.	.	.	.	.
ST6	.	.	.	.	.	T	T	C	C	G	T	T	C	C	.	T	G	C	T	.	A	T	.	G	T	T	.	A	T	.
ST2525	.	.	.	.	.	.	.	C	C	C	T	.	.	.	.	.	.	.	T	G	A	.	A	.	.	.	.	.	A	.
ST2602	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST152	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.
ST69	.	T	.	.	.	.	.	C	.	C	T	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Sequence Type (ST)	*fumC* gene																																						
	14	62	63	80	120	123	152	158	167	194	212	227	248	251	257	278	281	282	296	302	307	311	317	320	323	338	344	350	353	365	368	375	380	383	391	405	416	420	437
ST4267	G	G	C	C	C	G	C	C	A	G	C	G	G	G	G	G	G	C	A	G	G	A	G	T	A	C	C	A	C	C	C	C	A	A	A	G	C	C	C
ST10	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	A	A	G	C	.	T	.	T	T	T	T	.	T	G	G	.	.	.	.	.
ST155	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST245	.	.	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST2332	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST3857	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST279	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	A	A	.	.	.	.	A	.	.	.	.	.	.	.	G	.	.	.	.	.
ST4	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	A	A	G	C	.	T	.	T	T	T	T	.	T	G	G	C	.	.	.	.
ST641	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST2758	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	A	A	G	C	.	T	.	T	T	T	T	.	T	G	G	.	A	.	.	.
ST328	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	G	.	.	.	.	.
ST630	.	.	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST99	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST4268	.	.	.	.	T	.	.	.	.	.	.	.	A	A	.	A	A	.	C	.	.	.	C	C	.	.	.	T	.	.	G	T	G	G	.	.	G	.	.
ST6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	T	.
ST2525	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	T	.	.	.	A	A	G	C	.	T	.	T	T	T	T	.	T	G	G	.	.	.	.	.
ST2602	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	G	.	.	.	.	.
ST152	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.
ST69	.	A	T	T	.	.	.	T	G	.	T	A	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	G	.	.	.	.	T

Sequence Type (ST)	*gyrB* gene																					*purA* gene												
	27	63	117	121	174	180	282	288	291	294	321	330	372	375	387	396	405	411	417	427	444	83	263	278	320	329	359	377	389	420	443	447	470
ST4267	A	C	C	C	C	C	C	C	C	C	C	G	C	G	C	T	C	T	C	T	C	A	G	C	T	G	T	T	C	C	C	C	C
ST10	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST155	.	.	.	.	T	.	.	T	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST245	.	T	T	.	.	.	.	T	.	.	.	A	.	A	.	C	.	.	T	C	.	G	.	.	C	.	.	.	.	.	.	.	A
ST2332	.	.	A	T	.	.	.	.	.	.	.	.	T	.	T	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST3857	T	.	T	.	.	.	T	T	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST279	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	C	.	.	.	.	.	.	.	.
ST4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST641	T	.	T	.	.	.	.	T	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST2758	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST328	.	.	T	.	.	.	.	T	.	.	.	.	.	.	.	C	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.
ST630	.	T	T	.	.	.	.	T	.	.	.	A	.	A	.	C	.	.	T	C	.	G	.	.	C	.	.	.	.	T	.	.	A
ST99	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	G	.	.	.	.	.	.	.	.	T	.	A
ST4268	.	.	.	.	A	T	.	.	T	.	.	.	.	.	.	C	.	.	.	.	.	G	T	T	.	.	C	.	T	.	.	T	.
ST6	.	.	T	.	.	.	.	T	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST2525	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.
ST2602	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST152	.	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	A
ST69	.	.	T	.	A	.	.	.	.	T	.	.	.	.	.	C	T	C	.	.	.	G	.	.	C	T	.	A	.	.	.	.	.

Sequence Type (ST)	*recA* gene											*mdh* gene																
	1	19	25	28	49	52	55	421	451	472	505	3	54	63	85	92	93	138	141	144	147	156	219	237	348	384	396	405	420	429
ST4267	C	C	C	T	G	C	C	C	A	T	C	T	C	G	G	C	C	C	T	T	A	T	T	C	T	T	G	G	A	C
ST10	.	.	.	C	.	.	.	.	G	C	.	C	.	.	.	.	.	.	A	.	.	.	.	.	.	C	.	A	G	.
ST155	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST245	.	.	.	.	.	T	.	.	.	.	.	C	.	.	.	A	.	.	A	.	.	.	.	.	.	.	.	.	.	.
ST2332	.	.	.	.	.	.	.	.	G	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST3857	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST279	.	.	.	C	.	.	.	.	G	C	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
ST4	.	.	.	C	.	.	.	.	G	C	.	C	.	.	.	.	.	.	A	.	.	.	.	.	.	C	.	A	G	.
ST641	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ST2758	.	.	.	C	.	.	.	.	G	C	.	C	.	.	.	.	.	.	A	.	.	.	.	.	.	C	.	A	G	.
ST328	.	.	.	.	.	.	.	.	G	C	.	.	.	.	T	.	.	.	A	.	G	.	.	.	.	.	A	A	.	.
ST630	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
ST99	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	G	.	.	T	.	.	A	A	.	T
ST4268	.	T	.	C	T	.	.	.	G	C	.	C	T	A	.	.	T	T	A	G	T	A	.	.	.	.	.	.	.	.
ST6	.	.	.	.	.	.	.	.	G	C	.	.	.	.	.	.	.	.	A	.	G	.	.	.	.	C	.	A	.	.
ST2525	.	.	.	C	.	.	.	.	G	C	.	.	.	.	.	.	.	.	A	.	G	.	.	.	.	C	.	A	G	.
ST2602	.	.	.	C	.	.	.	.	G	C	.	.	.	.	.	.	.	.	A	.	G	.	.	T	.	.	A	A	.	T
ST152	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	G	.	C	T	.	C	.	A	.	.
ST69	T	.	.	C	.	.	T	T	G	C	T	.	.	.	.	.	.	.	A	.	G	.	.	.	C	.	.	.	.	.