Genetic, In Vitro Phenotypic, and Clinical Characterization of Atypical Enteropathogenic E. coli Infection and Pathogenesis

Sarah E. Kralicek
Loyola University of Chicago Graduate School

Follow this and additional works at: https://ecommons.luc.edu/luc_diss

Part of the Microbiology Commons

Recommended Citation
https://ecommons.luc.edu/luc_diss/4093

This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.
LOYOLA UNIVERSITY CHICAGO

GENETIC, *IN VITRO* PHENOTYPIC, AND CLINICAL CHARACTERIZATION OF ATYPICAL ENTEROPATHOGENIC *E. COLI* INFECTION AND PATHOGENESIS

A DISSERTATION SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL IN CANDIDACY FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

PROGRAM IN BIOCHEMISTRY, MOLECULAR, AND CANCER BIOLOGY

BY

SARAH E. KRALICEK

CHICAGO, ILLINOIS

MAY 2024
ACKNOWLEDGEMENTS

First, I would like to thank my mentor Dr. Gail Hecht for her continued support and guidance. We developed a great trust over the years, continually questioning and pushing me further while also giving me the freedom to explore scientific questions. For her encouragement to return to school and being understanding when life takes over sometimes. For her extensive network of researchers to whom she has introduced me to over the years, allowing me to establish collaborations that will hopefully continue going forward. For honing my writing and communications skills. It has been a pleasure to work with you Gail and I am forever grateful. I would like to thank former members of the Hecht lab who without their help this project would have never gotten where it is today, especially Max Carlino, Lalitha Sitaraman, and Sarah Santiago. In addition, I would like to thank Bill Adams for his responsiveness and careful consideration and review of the statistical approaches applied in this work. I would like to thank Susan Zelisko, from the clinical research office, whose help and efficient extraction of data was essential to completing this work. In addition, I would like to thank Amanda Harrington, from the LUMC Clinical Microbiology lab, for her and the staff’s extensive efforts to save and sort stools for us. I’d also like to thank Majdi Osman and the rest of the OpenBiome team, for their continued effort supplying us with stool samples and data.

I would like to thank my committee members, Dr. Mashkoor Choudhry, Dr. Abby Kroken, Dr. Joe Zackular, and my chair Dr. Nancy Zeleznik-Le, for always asking thoughtful questions and offering helpful discussions regarding my work and for their honest guidance on my future career goals. I would especially like to thank Nancy for her continued support in all
things lab related. To Dr. Mitchell Denning, Associate Dean for Graduate Education, thank you for your understanding nature and your thoughtful insights. I would like to thank the BMCB program, especially the director, Dr. Clodia Osipo for her continual support and effort to bring the department together. To the BMCB program coordinator, Anna Sofinski, thank you for always keeping things organized with a smile on your face.

I am thankful for the large support system created by my loving family and friends. To my brothers, John and Matt, your love, support, and willingness to help is and will always be appreciated. To my in-laws, Ron, Gary and Amy and the kids, and Diane, thank you for your continued support and interest in my work and your devotion to helping with the kids. Special thanks to my mom, Charlotte, and my sister, Maria, for the extra check-ins, little gifts, meal cooking, and the continued help with the kids, without your help I could not have finished.

Of course, special thanks to my husband, Joel, who even though I told him years ago to remind me to never go back to school, fully supported my decision and stuck through it with me. Your pride in my work shows just how much you care and gives me the strength to continue. Your support is also a valuable lesson for our two children, and I love you more for it every day. Most of all I want to thank my children, Ivan and Adelaide, you were my motivation on so many levels. For all the questions of “Are you the boss yet?” to “How many pages now Mom?” or “What’s EPEC?”. You constantly brought me joy on the tough days with your hugs and kept me grounded on the good days by coming together at the end of the day. I am thankful you are such good and patient kids and kept me company by sitting on my lap or just by being in the same room while I worked. You are a major reason I did this, and I hope you have learned from my experience that if you persevere, your hard work will reward you with happiness, a sense of accomplishment, and knowledge that you can do anything you set your mind to.
For my children, Ivan and Adelaide, remember that no matter your age, the pursuit of knowledge and striving to better yourself should always be a part of your life’s journey.
“And just like water on the ground, we will find our way.”

– Chris Robinson
TABLE OF CONTENTS

ACKNOWLEDGEMENTS .................................................................................. iii
LIST OF FIGURES .......................................................... xi
LIST OF TABLES ........................................................... xiv
LIST OF ABBREVIATIONS ........................................................................ xv
ABSTRACT ........................................................................................................ xix

CHAPTER 1: INTRODUCTION ........................................................................... 1

CHAPTER 2: LITERATURE REVIEW ................................................................. 4
  Structure of the Human Intestine and Intestinal Epithelium .......................... 4
  Absorption and Secretion in the Intestine and Intestinal Epithelial Cells ....... 6
  Pathophysiology of Diarrhea ..................................................................... 9
  Etiological Agents of Infectious Diarrhea .................................................. 12
  E. coli Lineages and Pathotypes ................................................................ 15
  Diarrheagenic E. coli .............................................................................. 19
    Enterotoxigenic E. coli ......................................................................... 20
    Enteroaggregative E. coli .................................................................... 21
    Shigella and Enterotoxigenic E. coli ................................................... 22
    Enterohaemorrhagic E. coli and Shiga-toxin producing E. coli ............... 23
    Diffusely Adherent E. coli .................................................................. 25
  Detection and Epidemiology of EPEC ..................................................... 26
  Clinical Manifestations of EPEC Infection .............................................. 31
  Confounding Factors of EPEC Infection ................................................. 34
    Asymptomatic Detection ..................................................................... 34
    Presence of Co-Infecting Pathogens .................................................... 35
    Host Susceptibility Due to Age or Pre-Existing Conditions ................. 36
    Summary of Confounding Factors of EPEC Infection ......................... 38
  Typical EPEC Pathogenesis .................................................................. 38
  Atypical EPEC Pathogenesis .............................................................. 45
    Virulence Gene Profiling of aEPEC .................................................. 46
    Profiling of aEPEC for In Vitro Virulence Phenotypes ......................... 48
    Summary .............................................................................................. 51

CHAPTER 3: CLINICAL MANIFESTATIONS OF ATYPICAL EPEC INFECTION IN
ADULTS AND CHILDREN FROM A US MIDWEST HOSPITAL .......................... 53
  Introduction ............................................................................................. 53
  Materials and Methods ......................................................................... 55
  Case-Control Study ............................................................................... 55
  Collection of Demographic and Clinical Data ......................................... 56
  Diarrhea Severity Score ....................................................................... 58
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>EspZ and EspD</td>
<td>181</td>
</tr>
<tr>
<td>Specific Non-LEE Genetic Factors Associated with Symptomatic aEPEC Infection</td>
<td>182</td>
</tr>
<tr>
<td>EspL2</td>
<td>182</td>
</tr>
<tr>
<td>EspJ</td>
<td>182</td>
</tr>
<tr>
<td>NleE, NleB1, and NleG</td>
<td>183</td>
</tr>
<tr>
<td>Summary</td>
<td>185</td>
</tr>
<tr>
<td>Specific Non-LEE Effectors Associated with Tir Nodes of aEPEC Isolates</td>
<td>186</td>
</tr>
<tr>
<td>NleA and NleF</td>
<td>186</td>
</tr>
<tr>
<td>NleC and NleD</td>
<td>186</td>
</tr>
<tr>
<td>Cif</td>
<td>187</td>
</tr>
<tr>
<td>EspV</td>
<td>187</td>
</tr>
<tr>
<td>EspR, EspX, and EspY</td>
<td>188</td>
</tr>
<tr>
<td>Summary</td>
<td>190</td>
</tr>
<tr>
<td>Adhesins Associated with Symptomatic aEPEC Infection</td>
<td>190</td>
</tr>
<tr>
<td>Emerging Strains of Hybrid Pathogenic <em>E. coli</em></td>
<td>192</td>
</tr>
<tr>
<td>Possible Bacterial and Host Factors Which May Contribute to aEPEC Virulence</td>
<td>194</td>
</tr>
<tr>
<td>Summary</td>
<td>198</td>
</tr>
</tbody>
</table>

REFERENCE LIST ........................................................................................................... 202

VITA ................................................................................................................................. 231
LIST OF FIGURES

Figure 1: BioFire Gastrointestinal Panel (GIP) Data from LUMC .................................................. 28
Figure 2: Localized Adherence and Microcolony Formation by tEPEC .............................................. 39
Figure 3: Features of the Locus of Enterocyte Effacement (LEE) Pathogenicity Island ............... 40
Figure 4: Major Pathogenic Mechanisms of Typical EPEC ................................................................. 43
Figure 5: Comparison of Pathogenic Mechanisms between Typical and Atypical EPEC .......... 46
Figure 6: Cohorts and Main Objectives of Clinical Analysis of EPEC Infection ......................... 55
Figure 7: Analysis of EPEC Infection During International Travel .................................................... 61
Figure 8: Diarrhea Severity Associated with EPEC Infection ............................................................ 62
Figure 9: Symptoms of EPEC Infection ............................................................................................... 63
Figure 10: Abdominal Pain Location and Quality in Individuals Infected with EPEC .......... 64
Figure 11: Laboratory Abnormalities Associated with EPEC Infection ............................................. 66
Figure 12: Other Serological Analysis of Individuals Infected with EPEC .................................. 67
Figure 13: Vital Signs of Individuals Infected with EPEC ................................................................. 68
Figure 14: Contribution of Co-infection Status to Symptoms of EPEC Infection ....................... 69
Figure 15: Analysis of EPEC Co-infections During International Travel ....................................... 70
Figure 16: Hydration and Antibiotic Treatment for EPEC Infections .............................................. 72
Figure 17: Longitudinal Timelines of EPEC Infection ..................................................................... 74
Figure 18: Diagram of the Source and Number of Stools Used in EPEC Load Analysis ........... 78
Figure 19: EPEC Loads Are Higher in SYMP Individuals ................................................................. 81
Figure 20: Final aEPEC Isolates ........................................................................................................ 92
Figure 21: BFP and Per Operon Protein Homology ................................................................. 92

Figure 22: Phylogroup Does Not Predict Symptomatic Status for aEPEC Isolates ............ 93

Figure 23: Core Genome Phylogenetic Tree of aEPEC Isolates ........................................ 95

Figure 24: Heatmap of LEE Protein Homology of aEPEC Isolates to tEPEC/EHEC Reference Strains ........................................................................................................... 96

Figure 25: Differences in LEE Homology of aEPEC Isolates ............................................ 97

Figure 26: Intimin Subtypes are Not Associated with Symptomatic Status. ....................... 98

Figure 27: Tir Nodes Are Significantly Associated with Symptomatic Status ................. 99

Figure 28: Phylogenetic Protein Trees of EspA-B-D, EspH, and EspZ Are Associated with Tir Nodes .................................................................................................................. 101

Figure 29: LEE-Encoded Protein Homology Correlation Matrix ........................................ 102

Figure 30: Heatmap of LEE Protein Homology of aEPEC Isolates to Reference Strains tEPEC/EHEC Clustered According to Tir Nodes ........................................................................... 102

Figure 31: Disease Manifestations of aEPEC Infection Associated with Tir Nodes .......... 107

Figure 32: Non-LEE Virulence Factors Are More Prevalent in SYMP and Tir Node 2 aEPEC Isolates ......................................................................................................................... 108

Figure 33: tEPEC Non-LEE Effectors of aEPEC Isolates ..................................................... 109

Figure 34: Other Non-LEE Effectors of aEPEC Isolates Associated with Tir Node, but Not SYMP Status ................................................................................................................. 111

Figure 35: Correlation Matrix of Non-LEE Effectors Present Versus Clinical Manifestations of aEPEC Infection ........................................................................................................... 112

Figure 36: Adhesins of aEPEC Isolates ................................................................................ 114

Figure 37: Autotransporter Adhesins of aEPEC Isolates ..................................................... 115

Figure 38: Fim Operon (Type I Fimbriae) and Other Fimbrial Protein Genes of aEPEC Isolates ............................................................................................................................... 116

Figure 39: Fimbrial Protein Genes of aEPEC Isolates ......................................................... 117
LIST OF TABLES

Table 1: International Classification of Diseases Codes .......................................................... 56
Table 2: Modified Vesikari Scoring System ............................................................................. 58
Table 3: Symptomatic Subjects per Age Group ....................................................................... 60
Table 4: Risk Factors for EPEC Infection .................................................................................. 60
Table 5: Characterized Adhesin Genes .................................................................................... 84
Table 6: Strain and Clonal Determinant Analysis ..................................................................... 90
Table 7: LEE Lineage and Sequence Type Determination of aEPEC Isolates ......................... 103
Table 8: Non-LEE Effectors and Adhesins Associated with Symptomatic Status .................. 109
Table 9: Autotransporter Adhesin Genes Significantly Associated with Tir Node .................. 115
Table 10: Fimbrial Protein Genes Significantly Associated with Tir Node .............................. 118
Table 11: Summary of Principal Component Analyses ........................................................... 153
LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>aggregative adherence</td>
</tr>
<tr>
<td>afaBC</td>
<td>Dr. family adhesin</td>
</tr>
<tr>
<td>Agn43</td>
<td>antigen 43</td>
</tr>
<tr>
<td>A/E</td>
<td>attaching and effacing</td>
</tr>
<tr>
<td>aEPEC</td>
<td>atypical enteropathogenic <em>Escherichia coli</em></td>
</tr>
<tr>
<td>aPCLV</td>
<td>absolute principal component loading value</td>
</tr>
<tr>
<td>ASBT</td>
<td>apical sodium-dependent bile acid transporter</td>
</tr>
<tr>
<td>ASYM</td>
<td>asymptomatic</td>
</tr>
<tr>
<td>BFP</td>
<td>bundle-forming pilus</td>
</tr>
<tr>
<td>BM</td>
<td>bowel movement</td>
</tr>
<tr>
<td>CD</td>
<td>Crohn’s disease</td>
</tr>
<tr>
<td>CDC</td>
<td>Centers for Disease Control and Prevention</td>
</tr>
<tr>
<td>CFA</td>
<td>colonization factor antigens</td>
</tr>
<tr>
<td>CFU</td>
<td>colony forming units</td>
</tr>
<tr>
<td>CFTR</td>
<td>cystic fibrosis transmembrane regulator family channels</td>
</tr>
<tr>
<td>Cif</td>
<td>cycle-inhibiting factor</td>
</tr>
<tr>
<td>DA</td>
<td>diffuse adherence</td>
</tr>
<tr>
<td>daaC</td>
<td>Dr family adhesin</td>
</tr>
<tr>
<td>DRA</td>
<td>down-regulated in adenoma carcinoma</td>
</tr>
<tr>
<td>DEC</td>
<td>diarrheagenic <em>E. coli</em></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>DAEC</td>
<td>diffusely adherent <em>E. coli</em></td>
</tr>
<tr>
<td>EAEC</td>
<td>enteroaggregative <em>E. coli</em></td>
</tr>
<tr>
<td>ECP</td>
<td><em>E. coli</em> common pilus</td>
</tr>
<tr>
<td>ENaC</td>
<td>epithelium sodium channel</td>
</tr>
<tr>
<td>EPEC</td>
<td>enteropathogenic <em>Escherichia coli</em></td>
</tr>
<tr>
<td>EHEC</td>
<td>enterohaemorrhagic <em>E. coli</em></td>
</tr>
<tr>
<td>ETEC</td>
<td>enterotoxigenic <em>E. coli</em></td>
</tr>
<tr>
<td>ExPEC</td>
<td>extraintestinal pathogenic <em>E. coli</em></td>
</tr>
<tr>
<td>FADD</td>
<td>Fas-associated protein with death domain</td>
</tr>
<tr>
<td>FoodNet</td>
<td>Foodborne Diseases Active Surveillance Network</td>
</tr>
<tr>
<td>Gb3</td>
<td>globotriaosylceramide</td>
</tr>
<tr>
<td>GI</td>
<td>gastrointestinal</td>
</tr>
<tr>
<td>GIP</td>
<td>gastrointestinal panel</td>
</tr>
<tr>
<td>GEMS</td>
<td>Global Enteric Multicenter Study</td>
</tr>
<tr>
<td>GrlA/R</td>
<td>global regulator of LEE activator/repressor</td>
</tr>
<tr>
<td>HCO3-</td>
<td>sodium bicarbonate</td>
</tr>
<tr>
<td>HUS</td>
<td>hemolytic uremic syndrome</td>
</tr>
<tr>
<td>HyPEC</td>
<td>hybrid pathogenic <em>E. coli</em></td>
</tr>
<tr>
<td>IBD</td>
<td>inflammatory bowel disease</td>
</tr>
<tr>
<td>IBS</td>
<td>irritable bowel syndrome</td>
</tr>
<tr>
<td>IEC</td>
<td>intestinal epithelial cell</td>
</tr>
<tr>
<td>Iha</td>
<td>IrgA homologue adhesin</td>
</tr>
<tr>
<td>IV</td>
<td>intravenous</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>LA</td>
<td>localized adherence</td>
</tr>
<tr>
<td>LAL</td>
<td>localized adherence-like</td>
</tr>
<tr>
<td>lda</td>
<td>locus of diffuse adherence</td>
</tr>
<tr>
<td>LEE</td>
<td>locus of enterocyte effacement</td>
</tr>
<tr>
<td>LPS</td>
<td>lipopolysaccharide</td>
</tr>
<tr>
<td>LT</td>
<td>heat-labile toxin</td>
</tr>
<tr>
<td>LUMC</td>
<td>Loyola University Medical Center</td>
</tr>
<tr>
<td>M cells</td>
<td>microfold cells</td>
</tr>
<tr>
<td>MCT1</td>
<td>monocarboxylate transporter 1</td>
</tr>
<tr>
<td>MLST</td>
<td>multi-locus sequence type</td>
</tr>
<tr>
<td>MTP</td>
<td>membrane transport proteins</td>
</tr>
<tr>
<td>mVS</td>
<td>modified Vesikari Score</td>
</tr>
<tr>
<td>NHE</td>
<td>Na$^+$/H$^+$ exchangers</td>
</tr>
<tr>
<td>NHERF2</td>
<td>Na$^+$/H$^+$-exchanger regulatory factor 2</td>
</tr>
<tr>
<td>NKCC1</td>
<td>Na$^+$/K$^+$/2Cl$^-$ cotransporter</td>
</tr>
<tr>
<td>NORS</td>
<td>National Outbreak Reporting System</td>
</tr>
<tr>
<td>NRTK</td>
<td>nonreceptor tyrosine kinases</td>
</tr>
<tr>
<td>OB</td>
<td>OpenBiome and OB isolates from asymptomatic healthy individuals</td>
</tr>
<tr>
<td>paa</td>
<td>porcine A/E associated adhesin</td>
</tr>
<tr>
<td>papC</td>
<td>P fimbriae</td>
</tr>
<tr>
<td>PAT1</td>
<td>putative anion transporter 1</td>
</tr>
<tr>
<td>PCA</td>
<td>principal component analysis</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td>pEAF</td>
<td>EPEC adherence factor plasmid</td>
</tr>
<tr>
<td>PMNL</td>
<td>polymorphonuclear leukocyte</td>
</tr>
<tr>
<td>SCFA</td>
<td>short chain fatty acid</td>
</tr>
<tr>
<td>sfaDE</td>
<td>S fimbriae</td>
</tr>
<tr>
<td>SGLT</td>
<td>sodium/glucose transporters</td>
</tr>
<tr>
<td>ShET1/2</td>
<td>Shigella enterotoxins 1 and 2</td>
</tr>
<tr>
<td>ST</td>
<td>heat-stable toxin</td>
</tr>
<tr>
<td>STEC</td>
<td>Shiga-toxin producing <em>E. coli</em></td>
</tr>
<tr>
<td>Stx1/2</td>
<td>Shiga-toxin 1 and 2</td>
</tr>
<tr>
<td>SYMP</td>
<td>symptomatic</td>
</tr>
<tr>
<td>T3SS</td>
<td>type three secretion system</td>
</tr>
<tr>
<td>tEPEC</td>
<td>typical enteropathogenic Escherichia coli</td>
</tr>
<tr>
<td>TJ</td>
<td>tight junction</td>
</tr>
<tr>
<td>UND</td>
<td>undefined adherence</td>
</tr>
<tr>
<td>UPEC</td>
<td>uropathogenic <em>E. coli</em></td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
</tbody>
</table>
ABSTRACT

Enteropathogenic *E. coli* (EPEC) are enteric pathogens that are non-invasive, lack Shiga toxin, and use a type 3 secretion system encoded on the locus of enterocyte effacement (LEE) pathogenicity island to translocate bacterial effector proteins into host intestinal epithelial cells leading to diarrhea. Atypical EPEC (aEPEC), in contrast to typical EPEC (tEPEC), lack bundle-forming pili, likely altering adherence and pathogenicity. Detection of aEPEC with co-infecting pathogens and in some asymptomatic individuals leads to questions regarding aEPEC virulence, especially in adults. I aimed to characterize the clinical manifestations of aEPEC infection and the genetic and *in vitro* phenotypic factors that contribute to aEPEC virulence.

aEPEC are associated with a wide array of symptoms, ranging from asymptomatic carriage to severe diarrhea with up to 10-40 bowel movements/day and persistent/chronic diarrhea in some. Co-infecting pathogens did not alter diarrhea severity. EPEC loads were higher in symptomatic individuals but did not predict diarrhea severity.

aEPEC isolates from asymptomatic and symptomatic individuals originated from sporadic infections and diverse lineages. Translocated intimin receptor (Tir), an effector involved in intimate host attachment and actin accumulation under attached bacteria termed pedestals, was a major virulence determinant with Tir subtypes correlating with all examined genetic and *in vitro* phenotypic virulence factors. Principal component analyses revealed distinct clusters of aEPEC isolates based on virulence determinants. The most virulent aEPEC isolates were characterized by having the greatest homology to tEPEC/EHEC in EspA, an adhesin and needle protein, and the least homology in Tir and other LEE effectors. These isolates also had the greatest number of non-LEE effectors and adhesins, the greatest adherence and pedestal formation on intestinal epithelial cells, and most robust diarrheal symptoms and...
severity. The least virulent aEPEC isolates had the opposite genetic and phenotypic virulence factors. Those isolates with variable genetic virulence factors correspondingly had variability in phenotypic and clinical manifestations. A subset of aEPEC isolates originating from symptomatic individuals did not fit this trend and likely possess unique pathogenic mechanisms. This is the first study to correlate *in vitro* phenotypes and clinical manifestations with genetic virulence factors of aEPEC isolates from children and adults in the US.
CHAPTER 1
INTRODUCTION

Enteric pathogens are a threat to all with at least 47.8 million cases of infectious diarrhea reported annually.\textsuperscript{1,2} Infectious gastroenteritis is caused by a diverse array of etiologic agents, including diarrheagenic \textit{E. coli}; however, conventional culturing techniques, often fail to identify the specific causative agent or differentiate \textit{E. coli} pathovars. One such pathovar, enteropathogenic \textit{E. coli} (EPEC), is historically defined as a pathogen of malnourished infants in developing nations and is sub-divided into typical (tEPEC) and atypical (aEPEC) by the presence or absence of the virulence associated EPEC adherence factor (EAF) plasmid encoding bundle forming pili (BFP). EPEC is now one of the most commonly detected enteric pathogens, outnumbering other \textit{E.coli} pathotypes representing 6.2\% of total cases and 18-22\% of positive cases of children and adults in the US.\textsuperscript{3,4} Recent data indicates aEPEC is the causative agent for several outbreaks and that occurrence is emerging over tEPEC worldwide among all ages.\textsuperscript{3,5} However, controversy surrounds aEPEC pathogenicity because of several factors: (1) inconsistent reporting of atypical or typical EPEC designation, (2) lack of information regarding symptoms and severity of aEPEC infection in adults, (3) aEPEC asymptomatic carriage in some healthy individuals,\textsuperscript{5,6} and (4) detection of aEPEC in the presence of co-infecting pathogen.\textsuperscript{4,7,8} tEPEC pathogenic mechanisms are mostly understood. In contrast, aEPEC pathogenic mechanisms are largely unknown, and strains have high genomic heterogeneity often harboring virulence factors from other diarrheagenic \textit{E. coli}.\textsuperscript{9} Therefore, the need to discriminate between virulent or less-virulent aEPEC strains is needed. This study proposed to address several of these
unanswered questions by investigating **three major aims**: (1) determine the clinical manifestations of aEPEC infection in a primarily adult population in the US, (2) purify aEPEC isolates from stools of asymptomatic and symptomatic individuals for whole genome sequencing and bacterial genetic risk factor analyses, and (3) assess *in vitro* virulence phenotypes of aEPEC isolates. Correlations were examined between bacterial genetic risk factors and *in vitro* disease-associated phenotypic virulence factors of aEPEC isolates and then both were correlated to clinical manifestation data associated with EPEC infection in children and adults. I hypothesized that aEPEC possess adhesins or virulence factors that compensate for the lack of BFP to initiate attachment and downstream effects leading to diarrhea. Variable clinical manifestations are likely dictated by the specific virulence factors present. The **overarching goal** of this project was to define aEPEC pathogenic mechanisms by examining and correlating aEPEC clinical manifestations to virulent genetic and *in vitro* disease-associated phenotypic factors of aEPEC isolates. aEPEC isolates were collected from and compared between healthy-asymptomatic (OB), hospitalized-asymptomatic (ASYM), and symptomatic (SYMP) individuals. One of the long-term goals of the laboratory is to establish bacterial genetic risk factors which could distinguish virulent from less or avirulent strains of aEPEC in hopes of improving the diagnostic paradigm for EPEC and to better inform treatment strategies and the clinical care of patients.

The vast majority of studies describing EPEC pathogenicity have focused on infantile diarrhea in developing nations and on genetic rather than phenotypic propensity of strains to cause diarrhea or not. In addition, clinical manifestations of aEPEC infection are not well understood, especially in adults. Now that the prevalence of aEPEC in both children and adults in the US is known to be significant, it is important to investigate the clinical significance. This approach was innovative in that both bacterial genomic and *in vitro* phenotypic virulence factors
were analyzed and correlated to clinical manifestations of aEPEC infection from a primarily adult population in a developed nation. Associations between EPEC bacterial load and diarrheal severity were also determined. Together, these data will help define the relevance of EPEC detection in this population.
CHAPTER 2
LITERATURE REVIEW

This study focuses on atypical enteropathogenic *E. coli* (aEPEC), an enteric pathogen for which there is very little information known regarding its pathogenic mechanisms. In general, *E. coli* are characterized by great genetic diversity and plasticity due to a high level of genetic recombination and exchange often acquiring genetic factors from various species in diverse ecological niches. Similarly, virulence factors originating from several different pathogens have been detected in aEPEC. Often orthologous genes are detected in different enteric pathogens and help inform function when it is known in one and not the other. aEPEC are also commonly associated with co-infecting pathogens. Therefore, it is important to understand the different virulence mechanisms of diverse enteric pathogens to gain insight into function of aEPEC virulence factors and into the interplay between microorganisms. Furthermore, the clinical manifestations of aEPEC infection, especially in adults, are largely unknown. Therefore, it is imperative to understand the structure and physiology of the intestine, the etiologies of diarrhea as well as the prevalence and clinical manifestations of diverse enteric pathogens. This will aid in the description of the clinical manifestations caused by aEPEC infection and give perspective and relevance to disease caused by aEPEC infection. These topics and more will be discussed in the following chapter.

**Structure of the Human Intestine and Intestinal Epithelium**

The human intestine is comprised of the small intestine divided into three sections (duodenum, jejunum, ileum) and the large intestine divided into 4 sections (cecum and ascending
colon, transverse colon, descending colon, and sigmoid colon). Both the small and large intestines have mechanisms for secretion and absorption of water, electrolytes, vitamins, and other nutrients. Their absorptive capacity is largely due to the inner most layer of cells closest to the lumen termed the mucosa epithelium. The predominant cell type of the intestinal epithelium are absorptive cells called enterocytes or colonocytes of the small intestine and colon, respectively, with their apical surfaces comprised of characteristic microvilli creating a brush border. The small intestine is covered in finger-like villous projections with deep crypts between villi, increasing surface exposure and absorptive capacity; the small intestine is the major site of digestion and absorption of nutrients and extracts 90% of the water consumed or secreted by the host. The large intestine is essential for absorbing water, vitamins, and electrolytes from indigestible food material passed from the small bowel and contains deep tubular pits or crypts without villi that increase in depth toward the rectum.

Along with the estimated 60 tons of food which pass through the human gastrointestinal (GI) tract within an average lifespan, an abundance of environmental microorganisms transit the GI tract posing a huge threat to the integrity of the gut. However, over thousands of years these microorganisms defined as human gut microbiota have established an intricate and mutually beneficial relationship with the host providing essential nutrients and vitamins, impacting and maintaining immune and metabolic homeostasis in the host, and providing colonization resistance limiting pathobiont expansion and infection by exogenous pathogens.

The intestinal barrier created by specialized epithelial cells allow for commensal bacteria colonization and intestinal homeostasis. The intestinal epithelium is under a constant state of proliferation and differentiation with undifferentiated intestinal epithelial stem cells located within the crypts and differentiated cells spanning the length of villi and tubular pits. In
addition to the absorptive enterocytes and colonocytes, other differentiated cells in the intestinal epithelium include: Paneth cells which act as defense cells secreting lysozyme and other antimicrobial agents, enteroendocrine cells which secrete hormones coordinating digestion, absorption, secretion and appetite, tuft cells which serve a role in chemoreception and potentially parasitic defense, microfold (M) cells involved in antigen sampling and are found covering in the gut associated lymphoid tissue follicles called Peyer’s patches, and goblet cells that serve a protective and lubricating role by secreting mucin.\textsuperscript{14, 21, 22} Adjacent epithelial cells are held together at their apical pole by tight junctions call zonula occludens and are important regulators of transcellular transportation and barrier function.\textsuperscript{20} The diverse functions of the intestinal epithelial cells (IECs) provide a physical and biochemical barrier that segregates host tissue and commensal bacteria essential for maintaining intestinal homeostasis and protection from potentially inflammatory stimuli and enteric pathogens.

**Absorption and Secretion in the Intestine and Intestinal Epithelial Cells**

The main functions of the intestines are to digest and absorb nutrients, vitamins, and water. Absorption is the movement of molecules from the lumen of the GI tract into IECs (enterocytes or colonocytes) and then into the bloodstream or lymph. The GI tract is responsible for handling approximately 8 – 10 L of fluid containing approximately 800 mmol of sodium (Na\textsuperscript{+}), 700 mmol of chloride (Cl\textsuperscript{−}), and 100 mmol of potassium (K\textsuperscript{+}) that passes through the intestinal lumen every day. The majority of this fluid comes from endogenous secretions in the upper intestine (~7.5 L) which is reabsorbed within the small intestine itself; the remaining comes from fluid intake. After absorption in the small intestine, approximately 1.5 – 2.0 L of fluid remains for the colon to absorb resulting in excretion of a stool containing approximately 100 – 200 mL of fluid with <5 mM Na\textsuperscript{+}, 2 mM Cl\textsuperscript{−}, and 9 mM K\textsuperscript{+} that is lost every day.\textsuperscript{23, 24}
Absorption and secretion occur simultaneously through an array of complex transcellular mechanisms involving membrane transport proteins (MTPs) including active and passive ion and solute transporters, exchangers, pumps, and channels. Water molecules surround ions and nutrients in the lumen, and as they are transported by MTPs into IECs they carry water along with them causing water absorption. In addition, driven by osmotic gradients water passively diffuses through the IEC phospholipid bilayer, and increasing evidence indicates the role of specific water channel proteins or aquaporins in osmotically driven transmembrane water movements. Absorption and secretion can also occur through the paracellular route facilitated and controlled by different TJ proteins. Together, these mechanisms result in net water absorption across the intestinal epithelium with MTPs as the major players of this process.

Expression of MTPs varies along the length of the intestine, along the crypt-villus axis, and along the apical-basolateral membrane determined by the function of the intestinal segment as well as the role of the transporter involved. MTPs facilitate absorption of sodium, anions (chloride, sulfate, oxalate), carbohydrates, amino acids and peptides, lipids, lipid and water-soluble vitamins, as well as minerals and micronutrients. In addition, MTPs have direct and indirect roles in ionic balance, fluid absorption and secretion, epithelial cell viability, maintenance of TJs and mucus barrier, microbial distribution, and mucosal blood flow.

There are a vast number of MTPs of the intestinal epithelium, but those involved in active transport of Na\(^+\), Cl\(^-\), and bicarbonate (HCO\(_3^+\)) are the main contributors to net fluid movement. On the basolateral membrane of IECs, resides the Na\(^+\)/K\(^+\)-ATPase pump which is largely responsible for maintaining the electrochemical Na\(^+\) concentration gradient. Using energy from the electrochemical gradient, Na\(^+\) absorption is facilitated by Na\(^+\)/H\(^+\) exchangers (NHEs) in the apical membrane of both the small and large intestine. Three major isoforms have been
identified in IECs, NHE2, NHE3, and NHE8; NHE3 makes the greatest contribution to Na⁺ and water intake. NHEs are coupled to Cl⁻/HCO₃⁻ exchangers for a net result of Na⁺/Cl⁻ absorption in exchange for H⁺/HCO₃⁻ exit. The two main Cl⁻/HCO₃⁻ exchangers are DRA (down-regulated in adenoma carcinoma) expressed throughout the intestine with greater predominance in the colon, and PAT1 (putative anion transporter 1) with more predominance in the small intestine. Influx of Cl⁻ from DRA and PAT1 and from the basal Na⁺/K⁺/2Cl⁻ cotransporter (NKCC1) is counterbalanced by secretion of Cl⁻ through the apically located CFTR (cystic fibrosis transmembrane regulator family channels) and/or by Ca²⁺-activated chloride channels. The colon also harbors the epithelial sodium channel (ENaC), located in the apical membrane, it mediates electrogenic Na⁺ absorption. Also, specific to the colon is the short-chain fatty acid (SCFA) carrier protein MCT1 (monocarboxylate transporter 1) which is a major regulator of colonic sodium transport by enhancing Na⁺/H⁺ exchange. Several cotransporters for amino acids, peptides, vitamin C, and biotin in the small intestine are also coupled to the proton gradient created predominately by NHE3 and the Na⁺/K⁺-ATPase. Similarly, sodium/glucose transporters (SGLT1s) responsible for absorption of the monosaccharides, glucose and galactose, are dependent on the Na⁺ gradient. Together the balance between absorption and secretion of ions works in concert with absorption of water, nutrients, lipids, vitamins, and micronutrients and varies along the length of the GI tract. Secretion of Cl⁻ and HCO₃⁻ are essential for mucus production and for pH balance within the cell and in the lumen. This fine balance helps maintain barrier function, intestinal homeostasis, and regulate intestinal microenvironments. Disruption of this intricate balance between absorption and secretion along the length of the GI tract results in diarrhea of different etiologies which can influence the severity of disease.
Pathophysiology of Diarrhea

Diarrheal disease is characterized by the onset of increased passage of stools that have increased volume and altered consistency. Vomiting and systemic manifestations, such as fever and abdominal cramps or pain, may be present. A stool output of more than 10 g/kg/day in infants and children and 200 g/day in an adolescent or adult is considered abnormal. However, measurement of stools and fluid is impractical and assessment of changes in frequency and consistency are preferred for diagnostic purposes. The World Health Organization (WHO) defines diarrhea as the passage of at least 3 or more unusually loose or watery stools in a 24 hour period but emphasizes that changes in stool consistency rather than frequency are most important. Frequent passing of formed stools is not considered diarrhea.

Diarrhea can be further subdivided into acute, persistent, or chronic. Acute diarrhea typically has an abrupt onset, is self-limited lasting less than 2 weeks, and will resolve with or without medication. Enteric infection is the most common cause of acute diarrhea. Persistent diarrhea lasts for 2-4 weeks and chronic diarrhea is any that lasts for 4 weeks or longer. Important considerations need to be made when diagnosing diarrhea as treatment varies greatly depending on the etiological agent. Also, a wide range of conditions can cause chronic diarrhea with some of the most common including but not limited to irritable bowel syndrome (IBS), inflammatory bowel disease (IBD) including Crohn’s disease (CD) and ulcerative colitis (UC), malabsorption syndromes like celiac disease, food allergy or sensitivity, and chronic infections.

The pathophysiology of diarrhea results from two main types, osmotic or secretory. Osmotic diarrhea occurs when unabsorbable, water-soluble solutes remain in the lumen causing more fluid to passively flow into the lumen down the osmotic gradient. If the luminal water
exceeds the absorptive capacity of the gut, then diarrhea occurs. Osmotically active particles can be present for several reasons: ingestion of solutes, which can include polyethylene glycol, magnesium salts, sodium phosphate, and lactulose which are used as laxatives\textsuperscript{37}; motility disorders related to IBS, gastroschisis, and hyperthyroidism; and sugar intolerances, such as lactose, malabsorption of specific solutes, or over ingestion of high fructose corn syrups and certain sugar substitutes hexitols (sorbitol, mannitol, xylitol).\textsuperscript{37} Damage to the absorptive area of the mucosa, caused by acute gastroenteritis, cow’s milk protein allergy, celiac disease, and Crohn’s disease, can also result in less fluid absorption in osmotic diarrhea. Osmotic diarrhea typically will cease if an individual has fasted.\textsuperscript{24, 37}

Secretory diarrhea occurs when more electrolytes and water are secreted by the bowel than they absorb.\textsuperscript{24, 37} In contrast to osmotic diarrhea, secretory diarrhea will continue even after an individual has fasted.\textsuperscript{24} Diarrhea that originates from the small bowel is often watery in appearance, is of large volume with increased frequency, trace blood may be present and serum white blood cell counts are typically normal with few detected in stools.\textsuperscript{35} One of the most common causes of secretory diarrhea is infection by an enteric pathogen. In secretory diarrhea, absorptive mechanisms are often still in place but are overwhelmed by the amount of fluid secreted. This has been demonstrated for celiac disease, cholera, and toxigenic \textit{Escherichia coli} infection, where inflammatory mediators elicit a net sodium, chloride, and fluid secretion.\textsuperscript{36} However, changes to ion transport and MTPs are largely dependent on the different mechanisms employed by the specific enteric pathogen (as discussed in the section “Etiological Agents of Infectious Diarrhea”).

Inflammatory diarrhea can be characterized by loss of resorptive area, destruction of epithelial cells, leaky tight junctions, and release of inflammatory mediators and products from
immune cells that stimulate fluid secretion. Inflammatory diarrhea is often mucoid or bloody, is of smaller volume than that of small intestinal diarrhea, and grossly bloody stools, white blood cell detection in stool and leukocytosis is often present.\textsuperscript{35} In contrast to secretory diarrhea, the abnormalities in ion transports result from decreased sodium and chloride absorption in the absence of increased chloride secretion and are often times attributed to altered expression and regulation of various MTPs, such as ENaC, NHEs, DRA, and the Na+/K+-ATPase.\textsuperscript{36}

Diarrheal illnesses in the US have an economic burden of $150 million to the health-care sector\textsuperscript{36}, and can have a substantial impact on the quality of life and overall health of an individual ranging from mild diarrhea which can be a mere inconvenience to severe or chronic diarrhea resulting in dehydration, malabsorption, and sometimes death. According to the Global Burden of Disease in 2019, there were ~99 million who suffered from diarrhea worldwide among all ages with ~1.5 million deaths.\textsuperscript{38} Prevalence of diarrheal disease in the US has increased since 1990 with an annual 20% increase. In the US in 2019, there were ~2.47 million cases of diarrhea and 11,272 deaths with the majority of deaths (99.3%) in those aged 20 years or older.\textsuperscript{38} Even acute diarrhea can have long term effects. Linear growth deficits in children aged <5 years from developing nations are associated with bacterial causes of gastroenteritis.\textsuperscript{39} Acute infectious gastroenteritis can result in postinfectious functional GI disorders, with symptoms remaining for at least 8 years\textsuperscript{40}, can lead to IBD\textsuperscript{41}, or reactive arthritis\textsuperscript{42}. Although the etiology and resulting intestinal inflammation has different long-terms effects, gut inflammation has been correlated with carcinogenesis by disrupting gastrointestinal homeostasis.\textsuperscript{43} Together, these studies exemplify the need for more longitudinal studies to assess the long-term effects and disease indexes for acute and persistent/chronic diarrhea of different etiologic agents.
Etiological Agents of Infectious Diarrhea

Infectious diarrhea can be transmitted through many different routes, such as through oral-fecal contamination of waterborne or foodborne sources, person-to-person or animal contact, or environmental contaminations. Infectious diarrhea is caused by a diverse array of etiologic agents, including parasites, viruses, and bacteria, however, conventional culturing techniques often fail to identify the specific etiological agent. The multiplex PCR assay, FilmArray Gastrointestinal Panel (GIP) (BioFire Diagnostics, Salt Lake City, UT) screens diarrheal stool samples for 22 enteric pathogens. This assay was introduced in 2014, is now used in many clinical microbiology labs, and gives clinicians quick, accurate, and simultaneous detection of different diarrheal etiologies. The BioFire GIP assay is indicated as an aid in the diagnosis of specific agents of gastrointestinal illness and results are meant to be used in conjunction with other clinical, laboratory, and epidemiological data. BioFire GIP detects some of the most common causes of infectious diarrhea from both the US and worldwide as discussed here.

Common enteric parasitic infections include *Cryptosporidium*, *Cyclospora cayetanensis*, *Entamoeba histolytica* and *Giardia lambia*. Most parasitic infections arise following ingestion of fecal-contaminated drinking water, recreational water, or food and in rare cases from sexual transmission. Parasitic infections are detected worldwide, however, are more common in tropical, subtropical, and warm temperate regions; infections in the US are typically associated with traveler’s diarrhea. Symptoms of infection vary from asymptomatic carriage to self-limited watery or bloody diarrhea with some infections causing more severe manifestations such as short-term gastroenteritis infecting stomach and biliary ducts, chronic diarrhea, or liver abscesses.
Common enteric viral pathogens include adenovirus (a double-stranded DNA virus), and RNA viruses, such as astrovirus, norovirus, rotavirus, and sapovirus. Transmission is mostly through fecal-oral spread, with some like norovirus also spread through vomitus particles. Most are resistant to chemical and physical damage and have low infectious dose and therefore, are highly contagious. Infections with viral pathogens are more commonly associated with vomiting than bacterial pathogens; individuals also experience diarrhea, nausea, and fever. Infections typically are self-limited, lasting as short as 24-48 hrs for norovirus and up to 5-12 days for adenovirus. Norovirus is the leading cause of foodborne gastroenteritis in the US; however, rotavirus is the leading cause of severe diarrheal illnesses in infants and young children worldwide.

Common bacterial non-E.coli pathogens include gram-positive Clostridium difficile, and gram-negative Campylobacter jejuni, Salmonella, Vibrio, and Yersinia entercolitica. Of note, Vibrio or cholera infections are rare in the US but can be associated with traveler’s diarrhea. The gram-negative enteric pathogens are most commonly recognized as foodborne pathogens acquired through ingestion of contaminated food or water, often in under cooked meats like beef, pork, poultry, in unpasteurized dairy products or in unwashed fresh produce. In addition, Salmonella is found in a variety of animal reservoirs, such as bovine, ovine, porcine, poultry, dogs, cats, and reptiles. In contrast, C. difficile are spore-forming gram-positive bacteria that can be acquired from the environment or transmitted via the fecal-oral route, however, antibiotic treatment and hospitalization or residence in an elderly care facility are the major risk factors for infection. Symptoms of illness range from self-limited acute watery diarrhea sometimes with blood, fever, or abdominal pain with few having nausea or vomiting, to copious volumes of diarrhea to more serious complications such as terminal ileitis, mesenteric lymphadenitis, and
pseudoappendicitis\textsuperscript{47, 49}, systemic infections\textsuperscript{46}, long term health issues such as Guillain-Barré syndrome and reactive arthritis\textsuperscript{44}, dehydration leading to death, or severe and or fulminate cases of \textit{C. difficile} infection marked by leukocytosis, abdominal distention, and hypovolemia, or by hypotension or shock, ileus or megacolon.\textsuperscript{50} Despite these robust clinical manifestations, asymptomatic carriage particularly of \textit{Campylobacter} and \textit{C. difficile} has been detected. Diarrhea is caused by toxin production in \textit{C. difficile} and \textit{Vibrio} infection, and by invasion of enterocytes, M cells or macrophages among other immune cells and production of cytotoxins in some during infection by \textit{Campylobacter}, \textit{Salmonella}, and \textit{Yersinia}. Toxin production and cellular invasion elicit robust immune responses in the host and are the main cause of diarrhea in these infections except for cholera toxin.\textsuperscript{46, 51, 52} Cholera toxin via cyclic AMP increases CFTR activity and decreases NHE abundance with an overall net increase of both sodium and chloride in the lumen causing diarrhea.\textsuperscript{32} Others have been shown to affect NHE3, DRA, ENaC, CFTR, and Na\textsuperscript{+}/K\textsuperscript{+}-ATPase\textsuperscript{32, 48} \textit{Vibrio} houses one or two T3SS with differential functions depending on the species and strain, often enhancing virulence through secreted effector proteins.\textsuperscript{46, 52-54} \textit{Salmonella} and \textit{Yersinia} invasive properties are dependent on delivery of effector proteins into host cells via type 3 secretion systems (T3SS), while in \textit{Vibrio} T3SS only enhances virulence.\textsuperscript{46, 49, 52} T3SS will be discussed in more detail in CHAPTER 2 “Typical EPEC Pathogenic Mechanisms”.

The etiological agents that cause diarrhea and gastroenteritis are diverse. Although many symptoms are similar, there are also many distinguishing clinical manifestations resulting from the specific canonical virulence factors present in each pathogen. These virulence factors differentially affect ion transport, the host immune system, and epithelial structure and integrity. Conventional culturing techniques are often costly, time consuming, and fail to identify the etiological agent. Thus, the use of multiplex PCR diagnostic tests, such as the BioFire GIP assay,
have greatly improved the diagnostic capacity of clinicians. However, careful consideration of travel history, food consumption, and antibiotic use as well as pre-existing conditions and medications still needs to be implemented as detection of DNA from pathogenic organisms in stool alone does not indicate causation. Despite this drawback, the diagnostic power of multiplex PCR assays is especially important for determining the presence of different *E. coli* pathovars as conventional culturing systems cannot distinguish between commensal *E. coli* and the diverse array of pathogenic *E. coli*, including diarrheagenic *E.coli* and *Shigella*.24

**E. coli Lineages and Pathotypes**

*Escherichia coli* was first described by Theodor Escherich in 1885 as a Gram-negative rod-shaped bacterium in healthy individuals.55 *E. coli* is from the family *Enterobacteriaceae* and grows both aerobically and anaerobically, preferably at 37°C, and can be either nonmotile or motile, with peritrichous flagella.56 *E. coli* have many different sub-types with great genetic diversity and plasticity indicated by the relatively small core genome (<2000 genes) and the large pan-genome (>8000 genes)57 with commensal and disease-associated *E. coli* subtypes containing different genetic profiles.58, 59

*E. coli* is a predominant species of the normal human microbiota found in 75-90% of individuals with up to ~5% relative abundance.57, 60 *E. coli* grows in a complex, multi-species biofilm in the mucus layer that surrounds the gut epithelium where it competes for an array of nutrients.61 Interestingly, *E. coli* residency in the gut, in contrast to other more stable commensals, is dynamic with specific strains able to become residents while others are transient and lost at the rate of gut transit.62 Despite its high turnover, commensal *E. coli* have a symbiotic relationship with its host by producing vitamin K and B12 and preventing colonization of the intestine by pathogenic bacteria (colonization resistance).63 *E. coli* can tolerate acid and bile
salt\textsuperscript{64}, produce short chain fatty acids\textsuperscript{64}, produce indole at higher levels \textit{in vitro} than other commensal bacteria\textsuperscript{65, 66}, have two pathways to produce active vitamin B\textsubscript{6}\textsuperscript{67}, and certain commensal strains have the ability to promote TJ structure and function\textsuperscript{68, 69}. Thus, commensal \textit{E. coli} have the potential to contribute greatly to intestinal homeostasis and health of the host.

In contrast, pathogenic \textit{E. coli} have acquired mobile genetic elements with specific virulence factors that allow them to replicate and survive in the host often to the detriment of the host causing inflammation, cell and tissue damage, and a variety of diseases in healthy individuals, including watery diarrhea, dysentery, sepsis and meningitis, hemolytic uremic syndrome, and urinary tract infections. \textit{E. coli} can readily be isolated from fecal samples by plating on selective media such as MacConkey agar which is selective for Gram-negative bacteria and differentiates between those that are positive- or negative-lactose fermenters. In addition to lactose fermentation, morphological distinctions aid in the differentiation of \textit{E. coli} from other \textit{Enterobacteriaceae} species such as \textit{Klebsiella, Salmonella, Proteus, Yersinia} and others.\textsuperscript{70} Despite this effective isolation tool, traditional culture techniques which are often laborious, time-consuming and not cost effective, cannot distinguish avirulent from pathogenic \textit{E. coli}. Therefore, in an attempt distinguish commensal \textit{E. coli} from the many different pathogenic strains, various molecular based sub-typing and lineage markers have been established throughout the years.

Although more types of \textit{E. coli} classification systems have been developed, historically three major classifications schemes have been used: serotyping, phylogroup, and Achtman multi-locus sequence type (MLST). These systems were originally developed as techniques involving bacterial isolation and wet bench assays (agglutination assay or single/multiplex PCR) but can now be readily applied to whole genome sequencing data \textit{in silico}. Serotyping is based on the
method developed by Kauffmann in the 1940s involving agglutination of bacteria based on presence of specific immunogenic structures: the lipopolysaccharide (LPS) (O antigen), the capsular antigen (K), and the flagellar (H) antigen. Since then, O:H serotyping has become the gold standard for *E. coli* with at least 188 O groups (O1-O188) and 53 H groups (H1-H56) included in the scheme. Currently, instead of agglutination assays, web-based servers, such as SeroTypeFinder made available by the Center for Genomic Epidemiology, perform serotyping of whole genome sequence data based on homology and small nucleotide polymorphism (SNP) analysis of O-antigen genes *wzx*, *wzy*, *wzm*, and *wzt* for *in silico* O typing and the flagellin genes *fliC*, *flkA*, *flmA*, *flmA*, and *flfA* for *in silico* H typing. A designation of NM or H− indicates an absence of the H antigen, and that the isolate is nonmotile. An OX designation indicates that the O-serotype has not yet been confirmed. The number of *E. coli* serotypes are continually increasing with only a small subset consistently associated with disease. Therefore, other molecular tools were also developed such as single/multiplex PCR based assays that are now also done routinely *in silico*, including phylogroup and MLST classifications, to better distinguish commensal from pathogenic *E. coli*.

*E. coli* phylogroups were first determined using restriction-fragment length polymorphisms (RFLP) based on electrophoretic analysis of a 35-enzyme digest to classify the ECOR-*Escherichia coli* Reference Collection into six groups (A–F). Clermont and colleagues went on to develop a PCR based method that similarly grouped *E. coli* strains into six groups (A–F) and a newly identified G phylogroup based on presence/absence and SNPs of specific genes (*aceK_arpA*, *chuA*, *yjaA*, *tspE4.C3*, *arpAgpE*, *trpBA*, *ybgD*, and *trpAgpC*). Using this scheme, the EzClermont web application is available for *in silico* phylogroup designation. Similarly, Achtman MLST is based on SNPs found in seven housekeeping genes (*adk*, *fumC*, *gyrB*, *icd*,...
mdh, purA, recA) distributed through the E. coli chromosome and has been found to encompass almost all E. coli strains typed. However, they often fail to distinguish various pathogenic E. coli that have different infections sites, present with different clinical manifestations, and house distinct virulence factors. This abundance of variation has led to a classification system of pathovars or pathotypes of pathogenic E. coli based on certain genetic virulence factors and on pathologies and virulent phenotypes generated in the host or on cultured epithelial cells. E. coli pathotypes can be broadly grouped as extraintestinal pathogenic E. coli (ExPEC), intestinal pathogenic E. coli (of which six are known as diarrheagenic E. coli (DEC)), and hybrid pathogenic E. coli (HyPEC) containing virulence factors from different pathovars (as reviewed and discussed in CHAPTER 8).

Extraintestinal pathogenic E. coli (ExPEC) do not cause diarrhea and often translocate from the gut bacteria to cause infections in other sites of the human body. ExPEC is the most reported microorganism to cause hospital-associated diseases in the US between 2011 – 2017 and is also acquired in long-term care facilities. ExPEC cause infections such as urinary tract infection, skin and soft tissue infection, abdominal and pelvic infection, pneumonia, and meningitis, among others. Notably, ExPEC is also the most frequent cause of sepsis and bacteremia. Pathotypes of the ExPEC group include uropathogenic E. coli (UPEC), neonatal and meningitis-associated E. coli, and sepsis-associated E. coli, among others. Canonical pathovar-specific virulence factors have yet to be elucidated; however, specific virulence genes are often found in ExPEC subtypes. It has been proposed that detection of at least 2 of the following genes, in the absence of other established pathovar genes, is likely indicative of
ExPEC (including UPEC strains): P fimbriae (papA/H, papC), S and F1C fimbriae (sfa/focDE), Dr antigen-specific adhesin (afa/dra), aerobactin iutA, and group 2 capsules kpsMII; this was the only multivariable combination which correlated with disease outcome in a murine sepsis model. Much work is needed to understand the pathogenic mechanisms of the diverse group of ExPEC strains. Interestingly, P fimbriae (papA/H, papC), S and F1C fimbriae (sfa/focDE), and Dr antigen-specific adhesin (afa/dra) have all been detected in aEPEC isolates as well as some strains of diffusely adherent E. coli.

Diarrheagenic E. coli

Diarrheagenic E. coli (DEC) strains include enteropathogenic E. coli (EPEC), enterotoxigenic E. coli (ETEC), enteroaggregative E. coli (EAEC), Shigella/enteroinvasive (EIEC), enterohaemorrhagic E. coli (EHEC) and Shiga-toxin producing E. coli (STEC), and diffusely adherent E. coli (DAEC). Several different pathotypes of DEC (EPEC, Shigella/EIEC, and EHEC/STEC) employ a type 3 secretion system (T3SS) to translocate bacterial effectors directly into host epithelial cells subverting host cell physiology. Similar to Salmonella, Yersinia, and Vibrio, T3SS and effectors are major contributors to pathogenicity of Shigella, EIEC, EHEC and some STEC strains, however they are not the only virulence factors; instead, these subtypes also employ toxins or invasive strategies.

In contrast, EPEC do not produce secreted toxins like C. difficile, Vibrio, EHEC/STEC and ETEC, nor are they invasive like Campylobacter, Salmonella, Yersinia, and Shigella/EIEC. Instead, EPEC virulence is based solely on the T3SS and effectors which enable them to cause localized lesions by attaching tightly to the surface of IECs, disrupting cell surfaces, and finally leading to effacement of the microvilli. This characteristic attaching and effacing (A/E) lesion formation by EPEC in concert with other virulence mechanisms leads to diarrhea. Initial
attachment of EPEC to host cells is largely due to bundle-forming pili (BFP) expressed on the bacterial surface. However, not all EPEC produce BFP. Thus, EPEC are sub-divided into typical (tEPEC) and atypical (aEPEC) by the presence or absence of BFP, respectively. Many questions remain regarding the clinical significance of aEPEC infection and has led to increased interest in deciphering the pathogenic mechanisms of this emerging pathogen. Current data regarding the epidemiology, clinical manifestations, and pathogenic mechanisms of tEPEC and aEPEC are discussed in further detail in CHAPTER 2, section “Detection and Epidemiology of EPEC” and beyond. The following information regarding other DEC pathotypes is included to highlight the diverse pathogenic mechanisms of DEC which are largely dependent on specific canonical virulence factors.

**Enterotoxigenic E. coli**

Enterotoxigenic *E. coli* (ETEC) is a major cause of traveler’s diarrhea as well as childhood diarrhea in developing countries. ETEC infection arises after ingestion of fecal contaminated food and water, as well as person-to-person contact. ETEC infection manifests as watery diarrhea, sometimes accompanied with headache, fever, nausea, or vomiting, with symptoms typically lasting no longer than 5 days. Severe dehydration and lack of protective immunity has led to some lethal cases in children. ETEC pathogenic mechanism is largely explained by adherence to host IEC and secretion of enterotoxins. Adherence of ETEC to the small bowel epithelial mucosa is mediated by at least 26 fimbrial adhesive structures collectively called “coli surface antigens” (CS1-CS26) or historically as “colonization factor antigens” (CFAs), of which CFAI/II were the first two fimbrial types discovered. Upon adherence and replication on the mucosa, ETEC secrete two potent enterotoxins, heat-labile (LT) and heat-stable (ST) toxins, responsible for secretory diarrhea. Both LT and ST induce activation of the
CFTR channel to release Cl⁻ and water into the lumen through increases in cGMP and cAMP, respectively. In addition, intracellular cAMP activates protein kinase A which phosphorylates and inhibits NHE3, thus Na⁺ absorption is also diminished during ETEC infection. ST also may stimulate HCO₃⁻ secretion in a CFTR-independent but DRA or PAT1-dependent mechanism. The net result of LT and ST is hypersecretion of electrolytes and water into the lumen which manifests as watery diarrhea in the host.

**Enteroaggregative E. coli**

Enteroaggregative E. coli (EAEC) infection is commonly associated with self-limited watery diarrhea, often mucoid, with or without blood and abdominal pain, nausea, vomiting, and low fever. Some may develop persistent diarrhea (≥14 days), especially children or adults who are immunocompromised. EAEC infection was first associated with infections of children in developing countries but outbreaks and sporadic infections in adults in both developed and developing nations have been documented often associated with travel to developing countries and ingestion of contaminated food or water. Asymptomatic carriage of EAEC is also common and virulence is likely dictated by different strains of EAEC. This is reflected in a human volunteer study, in which high doses (10ⁱ⁰ CFU) of 4 different EAEC were given; only one EAEC sub-type (EAEC 042_O44:H18) caused diarrhea in 3 out of 5 subjects and symptoms reflected secretory diarrhea and 2 subjects had mucoid diarrhea. EAEC pathogenic mechanisms are not well understood but involve abundant adhesion to the intestinal mucosa facilitated by at least 5 aggregative fimbriae (AAFI-V) encoded on the pAA plasmid. AAFs along with other adhesins are responsible for the characteristic *in vitro* and *ex vivo* stack-brick pattern of adherence with bacteria adhering to each other, on the surface of epithelial cells, and in the absence of cells. Secretion of excessive mucus has been described with mucoid biofilm
formation important for EAEC infection. After biofilm formation, EAEC produces several enterotoxins including Pet, EAST-1 and ShET1 causing mucosal destruction such as microvillus vesiculation, enlarged crypt openings, and increased epithelial cell extrusions.\textsuperscript{79, 80} Lastly, an inflammatory response ensues, as demonstrated in both clinical and laboratory studies.\textsuperscript{80} The toxins EAST-1 and ShET1 produced by EAEC may be involved in aberrant ion transport due to their homology to ST of ETEC and Shigella enterotoxin 1, respectively.\textsuperscript{80} However, the precise mechanisms are yet to be determined.

**Shigella and Enterotoxigenic E. coli**

*Shigella* consists of four species including *S. dysenteriae*, the first to be isolated and described to cause bacillary dysentery in 1889 but is rarely isolated in modern day infections, *S. flexneri*, the most prevalent worldwide, followed by *S. sonnei* the most prevalent in the US, and *S. boydii*, hardly ever detected outside of the Indian subcontinent.\textsuperscript{82} Shigella infection results in severe mucosal and bloody diarrhea with abdominal cramps and fever which clinically has been termed shigellosis. Fecal-oral transmission from individuals with symptomatic infection is the most common mode of transmission, however outbreaks due to contamination of food, water, or untreated recreational water and sexual transmission occur.\textsuperscript{82} *Shigella* and *E. coli* genomes have an average nucleotide identity of >98% causing some to argue that they should be grouped within the same genus. In fact, enteroinvasive *E. coli* (EIEC) discovered in 1940 causes very similar clinical symptoms and houses nearly all the same virulence factors as *Shigella*. The bacteria also share similar characteristics such as the inability to ferment lactose, the lack of lysine decarboxylase, the lack of motility, the ability to invade epithelial cells without adherence or flagella factors, and the presence of an invasion plasmid (pINV) responsible for intracellular lifestyle.\textsuperscript{55} Invasion into host cells is a distinct characteristic compared to other *E. coli*
pathotypes, but similar to *Campylobacter, Salmonella, and Yersinia*. Despite their similarities with *Shigella*, EIEC strains primarily elicit watery diarrhea and cause invasive inflammatory colitis instead of severe shigellosis and are biochemically distinguishable from *Shigella* by their enhanced ability to ferment mucate and utilize serine, xylose, or sodium acetate, metabolic traits more similar to *E. coli*.\(^5\) Although *Shigella/EIEC* have at least two enterotoxins, *Shigella* enterotoxins 1 and 2 (ShET1/ShET2), which contribute to intestinal secretory activity virulence, these toxins are not present in every strain and their role during infection is minor compared to the role of the T3SS and more than 25 bacterial effector proteins.\(^8\)

**Enterohaemorrhagic *E. coli* and Shiga-toxin producing *E. coli***

The major virulence determinant of enterohaemorrhagic *E. coli* (EHEC) and Shiga-toxin producing *E. coli* (STEC) is the production of two toxins, Shiga toxin 1 (Stx1) and Shiga toxin 2 (Stx2). Stx1 is nearly identical to *S. dysenteriae* toxin ShET, whereas Stx2 is only about 57% homologous. Two almost simultaneous reports in 1983 described an outbreak in the US of *E. coli* serotype O157:H7 which caused bloody diarrhea in patients and the other described sporadic infections of *E. coli* of various serotypes associated with hemolytic uremic syndrome (HUS). Thus, the terms EHEC and STEC have been used separately and synonymously throughout the literature. EHEC is almost exclusively associated with serotype O157:H7, and in addition to Stx1/Stx2, it also houses a T3SS and effectors encoded on the locus of enterocyte effacement (LEE) pathogenicity island. STEC on the other hand is associated with non-O157 serotypes, can be positive or negative for the LEE, and can produce either Stx1, Stx2, or both toxins. Stx2 is considered to be the more toxic of the two, and in general, more severe disease is associated with EHEC O157:H7 than non-O157 STEC strains. Following exposure of EHEC/STEC, symptoms ensue within 1-3 days and high risk STEC (O157 strains or Stx2-positive STEC) may present first
with fever, abdominal pain, and vomiting before the onset of initially non-bloody diarrhea which precedes to bloody diarrhea approximately 1 - 3 days after diarrhea onset. Symptoms are typically self-limited, lasting up to 7 days, however, HUS, if it occurs, is usually established 5 to 13 days after diarrhea onset. Many variants of Stx1/Stx2 have been described with some variants linked to more severe disease. This may be due to varying enzymatic activity of its A subunit or differential binding affinity of subunit B for globotriaosylceramide (Gb3), a membrane bound glycosphingolipid involved in the cell membrane structure and lipid raft microdomains. Upon entry and retrograde transport to the ER, Stx subunit cleavage occurs. The A subunit translocates to the cytosol, where it inactivates the ribosome capacity of protein synthesis by its N-glycosidase activity, triggering pro-inflammatory cell stress and apoptosis. The colon is the primary infection site for EHEC/STEC where inflammation causes greater permeability of Stx through the intestinal barrier allowing for systemic circulation. Increased inflammation is then apparent in local endothelial cells and distal sites throughout the body. In addition to toxin production, EHEC and certain STEC can form A/E lesions leading to a loss of absorptive capacity of the microvilli contributing to diarrhea and disease severity. The T3SS and the vast repertoire of effectors encoded on the LEE pathogenicity island and on non-LEE genetic elements facilitate persistence, host-to-host spread, and virulence of EHEC/STEC. At least 62 LEE and non-LEE effectors and homologues have been described in EHEC/STEC. Excluding toxin production, EHEC and tEPEC have the most similar pathogenic mechanisms of the different DEC. Therefore, EHEC effectors will not be described here as many overlap with effectors involved in tEPEC pathogenesis and will be highlighted later.

EHEC/STEC infects both children and adults with children having increased risk of developing HUS. EHEC/STEC are zoonotic bacteria, largely distributed in the GI tract of a wide
array of animals, normally as asymptomatic carriers, such as livestock species, domestic and wild mammals, birds, and fish. Cattle and several other ruminant animals are considered the main natural reservoir of EHEC/STEC with sporadic and outbreak infections worldwide linked to direct contact with animals or their environment, person-to-person spread, and contamination in the food supply chain, especially involving beef products, contamination of water used for drinking or swimming, and contamination of water and fertilizer (manure) used for irrigation and growing leafy and fresh produce. In contrast to many of the other DEC strains, active surveillance, and reporting of EHEC/STEC infections are mandatory in the US and other nations worldwide.

**Diffusely Adherent *E. coli***

Diffusely adherent *E. coli* (DAEC) is the most recent DEC pathotype to be distinguished and is defined by a diffuse pattern of adherence pattern on cultured epithelial HEp-2 cells and are often associated with the adhesins Afa/Dr/F1845. DEAC infections are associated with symptoms ranging from asymptomatic carriage to watery diarrhea in children and adults worldwide with an increase in severity in children from the age of 18 months to 5 years; some children can develop persistent diarrhea. DEAC strains harboring Afa/Dr/F1845 adhesins which bind two specific receptors, DAF and CEACAM, located in M-cells, enterocytes, and undifferentiated crypt cells. Binding elicits primary and secondary IL-8 responses resulting in polymorphonuclear leukocyte (PMNL) migration, TNF-α and IL-1β cytokine production, upregulation of DAF expression facilitating greater bacterial binding, and eventually induces PMNL apoptosis. Afa/Dra adhesins binding also induce lipid raft mobilization promoting cell membrane finger-like projections around attached bacteria. Similar to tEPEC, DAEC remain extracellular and cause epithelial cell damage, loss of adherens junctions, and loss of
brush border microvillus structure decreasing absorptive capacity.\textsuperscript{84, 85} In addition, the secreted autotransporter toxin (Sat) induces damage on epithelial cells and alters TJ integrity.\textsuperscript{84, 85} Although this mechanism has been characterized, Afa/Dr adhesins are not always present in diffusely adhering \textit{E.coli} and they can be found in some EAEC and EPEC strains.\textsuperscript{84, 85} In addition DAEC are often detected in asymptomatic individuals leading to controversy about DAEC as a causative agent of diarrhea and the epidemiology of DEAC remains unclear. Therefore, questions remain about how to effectively differentiate DAEC from other pathovars and underscores the plasticity of virulence genetic profiles among \textit{E. coli} pathotypes. The emergence of DEAC and controversy surrounding its pathogenicity is reflective of circumstances surrounding aEPEC emergence and controversy regarding its virulence.

\textbf{Detection and Epidemiology of EPEC}

In 1955 the term enteropathogenic \textit{E. coli} (EPEC) was coined and distinguished it as the cause of many outbreaks in the US and UK of watery, infantile diarrhea between 1889-1955.\textsuperscript{86, 87} Today, EPEC infection remains a leading cause of diarrhea-related deaths of infants in developing nations.\textsuperscript{88} However, the presence of EPEC in developed countries and in adults has been underestimated due to the lack of rapid and routinely used screening tools as well as a lack of reporting or severe underreporting. In the US, there are currently two federal surveillance systems run by the Centers for Disease Control and Prevention (CDC) to track enteric infections, the Foodborne Diseases Active Surveillance Network (FoodNet) and the Foodborne Disease Outbreak Surveillance System which in 2009 merged with the National Outbreak Reporting System (NORS).\textsuperscript{89, 90} FoodNet conducts active surveillance of 10 states representing 15\% of the US population through surveys of laboratories, physicians, and the general population as well as performing population based epidemiological studies.\textsuperscript{90} FoodNet currently conducts surveillance
for 8 enteric pathogens including STEC, but importantly, not EPEC. NORS collects information from state and local health departments about enteric and non-enteric disease outbreaks of many different types, such as waterborne, foodborne, or person-to-person transmission and others. Similar to Illinois, in most states and US territories, the only E. coli pathotype mandated to be reported is STEC. Although EPEC may fall under the category of “foodborne or waterborne outbreaks” that are mandated reportable diseases, many local agencies do not have the means to determine if EPEC is the cause of an outbreak. Therefore, many times EPEC infection likely goes unreported. In fact, two meta-analyses of the prevalence of gastrointestinal pathogens (1980-2008) revealed that EPEC incidence in developed nations is not reported or severely underreported compared to other pathogens, especially in studies of adults. Therefore, until recently, very little information has been available about the numbers of EPEC in sporadic or outbreak infections.

Epidemiological data of EPEC, thus, has largely been dependent on individual academic or medical studies. Two meta-analyses which examined prevalence of enteric pathogens from 1980-2008 revealed that EPEC were detected in only 0.1-1.3% of diarrheal samples from adults in developed nations. In addition, a US specific study from 2002-2004 found EPEC in 4.3% of samples. Molecular epidemiological typing of EPEC from Australian patients from 2008-2011 revealed 6.8% (368/5430) of diarrheal specimens contained detectable EPEC DNA. Studies such as these prior to 2014 predominantly used conventional culturing techniques followed by direct PCR to identify EPEC. However, with the advent of multiplex PCR based detection, such as BioFire GIP, differentiation between diarrheagenic E. coli pathovars is possible and thus EPEC epidemiological data has increased in recent years. GIP samples are deemed EPEC positive if the eae gene that encodes intimin is detected and if the sample is
negative for Shiga toxin and EHEC antigen. A multicenter evaluation of GIP from 4 geographical distinct sites in the US revealed that EPEC were the top detected pathogen being present in as many as 22% (348/1556) of diarrheal samples indicating its increased detection levels. Our recent study determined that EPEC was the 2nd most frequently detected pathogen just after C. difficile in stool samples tested by Loyola University medical center (LUMC) clinical microbiology labs (Figure 1). EPEC was present in 6.2% of all samples tested, representing 18.3% (1229/6734) of enteric-positive specimens (Figure 1). EPEC was detected equally compared to all other E. coli pathovars combined (Figure 1).

![Figure 1: BioFire Gastrointestinal Panel (GIP) Data from LUMC. (A) Number of diarrheal specimens collected between Feb 2016 and Dec 2019 that are positive for the indicated enteric pathogen. Top pathogens, C. difficile (C. diff), EPEC, Norovirus (Noro), and enteroaggregative E. coli (EAEC), and other E. coli pathovars, enterotoxigenic E. coli (ETEC), Shiga-toxigenic/enterohemorrhagic E. coli (STEC/EHEC), and Shigella/enteroinvasive E. coli (EIEC), are reported. Percentage of positive samples is indicated for each pathogen. (Adapted from: Carlino, MJ*, Kralicek SE*, Santiago SA, et al. Gut Microbes. 2020 Nov 9;12(1):1-21.)](image)

Similarly, in a UK study of adults hospitalized with suspected gastroenteritis, EPEC was the 2nd most detected pathogen at 10.2% (14/137). In a Canadian study of children and adults with acute gastroenteritis, EPEC was the 2nd most detected pathogen found in 2.56% (2/78) of all samples tested by GIP and 8.7% (2/23) of enteric-pathogen positive samples. In a study from
2019-2020, EPEC was the 4th most detected pathogen in 7.6% (14/184) of pediatric patients from a hospital in Seoul, Korea. Recent data obtained from US and UK service members with acute watery diarrhea identified EPEC as the 4th most detected pathogen at 12.0% (19/158) of positive samples, after EAEC 46.8% (74/158), ETEC 32.9% (52/158), and norovirus 17.7% (28/158). Lastly, a multicenter study, involving 10 European countries in 2014, found EPEC to be the top detected enteropathogen representing 15% (107/709) total tested and 27.9% (107/384) positive samples. Together, these data indicate that with the utility of multiplex PCR based assays increased detection of EPEC has been observed worldwide and underscore the requirement for greater attention and research into EPEC infections as the causative agent for infectious diarrhea.

One disadvantage of the BioFire GIP assay is that it does not test for the presence of bfpA and therefore, does not distinguish typical from atypical EPEC infection. Specific studies have addressed the question of the prevalence of typical versus atypical EPEC infection. In the UK, England, and Australia, aEPEC was detected in 99% (142/143), 95.4% (104/109), and 95.1% (58/61), respectively, of EPEC-positive samples. In a study of traveler’s diarrhea in UK and US service members, most EPEC samples were aEPEC (83%). Similar reports from developing countries are also found in the literature. In a study of children aged 0-5 years with acute diarrhea in India, aEPEC occurrence (7%) was higher than tEPEC (2.5%) amongst total samples examined with similar reports from other studies from India. In 10 out of 13 studies of developing nations examined by Ochao et. al., aEPEC was more prevalent than tEPEC. Recent data reviewed by Gomes et. al. indicates high prevalence of aEPEC in developing nations (0.05 – 11.3% from 16 studies) and in developed nations (2.5 – 7.5% from 5 studies including 2 from the US). Although significantly more data are available for the aEPEC occurrence in developing
versus developed nations, these data indicate that aEPEC occurrence is emerging over tEPEC worldwide.

aEPEC has also been identified as the causative agent for several outbreaks worldwide in children and adults. A recent NORS study of enteric illness outbreaks from 2009-2019 reported that *E. coli* outbreaks represented 2% of all outbreaks and had the highest hospitalization rate (22.8%) and case fatality rate (0.5%). However, this study did not delineate STEC from EPEC. Upon examination of NORS dashboard data, the majority of cases reported are STEC with only a few EPEC reported. Despite the underreporting of EPEC, 58 EPEC outbreaks were reported during 1971–2018 to NORS; 43 (74%) of these outbreaks were detected during 2016–2018. However, no information is available for whether these outbreaks were from aEPEC or tEPEC or in children or adults. A specific study in the US in 1991 reported an outbreak of >100 adults; the implicated pathogen was aEPEC 039:NM. Outbreak studies from other countries are more revealing. Outbreaks have been reported in China and Japan: aEPEC strain O55:HNM was linked to a daycare infection, 75 students (aged 10-15 years) were infected by aEPEC EC3605, and 112 adults were infected with strain O127a:K63. A study from Finland, reported the 1987 outbreak of EPEC O111:B4 in children and adults and raised the notion that EPEC should be considered in the diagnosis of all diarrheal cases and not only infantile diarrhea. In a more recent report from Finland, an outbreak of EPEC in > 60 children and adults was detected. Interestingly, EPEC infections are still not reportable in Finland and the authors recommended to add EPEC to their national surveillance since PCR-based diagnostics are now routinely performed in clinical laboratories. These studies in Finland underscore the lack of consensus on the importance of EPEC infection as a causative agent of diarrhea which has led to very little if any surveillance worldwide.
The NORS *E. coli* outbreak study indicated that the majority of cases are spread through person-to-person contact. However, in contrast to typical EPEC strains which are rarely found in animals, additional potential sources for aEPEC human infections have been shown to be from wild and domesticated animals and the environment, as reviewed. aEPEC detection in animals with or without diarrhea has been reported for cats and dogs worldwide and at the Mexico-US border, Norway rats in New York, deer, monkeys, horse, and animals used for food production such as dairy cattle in Michigan, slaughter cattle, sheep, goats, pigs, and poultry and chicken-derived products. Through lineage analyses, animal and human aEPEC strains showed close proximity indicating animals are reservoirs and can likely cause zoonotic infections. aEPEC stains have also been isolated from pasteurized milk, meat samples, and vegetables. Moreover, aEPEC was often detected in well water used for drinking in Alabama and had higher prevalence than STEC in Mid-Atlantic untreated surface waters of the US used for irrigation including eight surface water sites, two wastewater reclamation facilities, and one vegetable processing plant. Most reported cases and outbreaks of EPEC have been associated with foodborne sources or person-to-person contact. However, with the increased detection in humans, multiple domesticated and wild animals, and different water sources, dissemination of aEPEC into the environment and the sources for human infection need to be carefully considered and studied.

**Clinical Manifestations of EPEC Infection**

For many decades, studies conducted worldwide have shown that typical EPEC infection was strongly associated with diarrhea in children < 1 year of age with particularly severe disease in infants less than 6 months of age. Based on the prospective, case-control Global Enteric Multicenter Study (GEMS) in infants and young children in developing countries, typical EPEC
was significantly associated with moderate to severe diarrhea in children < 2 years of age in one out of seven sites, and was associated with an increased risk of death in patients aged 0 to 11 months. Similarly, vomiting, watery diarrhea, and severe dehydration were observed to be the common symptoms with EPEC infection in children 0-5 years from a recent study in India. In Peruvian children, EPEC infection resulted in median duration of 5.5 days (range 1-24 days) with maximum number of stools or emesis per day at 6.0 (range 3-12) and 1.2 (range 0-7) times, respectively. In addition, ~8% had persistent diarrhea, 5% had bloody stool, and 34% had fever. EPEC infection was associated with the second highest diarrhea severity score (7.7 ± 3.8) as measured by the modified Vesikari score (mVS), just after rotavirus, and ~32% of the EPEC cases had moderate to severe diarrhea (mVS ≥ 9). In children < 2 years old in Melbourne, Australia, infection with EPEC was associated with vomiting in approximately 50% of patients, was generally not accompanied by fever, abdominal pain, or dehydration, and was not characterized by fecal blood or leukocytes, indicating that it was not inflammatory in nature. Similarly, tEPEC and aEPEC were not associated with fecal stool markers of inflammation such as MPO (myeloperoxidase), NEO (neopterin), or AAT (α-1-antitrypsin) as determined from the MAL-ED multisite birth cohort study based in eight developing nations aimed at estimating pathogen-specific burdens of mild to moderate diarrhea in children aged 0-24 months.

Although EPEC is not associated with inflammatory diarrhea, EPEC infection often results in prolonged or persistent diarrhea. For instance, EPEC resulted in prolonged diarrhea in children lasting on average 12.1 days (range 7.5–16.7 days), greater than the average 4 - 7 days observed for the other enteric pathogens examined (adenovirus, Salmonella, Campylobacter and rotavirus). Similarly, persistent diarrhea lasting more than 14 days in children < 5 years old in Norway and in infants of Brazil and India was associated with EPEC infection. Of
significance, prolonged enteric infections may become chronic, leading to malabsorption, failure to thrive, and malnutrition.\textsuperscript{121}

There is limited data available regarding disease manifestations of EPEC infections in adults. In a Finnish study of a foodborne EPEC outbreak in children and adults, the highest frequency of symptoms were stomach pain (64\%) and diarrhea (50\%) with fewer incidences of headache (33\%), fever and chills (24\%), and vomiting (12\%); only 55\% visited a healthcare facility.\textsuperscript{107} Median duration of symptoms was 48 hours with some having symptoms up to 6.5 days.\textsuperscript{104} However this study did not describe symptomology by age group or distinguish atypical from typical EPEC. A study of EPEC infections from 2010-2012 in England of children and adults, generally described clinical manifestations as acute, self-limited watery diarrhea sometimes accompanied by fever, vomiting, and abdominal pain which was more prevalent in adults.\textsuperscript{122} Interestingly, this study also observed that a high percentage of cases experienced bloody diarrhea (32\%) and had prolonged diarrhea with a median duration of symptoms lasting 11 days (range 1–52 days), compared to those infected with STEC ($n=2$) which lasted on average 7 days.\textsuperscript{122} A recent outbreak at a child care center in Oregon in 2021, identified six children (35\%) and one adult infected with reported diarrhea, vomiting, or other gastrointestinal symptoms.\textsuperscript{104} Although these studies are useful in determining a basic understanding of the clinical manifestations of EPEC infections in adults, more detailed analyses are needed to understand the severity of disease based on the number of bowel movements, duration of diarrhea, and prevalence of other clinical manifestations such as fever, abdominal pain, serum electrolyte imbalances, and others.
Confounding Factors of EPEC Infection

Detection of aEPEC in asymptomatic individuals and in the occasional presence of co-infecting agents confounds our understanding of its virulence. In addition, how host susceptibility might change based on age or in the presence of pre-existing health conditions or other host factors is poorly understood with little evidence in the literature for aEPEC. What little evidence we have is described here.

Asymptomatic Detection

Asymptomatic detection is common among EPEC infection, as well as other established enteric pathogens. Three studies from the US revealed that aEPEC is often detected in asymptomatic children (6.1% prevalence) from 3 counties in Alabama, in asymptomatic children <3 years old from Boston (12.2% prevalence), and aEPEC-detection was not significantly associated with children or adults with diarrhea from Maryland and Connecticut (3.4% prevalence in asymptomatic versus 4.25% in symptomatic). As reviewed by Ochoa et. al., asymptomatic carriage of EPEC in children was detected worldwide with 0.8 – 10.7% prevalence and 5 out of 9 studies found no significant association of EPEC infection with diarrheal subjects. In a recent Finish outbreak of aEPEC in children and adults, asymptomatic individuals accounted for 29% of samples. In the MAL-ED multisite birth cohort study based in eight developing nations, tEPEC was among the top ten pathogens associated with diarrhea, however, three frequently detected pathogens, namely aEPEC, EAEC, and Giardia, and were not significantly associated with diarrhea for any age group, site, or diarrheal syndrome. When one considers the presence of aEPEC in an asymptomatic carrier, however, the sensitivity of the assay used to detect EPEC and the inclusion/exclusion criteria of investigated samples must be carefully considered for each study as these can have profound effects on the interpretation of
detecting EPEC in asymptomatic individuals. For example, in the GEMS and MAL-ED studies, authors from both studies discussed the high prevalence of breast feeding, often carried on past the infant stage, among the population examined. This likely accounts for asymptomatic carriage due to the possible protective effect of secretory IgA and oligosaccharides in breastmilk. In addition, Barletta et al. determined that symptomatic and asymptomatic detection of EPEC was similarly prevalent among Peruvian children, however, EPEC bacterial load in stools was significantly higher in the diarrheal group versus controls. Also, individual aEPEC isolates may retain eae but have various genetic mutations or deletions in other virulence factors causing it to be less virulent and contribute to detection in asymptomatic carriers. My study focused on identifying strategies that would differentiate more virulent from less virulent isolates based on genetic and phenotypic traits.

**Presence of Co-Infecting Pathogens**

BioFire GIP analysis often identifies a high percentage of mixed infections in diarrheal samples compared to routine culture/PCR methods. GIP identified mixed infections in 21.1% of positive prospective samples compared to 8.3% by routine methods; EPEC tended to have the highest rates in mixed infections. Similarly, in a study at George Washington University Hospital co-detection of two or more organisms was common (24.2%), most commonly involving EPEC, EAEC, ETEC, and STEC. EPEC was one of the most common pathogens found in mixed infections from a study of adult hospitalized patients in the UK. In a study of traveler’s diarrhea in UK and US service members, 58% of EPEC infections were detected in the presence of other co-infecting enteric pathogens. In our recent study at LUMC, EPEC was identified in single (49%) and mixed (51%) infections, suggesting it is the cause for diarrhea in at least half of infections. Our study also determined that EPEC co-infections were most
common with *C. difficile* or EAEC, in accordance with published findings. In addition, published data demonstrate clear synergy between different microbial pathogens indicating that EPEC may contribute to the virulence of other pathogens and vice versa. Barletta et. al. determined that among Peruvian children EPEC bacterial load was significantly higher when detected as the sole pathogen versus when present with other co-infecting pathogens. However, it is uncertain if this is true for all ages and for infections worldwide. Ochoa et. al. determined the clinical characteristics and severity of diarrheal episodes caused by single versus multiple pathogens in Peruvian children. EPEC as a sole infection had longer duration of diarrhea but similar number of bowel movements per day and similar diarrhea severity as mixed infection. However, the mixed infection group included all etiologies and a direct comparison of EPEC with or without another co-infecting pathogen was not performed. Overall, although mixed EPEC infections have been reported in the literature, comparisons between sole and mixed EPEC infections in symptomology or severity of disease are scarce, especially in adults, underscoring the need to determine if co-infection status affects clinical manifestations of aEPEC infection.

**Host Susceptibility Due to Age or Pre-Existing Conditions**

tEPEC has long been associated with infection of infants and children, however, recent epidemiological data reveal increased detection in both children and adults of aEPEC. In a multicenter US study, Buss et. al. detected EPEC across all age groups in the US indicating its increased detection levels. Nataro et. al. detected aEPEC in children and adults, however, age stratification did not reveal any age-specific variation in pathogenicity. In contrast, Staples et. al. determined that EPEC-positivity was significantly associated with most age groups, however, patients 0–5 years old had the highest proportion of EPEC detection (99/577, 17.2%) and other age groups ranged from 4.1% to 8.6% with 2 adults groups (41–50 and 61+ years) also showing
a significant association with EPEC detection. Although detection of sporadic and outbreak infections of EPEC have been reported in both children and adults, it remains largely unknown if age contributes to variable disease manifestations of aEPEC infection.

In addition, host susceptibility likely contributes to the detection of aEPEC in asymptomatic hosts. As reviewed by Hu and Torres, the presence or lack of certain host surface receptors, acquired natural immunity, and intact or disrupted natural barriers in the host gut may contribute to presence of aEPEC in asymptomatic individuals. The risk for enteric infection is increased in individuals with intestinal dysbiosis, such as those with dysglycemia, IBD, or immunosuppression. Severity of diarrhea may also be influenced by host susceptibility as exemplified by the study by Dupont et. al. which determined that immunocompromised individuals have longer duration of diarrhea caused by protozoa or bacteria, however, specific details of the influence of EPEC infection were not included. Also, stress increases gut permeability, depletes goblet cells, and enhances bacterial adherence. In fact, a multiple-stressor military training environment was found to increase intestinal permeability, inflammation, and alter intestinal microbiota composition. Therefore, stressed individuals may be more susceptible to enteric infections such as aEPEC. Changes in the gut microbiota due to dietary intake is major factor what could affect aEPEC pathogenicity. For example, artificial sweeteners can cause changes in bacterial diversity with an increase in abundance of Enterobacteriaceae, can increase factors associated with a “leaky” gut, and promotes E. coli_10418 to form biofilms and to adhere to, invade, and kill host epithelial cells in vitro. No studies to date have assessed how dietary intake may affect aEPEC infection.
**Summary of Confounding Factors of EPEC Infection**

A plethora of host factors could influence asymptomatic carriage and variable clinical manifestations of aEPEC infection, such as pre-existing medical conditions, stress, immunosuppression, changes in diet or microbiota, and previous infection, among many others. However, these factors were beyond the scope of this study. Although aEPEC have been detected in the presence of co-infecting enteric pathogens or across different age groups, few studies if any have examined if these factors affect clinical manifestations of aEPEC infection. Our study examined and aimed to describe if differences in symptoms or diarrhea severity of aEPEC infection were attributable to age and co-infectious status. Risk factors for infection such as pre-exiting medical conditions and travel history were also assessed. Although these factors likely influence infection outcomes, data from the different *E. coli* pathotypes and their diverse pathogenic mechanisms indicate that bacterial virulence factors likely make the largest contribution to pathogenicity. aEPEC pathogenic mechanisms are poorly understood, therefore, this study aimed to elucidate virulence factors of aEPEC by examining genetic and *in vitro* phenotypic pathogenic mechanisms known to contribute to tEPEC virulence.

**Typical EPEC Pathogenesis**

In contrast, to other enteric pathogens and *E. coli* pathotypes, tEPEC are non-invasive, lack Shiga toxins and other toxins, and do not elicit a strong inflammatory response. tEPEC pathogenesis is a multistage process involving 3 major stages of infection: (1) initial attachment, (2) intimate adherence, and (3) delivery of effector proteins into host intestinal epithelial cells (IECs) via a type 3 secretion system (T3SS). There are at least 21 known effectors of tEPEC, a greatly protracted number compared to the at least 50 known EHEC effectors. tEPEC effectors are involved in attachment to intestinal epithelial host cells leading to formation of actin
pedestals (dense clusters of F-actin under attached bacteria) and attaching/effacing (A/E) lesions characterized by loss of brush border microvilli and membranous protrusions on host cells intimately attached to bacteria. tEPEC effectors also cause host cell signaling cascades and perturb intestinal epithelial physiology contributing to diarrhea, dehydration and sometimes death.

**Initial attachment** of tEPEC to IECs is facilitated by the EPEC adherence factor plasmid (pEAF) harbored in tEPEC that contains a 14 gene operon encoding bundle forming pili (BFP). Detection of bfpA or probes for the BFP operon have been used throughout the literature to detect tEPEC. BFP promote bacterial autoaggregation, initial adherence to host cells, and microcolony formation rendering the prototypical localized adherence (LA) pattern on host cells (Figure 2). LA pattern of adherence is characterized by tightly clustered bacterial microcolonies in either a 2D or 3D configuration (stacked on top of each other). BFP are also important for bringing tEPEC and host cells into close proximity via BFP retraction, thus increasing delivery of bacterial effector proteins into host cells.

![Image of localized adherence and microcolony formation by tEPEC.](image.png)

**Figure 2: Localized Adherence and Microcolony Formation by tEPEC.** The colonic-like epithelial cell line, SKCO-15, was infected with tEPEC_E2348/69 for 2.5 hrs, fixed, Giemsa stained, and visualized with phase-contrast light microscopy. 2D and 3D (yellow circles) microcolony formation defines the localized adherence (LA) phenotype.
In addition to BFP, tEPEC express many accessory adhesins involved in attachment of tEPEC. Oligomerization of EspC, an enterotoxin, gives rise to rope-like structures that serve as a substratum for adherence and biofilm formation.\textsuperscript{144, 145} Lymphocyte inhibitory factor (LifA), the nearly identical Efa1, and the homologue ToxB contribute to epithelial cell adherence \textit{in vitro}.\textsuperscript{146, 147} LifA/Efa1 are required for intestinal colonization of mice by the related A/E pathogen \textit{Citrobacter rodentium}.	extsuperscript{148} Flagellin is also involved in initial attachment to cultured intestinal epithelial cells but its role diminishes as infection progresses.\textsuperscript{149} Other characterized fimbriae, such as rod-like fimbriae, fibrillae, and \textit{E. coli} common pilus (ECP) are implicated in the interaction of tEPEC with host cells, but their role in pathogenesis is unknown.\textsuperscript{149}

tEPEC also harbor a pathogenicity island called the locus of enterocyte effacement (LEE) encoding the adhesins Intimin (encoded by \textit{eae}) and EspA, components of the T3SS, and many of the effectors responsible for disease-associated phenotypes (Figure 3A). In addition, the LEE encodes transcriptional regulators including: Ler – the central regulator activating all LEE genes as well as more than 50 other genes beyond the LEE, the global regulator of LEE activator (GrlA), and repressor of Grl (GrlR).

\begin{itemize}
\item \textbf{A.} regulators $\rightarrow$ T3SS comp $\rightarrow$ translocon $\rightarrow$ effectors $\rightarrow$ chaperones $\rightarrow$ Intimin $\rightarrow$ Tir
\item \textbf{B.} 
\item \textbf{C.} Intimin
\item \textbf{D.} 
\end{itemize}
**Figure 3: Features of the Locus of Enterocyte Effacement (LEE) Pathogenicity Island.**

(A) Schematic of LEE-encoded genes grouped by function with *n* number of genes or specific genes listed: transcriptional regulators (*n*=3), type 3 secretion system (T3SS) components (*n*=21), translocon proteins (EspA,B,D), effectors (*n*=5), chaperones (*n*=6), intimin (*eaeh* used for detection in GIP analysis), and the effector Tir – translocated intimin receptor. (B) Schematic depicting the chaperones needed for efficient transfer through the T3SS (not all T3SS components are depicted) spanning the inner (IM) and outer (OM) membrane of bacterial wall, the needle (EspA) and other translocator proteins (EspB/D) which contact and insert a pore into host intestinal epithelial cells (IEC) and provides the conduit for effectors to translocate into the host IEC, the adhesin intimin (INT) inserted into OM of bacterial wall with C-terminal exposed to the external environment, and the multiple effectors in host IEC which cause many downstream effects resulting in diarrhea. (C) Depiction of intimin (INT) and Tir interaction leading to intimate attachment between bacteria and host IEC. (D) Important features of N’ and C’ terminal domains of Tir, Y474 is absent in EHEC strains. (Figure created and adapted from 150, 151)

The T3SS is composed of a multiprotein structure embedded in the inner and outer membrane of the bacterial surface and produces a thin extracellular tube or needle which spans from the bacteria to the host IEC (Figure 3B). The needle complex acts as a molecular syringe allowing for the translocation of bacterial effectors directly into the host cytoplasm. The T3SS consist of 21 proteins with the major T3SS components including but not limited to: EscN – the ATPase, EscV – the export apparatus, EscD – major inner ring component, EscRST – channel of the export apparatus, EscC – major outer ring component, and EscF – basal protein of the needle complex (Figure 3B). A gatekeeper complex composed of the proteins SepL, SepD, and chaperone CesL regulates secretion of the translocon proteins including EspA, EspB, and EspD (Figure 3B). Lack of SepL/D results in hypersecretion of effectors. Elongation of EspA subunits from the base of EscF create the channel of the needle structure. At the tip of the needle are the oligomeric structures of EspB and EspD which form a pore in the host cell membrane. The length of the EspA filament is determined by EscP a “molecular ruler” which is secreted until the needle reaches its final length. In a mechanism that is still unknown and upon host cell contact, an activation signal is transmitted through the needle complex, turning off the gatekeeper complex which halts EscP secretion and activates effector secretion. Regulated
secretion of the translocon proteins is facilitated by the chaperones CesAB, CesD, and CesD2 and of different effectors by CesT (Tir) and CesF (EspF).\textsuperscript{151-153}

The interaction of EspA filaments and translocon proteins EspB/D with the host membrane are also part of the initial attachment stage of infection. This interaction leads to injection of the first effector translocated intimin receptor (Tir), \textit{initiating stage 2 of infection} characterized by intimate adherence between bacteria and host.\textsuperscript{141} Intimin, encoded by \textit{eae}, is expressed on the outer surface of the bacteria and interacts with Tir driving intimate attachment to host cells (Figure 3B and 3C). Intimin-Tir interaction initiates clustering of Tir driving downstream signaling events. Src family kinases and PKA phosphorylate Tir at multiple sites in its C-terminal domain (Figure 3D) with phosphorylation at tyrosine residue 474 (Y474) required for rearrangement of filamentous actin.\textsuperscript{154} Tir\textsubscript{–}Tyr474 phosphorylation binds the adapter protein Nck to enable the N-WASP-Arp2/3 complex to polymerize actin beneath attached bacteria.\textsuperscript{154} In addition, Tir\textsubscript{–}Y454 phosphorylation also nucleates actin in a Nck-independent mechanism; although the mechanism is less efficient, it likely works synergistically with the Y474-Nck pathway. Interestingly, EHEC lacks a Y474 equivalent. Instead, it uses the effector TccP, also known as EspFu, and its homologue TccP2 to cause actin polymerization and pedestal formation. Instead of binding Nck, EHEC-Tir binds IRTKS and IRSp53, members of the I-Bar subfamily of membrane deforming and remodeling proteins. TccP couples IRTHS/IRSp53 to N-WASP-Arp2/3 complex to cause actin assembly.\textsuperscript{154}

\textit{Stage 3 of infection} is indicated by translocation of other effectors and downstream events that interfere with several aspects of host cell signaling and physiology as summarized in Figure 4. EPEC effectors are translocated to host cells in a hierarchical order; after Tir, secretion of LEE-encoded effectors is as ordered, EspZ, EspF, EspH, Map, and EspG.\textsuperscript{155} There are also an
additional 14 non-LEE effectors translocated in an hierarchical manner, many of which are involved in dampening the host immune response.\textsuperscript{149} LEE and non-LEE effectors are often multifactorial affecting several processes within the host cell. Interestingly, the first effector to enter after Tir is EspZ and serves as a T3SS gatekeeper protecting host cells from effector overdose cytotoxicity.\textsuperscript{156} Mitochondrial disruption is characteristic of tEPEC infection through concerted action of multiple effectors, including EspF, Map, EspZ and EspH. EspF and Map are targeted to the mitochondria resulting in destruction of the mitochondrial membrane potential, release of cytochrome C and Ca\textsuperscript{2+} into the cytoplasm, cleavage of caspase 9 and 3, eventually leading to apoptosis. In early stages of infection, EspZ protects mitochondria by binding FIS1, a protein that induces mitochondrial fragmentation and mitophagy, promoting infection.\textsuperscript{157} In later stages of infection, EspH causes upregulation of FIS1 and increased apoptosis which may aid in dissemination of bacteria. In addition, non-LEE effectors EspL, NleH, NleF, and NleB suppress mechanism leading to apoptosis.\textsuperscript{158,159} EspF induces ER stress which also leads to apoptosis.\textsuperscript{160}

Figure 4: Major Pathogenic Mechanisms of Typical EPEC.
tEPEC infection leads to reduced expression/function of ion and solute transporters causing diarrhea through various effectors including, NleA (a non-LEE effector), EspF, EspG, EspG2, EspH, and likely others.\textsuperscript{161} Aberrant transporter expression or function include: altered function of NHERF2 (Na\textsuperscript{+}/H\textsuperscript{+}-exchanger regulatory factor 2)\textsuperscript{161}, inhibition NHE3\textsuperscript{161}, Na\textsuperscript{+}K\textsuperscript{+}-ATPase mislocalization\textsuperscript{162}, decreased apical expression and DRA inhibition\textsuperscript{163}, reduced expression of MCT1 (primary SCFA transporter)\textsuperscript{161}, decreased SGLT-1 activity\textsuperscript{164}, SERT inhibition (serotonin transporter)\textsuperscript{161}, aquaporin mislocalization\textsuperscript{149}, and reduced expression and activity of ASBT (apical sodium-dependent bile acid transporter)\textsuperscript{161}.

tEPEC also disrupt intestinal epithelial polarity and tight junction (TJ) structure and barrier function, thereby perturbing water and solute flux across the epithelium contributing to diarrhea.\textsuperscript{165-168} Three main protein complexes control epithelial polarity, Crumbs (Crb3/Pals1/Patj), PAR (Par3/Par6/aPKC\textgreczet/Cdc42), and Scribble (Scrib/Lgl/Dlg). EspF and Map cause mislocalization of the Crumbs polarity complex proteins (Crb3/Pals1/PatJ) away from the apical membrane into the cytoplasm.\textsuperscript{162} Map increases permeability to charged and non-charged molecules\textsuperscript{167}, NleA mislocalizes occludin and ZO-1 from the cell–cell contacts\textsuperscript{169}, and EspG1/G2 induce cytoplasmic accumulation of occludin and delays TJ recovery due to its perturbation of microtubule networks.\textsuperscript{168} EPEC-induced perturbation of Crumbs complex and TJ structure and barrier function are likely downstream effects of disruption of the PAR complex via aPKC\textgreczet mislocalization.\textsuperscript{170} At early infection timepoints, EspF binds the endocytic protein sorting nexin 9 (SNX9) contributing to recruitment of active aPKC\textgreczet to pedestals. Loss of aPKC\textgreczet from PAR, which tightly control TJs, likely leads to endocytosis of JAM-A, occludin, claudin-1, and ZO-1, and disruption of TJ barrier function.\textsuperscript{170} EspF, EspG, and NleA contribute to aberrant vesicle trafficking contributing to TJ and Golgi disruption.\textsuperscript{159, 168, 171}
To add to its many other functions, Map has GTPase GEF activity which induces Cdc42 causing formation of filopodia at the bacterial attachment site.\textsuperscript{83} Conversely, Esp is a RhoGEF inhibitor which reduces formation of filopodia.\textsuperscript{149} Map as well as other virulence factors activates MAPK signaling cascade activating pro-inflammatory responses such as ERK/p38 and NF-κB activation and production IL-8, IL-1β, TNFα, and IFNγ.\textsuperscript{87, 158, 172} To counteract inflammatory responses, EspH, NleC, NleD, NleE, and NleH have all been shown to inhibit NF-κB activation and MAPK through a variety of different mechanisms.\textsuperscript{158, 159, 172}

In conclusion, tEPEC pathogenesis is a delicate balance of processes that disrupt host intestinal epithelial physiology and structure, most of which are encoded on the LEE pathogenicity island. However, EspH and many of the non-LEE effectors counteract the pro-inflammatory, pro-apoptotic, and other processes resulting in non-inflammatory diarrhea.

**Atypical EPEC Pathogenesis**

aEPEC are identified by detection of eae and the absence of BFP. Studies of genetic virulent determinants have revealed that aEPEC represent a large heterogeneous group of organisms often with virulence factors from other \textit{E. coli} pathovars as reviewed.\textsuperscript{9} aEPEC can cause severe diarrhea despite the lack of BFP but can also be found in ~3-20% of asymptomatic individuals as described in the previous section.\textsuperscript{3, 5, 6} The pathogenic mechanisms of aEPEC are assumed to be similar to tEPEC, however, this is largely unsubstantiated and unknown as summarized in Figure 5. In human volunteer studies, ingestion of tEPEC caused diarrhea in 85-90\% of individuals compared to only 12.5-22\% of those infected with tEPEC lacking BFP or cured of EAF.\textsuperscript{173, 174} Interestingly, the response to ingestion of a clinical aEPEC strain was intermediate causing diarrhea in 55\% of volunteers.\textsuperscript{174} These two human volunteer studies indicate that aEPEC have acquired virulence factors to compensate for the lack of BFP.
Therefore, genetic profiles that discriminate between virulent or less-virulent aEPEC strains and characterization of defining pathogenic mechanisms of aEPEC are greatly needed.

<table>
<thead>
<tr>
<th>Stages of Pathogenesis:</th>
<th>Typical EPEC (tEPEC)</th>
<th>Atypical EPEC (aEPEC)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Step 1: Initial attachment</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| BFP | • Localized adherence (LA)  
• High levels of attachment to IEC  
• Autoaggregation & microcolony formation  
• Retraction increases bacteria/host cell proximity & effector delivery | Not present  
• Diverse adherence patterns (LA, AA, DA, UND,NA)  
• Unknown adherence mechanisms  
• Adherence level unknown |
| Other adhesins | Interaction/attachment but not fully elucidated | Present, variable amongst strains |
| EspA | Attach to host cell allowing translocator delivery | Present, sequence divergence |

| **Step 2: Intimate attachment** | |
| Intimin (eae) | Outer membrane protein, binds Tir | Present, many different sub-types  
• Some associated with diarrhea, Unknown |
| Tir | Intimin receptor translocated to host, pedestal formation | Present, sequence divergence |

| **Step 3: Effector delivery** | |
| Via T3SS LEE & non-LEE encoded | Disease-associated phenotypes  
• Pedestal formation  
• A/E lesions (microwilli loss)  
• Decrease SGLT-1 activity  
• Inhibition of DRA  
• Polarity/TJ disruption | LEE present, conserved  
Non-LEE present, highly variable  
• Disease-associated phenotypes  
• LEE, pedestal formation  
• A/E lesions, decrease SGLT-1 activity  
• Inhibition of DRA  
• Polarity/TJ disruption |

**Figure 5: Comparison of Pathogenic Mechanisms between Typical and Atypical EPEC.**

**Virulence Gene Profiling of aEPEC**

Although GIP testing is a rapid and accurate method to detect EPEC, GIP does not screen for BFP so does not classify EPEC as typical or atypical and does not predict the virulence potential of eae-positive isolates. Previously studies have performed virulence genetic profiling of aEPEC isolates including: regulatory and biofilm genes, toxins, different intimin subtypes\(^{175}\), Tir subtypes\(^{176,177}\) and the presence of non-LEE genes, as reviewed.\(^9\) However, to date, most have fallen short in their attempt to find distinguishing factors that dictate virulence. For instance, Staples et. al. determined the frequency of occurrence of putative individual virulence markers among 58 aEPEC strains with ~10-50% having the adhesins \(lpfA\) (3 variants examined), \(efa1/lifA\), or \(paa\), and ~35% having the non-LEE effectors \(nleB\) or \(nleE\).\(^93\) This study, however, only examined isolates from symptomatic individuals, thus their contribution to virulence cannot
be assessed from this study. Similarly, Bugarel et al. investigated 308 tEPEC and aEPEC strains for serotype and the presence and absence of non-LEE effectors; strains clustered into 2 groups according to presence of non-LEE effectors.\textsuperscript{178} Cluster 1 strains, comprised of EHEC/tEPEC/aEPEC, were associated with \textit{nleB}, \textit{nleE}, and \textit{espL2} from EHEC genetic island OI-122 (related to the integrative element 6 (IE6) of tEPEC_E2348/69) and to a lesser extent \textit{nleA}, \textit{nleF}, and \textit{nleH1-2} (OI-71_EHEC and prophage 6 (PP6) of tEPEC). Cluster 2 strains, comprised of tEPEC/aEPEC/LEE-negative STEC and 21 non-pathogenic \textit{E. coli} (\textit{eae}, \textit{bfpA}, \textit{stx1/2}-negative), were negative for all non-LEE effectors examined except for \textit{nleH1-2} which was only present in 23.3\%.\textsuperscript{178} This genetic analysis concluded that cluster 1 strains were more virulent because they had serotypes frequently associated with EHEC and with tEPEC serotypes frequently involved in severe illness and outbreaks in children.\textsuperscript{178} In addition, they concluded that \textit{nleB} may serve as diagnostic tool for identification of virulent EHEC and EPEC serotypes.\textsuperscript{178} However, it was not stated whether the examined strains originated from symptomatic or asymptomatic individuals. If one assumes that all isolates were from symptomatic individuals, diarrheal symptoms or severity were not described.\textsuperscript{178} Thus, presumptions of symptomology and that serotype dictates virulence do not create a strong foundation to build a genetic risk factor model for aEPEC pathogenesis.

In contrast to the study by Bugarel et al., Hazen et al. compared 70 EPEC isolates from the case-control GEMS study of moderate-severe diarrhea in children and compared isolates based on children classified as LI (lethal infection), NSI (non-lethal symptomatic), or AI (asymptomatic).\textsuperscript{179} Transcriptional regulators, \textit{espJ}, and \textit{espS} among many others were more prevalent with LI cases; when comparing symptomatic (LI and NSI) to asymptomatic cases, 39 - 198 different gene clusters were associated to one clinical outcome over another. Prevalent in
symptomatic cases only were many hypothetical proteins, and those in metabolic functions. However, this study assigned typical or atypical EPEC designation on detection of *bfpA* and did not account for other genes on the BFP operon, therefore, the majority of strains were designated as tEPEC and it is hard to discern which genetic features are associated with aEPEC isolates. This study did provide a valuable assessment of phylogenetic lineage of EPEC isolates emphasizing the genetic diversity and detection of 4 new sub-lineages of EPEC based on EcMLST for a total of 10 sub-lineages designated EPEC1-10. I used this classification system in the phylogenetic analysis of our aEPEC isolates. Similarly, an important study from the GEMS cohort, characterized the evolution of 185 aEPEC isolates based on variants of the LEE pathogenicity island and identified 30 distinct LEE subtypes based on *eae, tir, espA, espB, espD, espH*, and *espZ*. They also determined that certain non-LEE genes were associated with particular LEE subtypes. However, the authors did not describe whether the 185 aEPEC isolates originated from symptomatic or asymptotic children. While this study is useful from an evolutionary standpoint, it does not assess if certain LEE subtypes are associated with symptomatic status. In addition, the GEMS study encompassed children from developing nations; thus, it is unclear if aEPEC isolates from adults in developed nations would similarly have diverse EPEC and LEE subtypes. Therefore, I used both the Hazen et. al. and Ingle et. al. EPEC and LEE subtype designations in my assessment of the genomic and LEE diversity of our aEPEC isolates to determine if aEPEC these subtypes are associated with age and symptomatic status.

**Profiling of aEPEC for In Vitro Virulence Phenotypes**

Deletion of *bfpA* from tEPEC results in individually attached bacteria as opposed to localized adherence of microcolonies. In addition, deletion of *bfpA* or *bfpF* reduce delivery of
EPEC effectors into host cells and attenuate pedestal formation and TJ disruption.\textsuperscript{141, 143} Despite attenuated adherence, BFP mutants can still form A/E lesions but require longer infection times to do so.\textsuperscript{141} In addition, only 12.5-22\% of those infected with tEPEC lacking BFP or cured of EAF had diarrhea.\textsuperscript{173, 174} These data indicate that the lack of BFP in a tEPEC background greatly attenuates virulence.

Reports confirm that in aEPEC T3SS components are highly conserved and the entire LEE pathogenicity island is present.\textsuperscript{9, 176} In addition, there is sequence divergence among several genes encoding secreted bacterial effectors, as well as \textit{eae}, \textit{tir}, and the translocators.\textsuperscript{176, 180} However, few studies have focused on the functional analysis of these factors and how they may affect aEPEC virulence. The few studies that have assessed pedestal formation indicate that aEPEC isolates form actin-rich pedestals and A/E lesions, however, this is variable among strains with around 60\% forming pedestal and often require longer incubation to form pedestals.\textsuperscript{9, 181-184} Similarly, in our recent study, we detected pedestal formation in \textasciitilde70\% of our aEPEC isolates by 5 hours post-infection.\textsuperscript{3} These data suggest that in most aEPEC T3SS is functional and the ability to form pedestals is present. However, different adherence mechanisms, especially during initial attachment to compensate for the lack of BFP, likely account for the key differences in aEPEC virulence.

Several studies have examined the role of aEPEC attachment mechanisms in virulence, however, this remains unclear and is confounded by the diverse genetic make-up and phylogenetic distribution of aEPEC strains. A study of children in the UK indicated that adherent \textit{E. coli} had a greater association with diarrhea then non-adherent isolates.\textsuperscript{183} aEPEC display several adherence patterns defined as: localized-adherence-like (LAL) – loose circular clusters of bacteria; aggregative adherence (AA) – large clusters of bacteria in “stack-brick” pattern; diffuse
adherence (DA) – attached bacteria in scattered manner; undefined (UND) – single attached bacteria without a discernable pattern; others are non-adherent. In our recent study, we determined that all aEPEC adherence patterns were represented with diffuse adherence predominating. In contrast, in aEPEC isolates obtained from patients with diarrhea in Brazil, the majority displayed undefined adherence (43.9%) and 14.6% display LAL. Together, these data indicate that adherence patterns are highly variable among strains.

Since aEPEC lack BFP, what adhesins promote initial attachment? The only adhesin originally described in aEPEC, the locus of diffuse adherence (lda), confers DA on HEP-2 cells when expressed in E. coli K12. However, lda is not universal among aEPEC. Many other aEPEC adhesins were originally described in other pathovars. For example, from tEPEC and EHEC: lifA/efa1, toxB, ecpA, and lpfA, fimH (type 1 pilus), csgE (curli), hcpA (hemorrhagic coli pilus), elfA (E. coli laminin-binding fimbriae), and outer membrane proteins, paa (porcine A/E associated adhesin), iha (IrgA homologue adhesin), and agn43 (Antigen 43); from diffusely adherent E. coli and the afimbrial adhesion family: afaBC (Dr family adhesin) and daaC (F1845); from uropathogenic E. coli: papC (P fimbriae) and sfaDE (S fimbriae). Interestingly, a positive correlation between lifA/efa1 or paa and diarrhea is reported. In addition, ECP is expressed by 36% of aEPEC strains that attach to HeLa cells indicating a potential role in virulence. In view of the diversity of adherence factors detected in aEPEC thus far, it is no wonder that they have such diverse adherence patterns. Therefore, in-depth characterization is sorely needed to understand their contribution to aEPEC to virulence, such as their role in attachment, intestinal segment tropism, and the ability to cause in vitro disease-associated phenotypes.
Summary

In summary, increased detection of aEPEC in both symptomatic and asymptomatic adults and children has caused some to speculate about the causal relationship between diarrheal symptoms and aEPEC infection. However, most studies have not addressed confounding infection parameters such as age, co-infections status, preexisting conditions, or other host factor risks, nor have clinical manifestations been fully described for aEPEC infections, especially among adults. In addition, EPEC and especially aEPEC have been determined to be a very heterogenous group of bacteria in all manners, genetically, phenotypically, and clinically. tEPEC pathogenic mechanisms are well defined with a multistage process involving initial and intimate adherence and delivery of effectors into the host intestinal epithelial cells which results in A/E lesions causing destruction of microvilli structure and reducing the absorptive capacity of the intestine. tEPEC effector functions have been extensively described having direct effects on ion transport and actin and microtubule dynamics, disruption of polarity and TJ structure and function, ER stress, and mitochondrial disruption. Non-LEE effectors are largely responsible for counteracting the destructive capacity of the LEE effectors, as described above, ultimately suppressing host immune responses. Collectively, the effectors allow for efficient colonization and infection leading to diarrhea with a relatively low immune response.

aEPEC pathogenic mechanisms are not well understood. I speculate that variable virulent genetic propensity likely accounts for the majority of variable clinical manifestations observed with aEPEC infection. Therefore, the need to discriminate between virulent or less virulent aEPEC strains is needed. Although host factors likely contribute to variable disease manifestations, investigation of most host risks factor were beyond the scope of this study warranting future studies.
This study aimed to address many of the unanswered questions about aEPEC infection. First, we aimed to determine the clinical manifestations of aEPEC infection in a primarily adult population in the US, assess risk and confounding host factors such as age, preexisting conditions, and co-infection status, and compare EPEC loads between symptomatic and asymptomatic individuals among different ages and co-infection status. Second, aEPEC isolates were purified from stools of asymptomatic and symptomatic individuals for whole genome sequencing and bacterial genetic risk factor analyses such as those related to adherence and LEE and non-LEE effectors. Also, in vitro virulence phenotypes of aEPEC isolates were assessed related to adherence and pedestal formation, a proxy for T3SS functionality and A/E lesion formation. Lastly, correlation and principal component analyses were employed to determine the genetic and phenotypic characteristics of aEPEC isolates that were associated with asymptomatic and symptomatic individuals with varying clinical manifestations. I hypothesized that aEPEC possess adhesins or virulence factors that compensate for the lack of BFP to initiate attachment and downstream effects leading to diarrhea. Variable clinical manifestations are likely dictated by the specific virulence factors present. The overarching goal of this project was to define aEPEC pathogenic mechanisms by examining and correlating aEPEC clinical manifestations to virulent genetic and in vitro disease-associated phenotypic factors of aEPEC isolates. Despite increased detection of EPEC and the variable pathogenesis of aEPEC, there is little information to guide clinicians in the treatment of EPEC/aEPEC-associated diarrhea. A first step to improving clinical care is the development of a more specific classification system and improved diagnostic strategies for EPEC based on genetic determinants that correlate with virulence potential of clinical EPEC isolates. This study aimed to be one rung in the ladder in deciphering aEPEC pathogenic mechanisms and variable clinical manifestations.
CHAPTER 3

CLINICAL MANIFESTATIONS OF ATYPICAL EPEC INFECTION IN ADULTS AND CHILDREN FROM A US MIDWEST HOSPITAL


Introduction

EPEC has been the cause of many outbreaks of watery, infantile diarrhea beginning in 1889 and remains a leading cause of diarrhea-related deaths of infants in developing nations.\textsuperscript{83, 190, 191} Although several adult outbreaks have occurred worldwide, including the US, adult EPEC infections in developed nations are severely underreported.\textsuperscript{4, 87} The very limited data available regarding disease manifestations of EPEC infections in adults generally described it as acute, self-limited watery diarrhea sometimes accompanied by abdominal pain, fever, and vomiting.\textsuperscript{174} EPEC is detected in 3-20\% of asymptomatic individuals often with co-infecting pathogens\textsuperscript{2, 3, 117, 120, 122} and different EPEC serogroups do not cause diarrhea in all human adult volunteers.\textsuperscript{181} These observations have caused confusion and controversy about EPEC pathogenicity in adults leading some to disregard GIP-EPEC-positive results. However, in a human volunteer study, an aEPEC isolate caused diarrhea in 55\% of individuals compared to 22\% of those who ingested tEPEC lacking \textit{bfpA}\textsuperscript{192} indicating other virulence mechanisms are present in aEPEC. An in-depth characterization of symptoms and severity of disease caused by aEPEC infection is essential to understand its virulence potential and pathogenic mechanisms. Dysbiosis caused by certain
metabolic and immune conditions can increase the risk of enteric infection.\textsuperscript{39, 40, 99, 193} Thus, we expected that certain factors which cause intestinal dysbiosis may by more prevalent among EPEC-positive SYMP individuals. Similar to documented worldwide outbreaks,\textsuperscript{4, 83} we expected significant diarrheal symptoms in children and adults of sporadic aEPEC infection.

GIP-EPEC-positive cases and matched GIP-EPEC-negative controls were compared to determine if aEPEC can cause significant diarrhea with a particular focus on adults. Next, we aimed to assess symptoms and severity of diarrhea caused by aEPEC infection and to determine prevalent GI symptoms. We questioned whether certain co-morbidities or serum abnormalities are associated with aEPEC infection. In addition, we questioned if age or the presence of co-infecting pathogens affect symptoms or severity of aEPEC infection. We hypothesized that aEPEC causes significant diarrhea in children and adults as sole infections. These studies are essential to establish a baseline for the clinical manifestations associated with aEPEC symptomatic infection to begin to understand its pathogenic mechanisms.

The main objectives of the clinical study, as outlined in Figure 6, were to (1) determine the prevalence of symptoms and risk factors of aEPEC infection by comparing EPEC-negative controls and EPEC-positive cases; (2) to describe the spectrum of symptoms and disease manifestations associated with aEPEC infection in symptomatic individuals; (3) to determine among symptomatic individuals whether age or presence of co-infection contributes to variation in clinical manifestations; and (4) to describe treatments recommended by LUMC physicians.
Figure 6: Cohorts and Main Objectives of Clinical Analysis of EPEC Infection. Diagram of GIP-EPEC–negative and GIP-EPEC–positive cohorts from LUMC with the main objectives in yellow boxes. Demographics of clinical data are shown and include: distributions of females (F), age, and inpatient status in ASYM and SYMP cases and controls. (Figure adapted from: Kralicek SE*, Sitaraman LM*, Kuprys PV, et al. Gastroenterology 2022 Nov;163(5):1321-1333)

Materials and Methods

Case-Control Study

A retrospective case-control study, deemed exempt by the Loyola University Chicago Institutional Review Board, was performed of 380 randomly selected patients from 1229 GIP EPEC-positive cases at Loyola University Medical Center (LUMC) between February 2016 and December 2019. I identified 380 EPEC-negative, age-matched controls from a GIP-negative pool of LUMC individuals within the same timeframe as the EPEC-positive cohort by using a greedy matching algorithm. Paired cases and controls were of the same sex and season of diagnosis.
Collection of Demographic and Clinical Data

Each subject’s electronic medical record was accessed using the Loyola University Chicago Clinical Research Database and was manually reviewed, including inpatient and outpatient notes, clinical flow sheets, laboratory data (serologic and GIP), and medications as needed. International Classification of Diseases (ICD) codes (Table 1) and physician diagnoses were used to determine comorbidities. A hospital policy of performing GIP on all newly admitted patients to rule out *C. difficile* infection allowed a group of hospitalized, asymptomatic individuals to be identified. Patients were considered asymptomatic if they had no documentation of diarrhea or other gastrointestinal (GI) symptoms before or during admission, or if stool was documented as formed.

<table>
<thead>
<tr>
<th>Disease</th>
<th>ICD 9/10 codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBD</td>
<td>556, 556.1, 556.2, 556.3, 556.4, 556.5, 556.6, 556.8, 556.9, 579</td>
</tr>
<tr>
<td></td>
<td>K51, K51.0, K51.00, K51.01, K51.011, K51.012, K51.013, K51.014, K51.018, K51.019</td>
</tr>
<tr>
<td></td>
<td>K51.2, K51.20, K51.21, K51.211, K51.212, K51.213, K51.214, K51.218, K51.219</td>
</tr>
<tr>
<td></td>
<td>K51.3, K51.30, K51.31, K51.311, K51.312, K51.313, K51.314, K51.318, K51.319</td>
</tr>
<tr>
<td></td>
<td>K51.4, K51.40, K51.41, K51.411, K51.412, K51.413, K51.414, K51.418, K51.419</td>
</tr>
<tr>
<td></td>
<td>K51.5, K51.50, K51.51, K51.511, K51.512, K51.513, K51.514, K51.518, K51.519</td>
</tr>
<tr>
<td></td>
<td>K51.8, K51.80, K51.81, K51.811, K51.812, K51.813, K51.814, K51.818, K51.819</td>
</tr>
<tr>
<td></td>
<td>K51.9, K51.90, K51.91, K51.911, K51.912, K51.913, K51.914, K51.918, K51.919</td>
</tr>
<tr>
<td></td>
<td>K50, K50.0, K50.00, K50.01, K50.011, K50.012, K50.013, K50.014, K50.018, K50.019</td>
</tr>
<tr>
<td></td>
<td>K50.1, K50.10, K50.11, K50.111, K50.112, K50.113, K50.114, K50.118, K50.119</td>
</tr>
<tr>
<td></td>
<td>K50.8, K50.80, K50.81, K50.811, K50.812, K50.813, K50.814, K50.818, K50.819</td>
</tr>
<tr>
<td></td>
<td>K50.9, K50.90, K50.91, K50.911, K50.912, K50.913, K50.914, K50.918, K50.919</td>
</tr>
<tr>
<td>IBS</td>
<td>564.1, K58.2, K58.9, K58.8, K58.1, K58.0, 579.8, K90.41</td>
</tr>
<tr>
<td></td>
<td>250, 250.01, 250.02, 250.03, 250.1, 250.11, 250.12, 250.13, 250.2, 250.21, 250.22, 250.23</td>
</tr>
<tr>
<td></td>
<td>250.3, 250.8, 250.81, 250.82, 250.83, 250.9, 250.91, 250.92, 250.93</td>
</tr>
<tr>
<td></td>
<td>E10.6, E10.8, E10.9, E11.0, E11.1, E11.6, E11.8, E11.9, E13.0, E13.1, E13.6, E13.8, E13.9</td>
</tr>
<tr>
<td></td>
<td>250.4, 250.41, 250.42, 250.43, 250.5, 250.51, 250.52, 250.53, 250.6, 250.61, 250.62, 250.63, 250.7, 250.71, 250.72, 250.73</td>
</tr>
<tr>
<td>Liver disease</td>
<td></td>
</tr>
<tr>
<td>70.22, 70.23, 70.32, 70.33, 70.44, 70.54, 70.6, 70.9</td>
<td></td>
</tr>
<tr>
<td>570, 571, 571.1, 571.2, 571.3, 571.4, 571.41, 571.49, 571.5, 571.6, 571.8, 571.9, 573.3, 573.4, 573.8, 573.9, V42.7</td>
<td></td>
</tr>
<tr>
<td>456, 456.1, 456.2, 456.21</td>
<td></td>
</tr>
<tr>
<td>572.2, 572.3, 572.4, 572.8</td>
<td></td>
</tr>
<tr>
<td>I85.0, I86.4</td>
<td></td>
</tr>
<tr>
<td>K70.4, K71.1, K72.1, K72.9, K76.5, K76.6, K76.7</td>
<td></td>
</tr>
<tr>
<td>B18, B18.0, B18.1, B18.2, B18.8, B18.9</td>
<td></td>
</tr>
<tr>
<td>K70.0, K70.1, K70.10, K70.2, K70.9, K71.3, K71.4, K71.5, K71.7, K73, K73.0, K73.1, K73.2, K73.8, K73.9, K74, K74.0, K74.1, K74.2, K74.3, K74.4, K74.5, K74.6, K74.60, K74.69, K76.0, K76.2, K76.3, K76.6, K76.8, K76.9, Z94.4</td>
<td></td>
</tr>
<tr>
<td>Peptic ulcer disease</td>
<td></td>
</tr>
<tr>
<td>531, 531.01, 531.1, 531.11, 531.2, 531.21, 531.3, 531.31, 531.4, 531.41, 531.5, 531.51</td>
<td></td>
</tr>
<tr>
<td>531.6, 531.61, 531.7, 531.71, 531.9, 531.91, 532, 532.01, 532.1, 532.11, 532.2, 532.21, 532.3, 532.31, 532.4, 532.41, 532.5, 532.51, 532.6, 532.61, 532.7, 532.71, 532.9, 532.91, 533, 533.01</td>
<td></td>
</tr>
<tr>
<td>533.1, 533.11, 533.2, 533.21, 533.3, 533.31, 533.4, 533.41, 533.5, 533.51, 533.6, 533.61, 533.7, 533.71, 533.9, 533.91, 534, 534.01, 534.1, 534.11, 534.2, 534.21, 534.3, 534.31, 534.4, 534.41, 534.5, 534.51, 534.6, 534.61, 534.7, 534.71, 534.9, 534.91</td>
<td></td>
</tr>
<tr>
<td>Renal disease</td>
<td></td>
</tr>
<tr>
<td>403.01, 403.11, 403.91, 404.02, 404.03, 404.12, 404.13, 404.92, 404.93</td>
<td></td>
</tr>
<tr>
<td>582, 582.1, 582.2, 582.4, 582.8, 582.81, 582.89, 582.9</td>
<td></td>
</tr>
<tr>
<td>583, 583.1, 583.2, 583.4, 583.6, 583.7</td>
<td></td>
</tr>
<tr>
<td>585, 585.1, 585.2, 585.3, 585.4, 585.5, 585.6, 585.9</td>
<td></td>
</tr>
<tr>
<td>586</td>
<td></td>
</tr>
<tr>
<td>588, 588.1, 588.8, 588.81, 588.89, 588.9</td>
<td></td>
</tr>
<tr>
<td>I12.0, I13.1, N03, N03.3, N03.4, N03.5, N03.6, N03.7</td>
<td></td>
</tr>
<tr>
<td>N05.2, N05.3, N05.4, N05.5, N05.6, N05.7</td>
<td></td>
</tr>
<tr>
<td>N18, N18.1, N18.2, N18.3, N18.4, N18.5, N18.6, N18.9</td>
<td></td>
</tr>
<tr>
<td>N19, N25.0, Z49.0, Z49.01, Z49.02, Z94.0, Z99.2</td>
<td></td>
</tr>
</tbody>
</table>
Diarrhea Severity Score

Diarrhea severity was measured using the modified Vesikari Score (mVS). The mVS incorporates the presence and maximum recorded fever, presence and maximum number of emesis episodes, number of bowel movements (BMs), diarrhea duration, and treatment at a future health care visit for diarrhea (Table 2). Number of days with emesis was replaced with presence or absence of vomiting because duration was often not cited. The mVS was calculated for all symptomatic patients for which all mVS components were available. An mVS of <8 was defined as mild, 9 to 10 as moderate, and >11 as severe. Acute diarrhea was defined as ≤14 days, persistent as 15 to 27 days, and chronic as ≥28 days.

<table>
<thead>
<tr>
<th>Points</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diarrhea duration (days)</td>
<td>0</td>
<td>≤4</td>
<td>5</td>
<td>≥6</td>
</tr>
<tr>
<td>Max no. of diarrheal stools/24 hr period</td>
<td>0</td>
<td>1-3</td>
<td>4-5</td>
<td>≥6</td>
</tr>
<tr>
<td>Presence of Vomiting duration</td>
<td>No</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max no. of vomiting episodes/24 hr period</td>
<td>0</td>
<td>1</td>
<td>2-4</td>
<td>≥5</td>
</tr>
<tr>
<td>Max recorded fever</td>
<td>&lt; 37.0°C</td>
<td>37.1-38.4 °C</td>
<td>38.5-38.9°C</td>
<td>&gt; 39.0°C</td>
</tr>
<tr>
<td>Future healthcare visit</td>
<td>0</td>
<td>-</td>
<td>Primary Care</td>
<td>Emergency Dept.</td>
</tr>
<tr>
<td>Treatment at future healthcare visit</td>
<td>None</td>
<td>IV Rehydration</td>
<td>Hospitalization</td>
<td>-</td>
</tr>
</tbody>
</table>

Diarrheal severity score: 0-8 mild; 9-10 moderate; ≥11 severe illness with max score of 18.

Definition of Outpatient vs. Inpatient

Patients were categorized as an outpatient if the GIP was ordered from an outpatient medical office, urgent care facility, or emergency department when the patient was not admitted, and as an inpatient if they were admitted to the hospital from the emergency department or were already inpatients on wards or in intensive care units.

Longitudinal Data Analysis

We acquired longitudinal data by compiling and analyzing the entirety of GIP results for those subjects who had additional GIP analyses in their electronic medical record.
Statistical Analyses

Statistical analysis and figure creation were performed with GraphPad Prism 10 (GraphPad Software, San Diego, CA) and Illustrator 2021 (Adobe Inc, San Jose, CA), respectively. Statistical analyses used are indicated in figure legends. Data are presented where applicable as mean ± standard error of mean. Statistical significance was defined as $P < .05$, and “ns” indicates non-significant.

Results

Demographics and Risk Factors for EPEC Infection

To determine the clinical manifestations and symptoms associated with EPEC infection, a retrospective, matched case-control, exploratory study was performed of 380 EPEC-negative controls and 380 EPEC-positive individuals at LUMC (Figure 6). Cases and controls were matched by sex, age, and season. Half of the subjects were female. Individuals <18 years comprised 13.4%, 18 to 64 years, 54.5%; and those ≥65 years, 32.1% of the study cohort; and 70% of cases-controls occurred between June and October.

EPEC-negative and EPEC-positive individuals were categorized as asymptomatic (ASYM) or symptomatic (SYMP) for reported GI symptoms, including diarrhea, vomiting, or abdominal pain. The distribution of male and female subjects was comparable between the ASYM and SYMP EPEC-negative (Fisher’s exact test, $P = 0.6046$) and EPEC-positive cohorts (Fisher’s exact test, $P = 0.13$) (Figure 6). Hospitalized individuals comprised most of the EPEC-negative (73.6%) and ASYM(EPEC+) (93.6%) cohorts. In contrast, most of the SYMP(EPEC+) individuals were outpatients most often seen by a primary care physician (Figure 6). EPEC-positive individuals were 4.03 (95% confidence interval, 2.83–5.84)–times more likely to have
any GI symptom than EPEC-negative subjects \( (P < .001) \). Prevalence of GI symptoms significantly decreased with age in both the EPEC-negative and -positive cohorts (Table 3).

**Table 3: Symptomatic Subjects per Age Group**

<table>
<thead>
<tr>
<th></th>
<th>&lt;18 n (%)</th>
<th>18-64 n (%)</th>
<th>≥65 n (%)</th>
<th>Sig.</th>
<th>P-value**</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPEC-</td>
<td>38 (74.5)</td>
<td>126 (60.6)</td>
<td>53 (43.8)</td>
<td>***</td>
<td>0.002</td>
</tr>
<tr>
<td>EPEC+</td>
<td>42 (82.4)</td>
<td>156 (75.4)</td>
<td>73 (59.8)</td>
<td>***</td>
<td>0.0019</td>
</tr>
</tbody>
</table>

**a Chi-square test**

To determine risk factors associated with GIP-EPEC positivity, conditions known to contribute to infection and intestinal dysbiosis were examined.\textsuperscript{197, 198} EPEC-positive individuals were 16.5-times more likely to have contact with an individual with known GI symptoms and 11.8-times more likely to have recently travelled internationally compared with those who were EPEC-negative (Table 4). The most predominant locations of international travel were Mexico (45.7%) and the Caribbean and Central/South America (21.7%) (Figure 7). Inflammatory bowel disease (IBD) and irritable bowel syndrome and other examined comorbidities were equally prevalent among EPEC-negative and EPEC-positive individuals (Table 4). A moderate association of renal disease existed in the EPEC-positive cohort (Table 4).

**Table 4: Risk Factors of EPEC Infection**

<table>
<thead>
<tr>
<th>Condition</th>
<th>EPEC + n (%)</th>
<th>EPEC - n (%)</th>
<th>Odds Ratio</th>
<th>95% C.I.</th>
<th>Sig.</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sick contact</td>
<td>35 (9.21)</td>
<td>4 (1.05)</td>
<td>16.5</td>
<td>4.220 - 141.9</td>
<td>****</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>International travel</td>
<td>48 (12.6)</td>
<td>5 (1.32)</td>
<td>11.75</td>
<td>4.296 - 57.67</td>
<td>****</td>
<td>0.0015</td>
</tr>
<tr>
<td>IBD</td>
<td>18 (4.74)</td>
<td>29 (7.63)</td>
<td>0.607</td>
<td>0.312 - 1.149</td>
<td>ns</td>
<td>0.136</td>
</tr>
<tr>
<td>IBS</td>
<td>14 (3.68)</td>
<td>19 (5.00)</td>
<td>0.722</td>
<td>0.325 - 1.559</td>
<td>ns</td>
<td>0.4725</td>
</tr>
<tr>
<td>Diabetes</td>
<td>92 (24.2)</td>
<td>111 (29.2)</td>
<td>0.743</td>
<td>0.515 - 1.068</td>
<td>ns</td>
<td>0.113</td>
</tr>
<tr>
<td>Liver disease</td>
<td>81 (21.3)</td>
<td>101 (26.6)</td>
<td>0.733</td>
<td>0.508 - 1.052</td>
<td>ns</td>
<td>0.0956</td>
</tr>
<tr>
<td>Peptic ulcer disease</td>
<td>25 (6.58)</td>
<td>35 (9.21)</td>
<td>0.759</td>
<td>0.501 - 1.142</td>
<td>ns</td>
<td>0.198</td>
</tr>
<tr>
<td>Renal disease</td>
<td>92 (24.2)</td>
<td>69 (18.2)</td>
<td>1.535</td>
<td>1.03 - 2.309</td>
<td>*</td>
<td>0.0351</td>
</tr>
</tbody>
</table>
Figure 7: Analysis of EPEC Infection During International Travel. Percentage of symptomatic cases with international travel with location as indicated. Number (n) displayed. (Figure adapted from: Kralicek SE, Sitaraman LM, Kuprys PV, et al. Gastroenterology 2022 Nov;163(5):1321-1333)

Gastrointestinal Symptoms and Severity of Diarrhea Caused by EPEC Infection

EPEC has long been viewed as an infection of children in developing nations. However, most cases in the current study are symptomatic adults in a developed country infected with aEPEC.7 To assess the severity of diarrhea caused by aEPEC, mVS$^{194-196}$ was calculated for each SYMP(EPEC+) case, and comparisons were made between age-groups (Figure 8). Diarrhea was characterized mild in 85.3% (220 of 258), whereas 14.7% (38 of 258) had moderate to severe diarrhea. The average mVS for those aged <18 years (6.33 mVS) and those 18 to 64 years (5.72 mVS) was significantly higher than those aged ≥65 years (4.82) (Figure 8A). Interestingly, moderate-severe cases decreased with age, with the fewest occurring in those aged ≥65 years (Figure 8B).
Multiplie GI symptoms were associated with aEPEC infection. EPEC-positive individuals were 3.8-times more likely to have diarrhea and ~2-times more likely to have vomiting, abdominal pain, or fever than those who were EPEC-negative (Figure 9A). Of those in the EPEC-negative cohort, a cause for diarrhea was determined in 39% (59 of 150), including IBD/irritable bowel syndrome, other GI diseases/complications, or medications, whereas a cause for diarrhea was undefined in 61% (91 of 150) (Figure 9B). In contrast, 86% (225 of 263) of the EPEC-positive group likely had diarrhea due to aEPEC infection, whereas only 14% (38 of 263) had other conditions that cause diarrhea.
Figure 9: Symptoms of EPEC Infection. (A) Percentage of patients with indicated symptom (number indicated) from the EPEC-negative (green) and EPEC-positive (red) cohorts (N = 380 total cases-controls). C.I., confidence interval. (B) Percentage of patients with a cause for diarrhea in EPEC-negative and EPEC-positive cohorts. Percentage of SYMP(EPEC+) patients for each age group with (C) vomiting, (D) abdominal pain, and (E) fever (n = with symptom). IBS, irritable bowel syndrome. Statistical analysis by χ² or χ² trend tests. (F) Number of BMs/24 hours and (G) diarrhea duration by age-group. Statistical analysis by Welch’s analysis of variance. Percentage of patients with continuing diarrhea returning to (H) primary care physician (PCP), emergency department (ED), or ED then hospitalized and (I) according to age. Number returning (inset); (N = total cases). Statistical analysis by χ² trend test. (J) Percentage of patients recommended oral (PO) hydration or hospitalized at follow-up visit for diarrhea. Data are presented as the mean ± standard error of the mean. (Figure source: Kralicek SE*, Sitaraman LM*, Kuprys PV, et al. Gastroenterology 2022 Nov;163(5):1321-1333)

To determine whether age contributes to variation in GI symptoms, SYMP (EPEC+) cases were analyzed further. Diarrhea was present in 97% (263 of 271) of SYMP cases, with no significant differences between age-groups (χ² test, P = 0.4892) (data not shown). Vomiting was present in 34% (93 of 271) of SYMP cases but was more common in individuals aged ≤40 years (Figure 9C). Abdominal pain was present in 53% (145 of 271) of SYMP cases, occurring most
frequently in those aged 18 to 64 years (Figure 9D). Abdominal pain was described as diffuse by most patients (63.85%), followed by lower abdominal pain (17.69%), with pain quality commonly described as cramping (56.15%) or sharp (16.15%) (Figure 10). Fever was present in 18% (49 of 271) of SYMP cases and was most frequent in those aged <18 years (Figure 9E).

![Abdominal pain location and quality](image)

**Figure 10: Abdominal Pain Location and Quality in Individuals Infected with EPEC.** Percentages of symptomatic cases with abdominal pain with the location and quality reported and number (n) displayed. (Figure source: Kralicek SE*, Sitaraman LM*, Kuprys PV, et al. *Gastroenterology* 2022 Nov;163(5):1321-1333)

The average number of BMs in all age groups was 4 to 5/d, however, nearly 25% (62 of 258) had ≥6/d, with a maximum of 12 to 40/d (Figure 9F). Acute diarrhea was present in 83.8% (223 of 266) of individuals, whereas 16% (43 of 266) had persistent or chronic diarrhea (Figure 9G). Average diarrhea duration was 7 days for those aged <18 years, 13 days for those 18 to 40 years, 11 days for those 41 to 64 years, and those ≥65 years had the longest diarrhea duration, lasting an average 18 days (Figure 9G).

Other mVS measures were examined in more detail. Of the SYMP cases, 18% (48 of 271) returned for a future health care visit because of continued diarrhea. Most (73% [35 of 48]) returned to their primary care providers; however, some went to the emergency department or were hospitalized, or both (Figure 9H). Return visits for diarrhea decreased with age (Figure 9I).
Oral hydration was most often recommended at the return visit; however, 23% (11 of 48) were hospitalized for diarrhea (Figure 9J).

**Key Laboratory Values and Vital Signs During EPEC Infection**

Little is known about the effect of EPEC infection on electrolyte levels in humans, especially adults. Serum electrolyte values were compared between SYMP(EPEC+) individuals and either matched EPEC-negative or ASYM(EPEC+) individuals. The average serum sodium level of the SYMP group was significantly lower than the matched EPEC-negative group (Figure 11A). Adult serum sodium levels, and potassium and chloride levels of all ages were not significantly different between SYMP(EPEC+) and ASYM(EPEC+) or the matched EPEC-negative cohort (Figure 11A-C). Next, we questioned whether abnormal electrolytes correlate with diarrhea severity in SYMP(EPEC+) individuals. An inverse correlation was found between low serum potassium and mVS in adults but not in children (Figure 11D). There was no correlation between mVS and serum sodium or chloride levels in adult or pediatric patients (data not shown).

For white blood cell count analysis, patients with extraintestinal infections or disease states associated with leukocytosis were excluded from analysis (Figure 11E). Leukocytosis was equally present between EPEC-negative and SYMP(EPEC+) groups (Figure 11F). However, among the EPEC-positive cohort, leukocytosis was more prevalent in SYMP cases compared with ASYM (Figure 11F). The average white blood cell count of SYMP(EPEC+) pediatric or adult patients, however, was not significantly different from the matched EPEC-negative cohort or ASYM(EPEC+) individuals (Figure 11G). Abnormal serum levels of bicarbonate, blood urea nitrogen, albumin, creatinine, and lactate were present in some SYMP(EPEC+) cases (27%–40%); however, these were not significantly different compared with ASYM(EPEC+) patients.
Similarly, systolic blood pressure, heart rate, and respiratory rate were abnormal in 20% to 40% of individuals, but there was no association with symptomatic status (Figure 13).

Figure 11: Laboratory Abnormalities Associated with EPEC Infection. Pediatric and adult serum levels of (A) sodium, (B) chloride, and (C) potassium for EPEC-positive ASYM and SYMP patients and EPEC-negative controls matched to SYMP(EPEC+) and SYMP(EPEC−) cohort. Grey shading represents normal range. Statistical analysis by Student ($) and paired (&) t tests. (D) Spearman’s correlation between serum potassium and mVS in pediatric and adult SYMP(EPEC+) patients. Linear regression line (solid red) and 95% confidence interval (red dashed lines) are displayed. (E) Percentage of patients with extraintestinal infections. Number with infection (inset) (N = total cases). (F) Percentage of EPEC-positive ASYM and SYMP cases and SYMP(EPEC−) controls with leukocytosis. Number with leukocytosis (inset) (N = total cases). Statistical analysis by Fisher’s exact tests. (G) White blood cell (WBC) count in pediatric and adult patients without extraintestinal infections. Grey shading indicates normal range. Student ($) and paired (&) t tests. Data are presented as the mean ± standard error of the mean unless indicated otherwise. (Figure source: Kralicek SE*, Sitaraman LM*, Kuprys PV, et al. Gastroenterology 2022 Nov;163(5):1321-1333)
Figure 12: Other Serological Analyses of Individuals Infected with EPEC. Serum (A) bicarbonate (Bicarb) (B) BUN, (C) creatinine, (D) albumin, and (E) lactate in asymptomatic (ASYM) and symptomatic (SYMP) pediatric (top panels) and adult (bottom panels). Student’s t-test for significance. Grey shading indicates normal range. (F) Percentage of ASYM and SYMP cases with indicated abnormal serological values. Fisher’s exact tests. (Figure source: Kralicek SE*, Sitaraman LM*, Kuprys PV, et al. Gastroenterology 2022 Nov;163(5):1321-1333)
Figure 13: Vital Signs of Individuals Infected with EPEC. (A) Systolic blood pressure, (B) heart rate, and (C) respiratory rate of individual asymptomatic (ASYM) or symptomatic (SYMP) pediatric (top panels) and adult (bottom panels) cases. Grey shading indicates normal range for adults and children. Yellow shading indicates normal range for infants. (D) Percentage of ASYM and SYMP cases with indicated abnormal vital sign. Fisher’s exact test. (Figure source: Kralicek SE*, Sitaraman LM*, Kuprys PV, et al. Gastroenterology 2022 Nov;163(5):1321-1333)

Contribution of Mixed Infection Status to EPEC Symptoms

We reported previously that EPEC is present in nearly equal proportions as the sole-infecting pathogen and with co-infecting enteric pathogens. We questioned whether there were clinical differences in EPEC sole or mixed infections among age-groups. The prevalence of sole infections tended to increase with age in ASYM(EPEC+) individuals, with the greatest proportion found in those aged ≥65 years (Figure 14A). In contrast, equal proportions of sole and mixed infections were found in all SYMP(EPEC+) patients (Figure 14A).

The composition of mixed infections was examined to determine predominant co-infecting pathogens. For EPEC plus 1 co-infecting pathogen, C. difficile, enteroaggregative E.
coli (EAEC), and norovirus were the top 3 pathogens in ASYM and SYMP cases (Figure 14B and C). However, a greater proportion of *C. difficile* mixed infections was present in ASYM (46.3% [19 of 41]) than in SYMP (21.8% [32 of 147]) cases (Fisher’s exact test, *P* = .003) (Figure 14B and C). Interestingly, EAEC was most prevalent in ASYM pediatric patients with co-infections, whereas SYMP pediatric patients had a nearly equal mix between the top 5 co-infecting pathogens (Figure 14D and E). *C. difficile* co-infection increased with age in both ASYM and SYMP groups (Figure 14D and E).

Figure 14: Contribution of Co-Infection Status to Symptoms of EPEC Infection. (A) Percentage of patients in each age-group with EPEC as the sole pathogen in ASYM and SYMP EPEC-positive patients. Number with sole infections (*inset*) (*N* = sole plus mixed). Statistical analysis by χ² and χ² trend tests; ns, not significant. Percentage of co-infecting pathogens in mixed infections in total (B) ASYM and (C) SYMP patients, and by age for (D) ASYM and (E) SYMP patients. ADENO, adenovirus; ASTRO, astrovirus; CAMP, *Campylobacter*; CRYPTO, *Cryptosporidium*; CYCLO, *Cyclospora*; EIEC, enteroinvasive *E. coli*; ETEC, enterotoxigenic *E. coli*; NORO, norovirus; SALM, *Salmonella*; SAPO, sapovirus; YERSI, *Yersinia*. (G) mVS for sole (blue circles) and mixed (black triangles) infections by age-group. Statistical analysis by multiple Student *t* tests. (H) Diarrhea duration in patients with EPEC as the sole pathogen (EPEC), and EPEC plus 1 (EPEC+1) or ≥2 (EPEC+ ≥2) co-infecting pathogens. Statistical analysis by Welch’s analysis of variance with Šidák’s multiple comparisons. Data are presented as the mean ± standard error of the mean. (Figure source: Kralicek SE*, Sitaraman LM*, Kuprys PV, et al. *Gastroenterology* 2022 Nov;163(5):1321-1333)
We also examined the co-infection status of SYMP individuals who had recently traveled internationally and determined that, in contrast to the larger cohort, mixed (82.6%) predominated over sole EPEC infections, with EAEC and enterotoxigenic *E. coli* being the most prevalent co-infecting pathogens (Figure 15A-C). These data are in contrast to the nontraveler cohort, in which *C. difficile* predominated in mixed infections (Figure 14C).

**Figure 15: Analysis of EPEC Co-Infections During International Travel.** (A) Percentage of symptomatic cases with international travel associated with sole or mixed infections. Number (*n*) displayed. (B) Percentage of mixed infections with 1, 2, 3 or >3 co-infecting pathogens. (C) Number of cases with the indicated co-infecting pathogen for mixed infections with 1, 2, or 3 co-infecting pathogens. (Figure adapted from: Kralicek SE*, Sitaraman LM*, Kuprys PV, et al. *Gastroenterology* 2022 Nov;163(5):1321-1333)

We then determined whether co-infection status affected symptoms or other disease manifestations. Co-infection status did not change the average mVS within or across age-groups (Figure 14F). Further breakdown of GI symptoms considered for the mVS revealed that diarrhea duration was significantly longer when EPEC was the sole pathogen compared with individuals in whom EPEC plus ≥2 pathogens were present (Figure 14G). No significant differences were found between sole and mixed infections for presence of fever, the number or presence of
emesis, the number of BMs, the presence of abdominal pain, or abnormal serum electrolyte levels (data not shown).

**Treatment Measures for EPEC Infection**

Few studies have reported treatment regimens prescribed by physicians for EPEC infection specifically.\textsuperscript{125, 198, 199} We questioned whether and which treatment measures were taken for EPEC-positive SYMP cases. Self-administered oral hydration was most often recommended to SYMP patients regardless of outpatient or inpatient status (Figure 16A). Intravenous (IV) hydration was associated with the presence of fever or vomiting and those admitted for GI symptoms rather than for other reasons (Figure 16B). The mVS does not include treatment at the initial visit; therefore, we questioned whether diarrhea severity, as determined by mVS, correlated with oral vs IV hydration. Indeed, the average mVS was higher in those who received IV rather than oral hydration (Figure 16C), indicating the severity of GI symptoms was likely an important consideration in the hydration treatment plan.

Antibiotic treatment was also assessed. Antibiotics were administered to 57.0\% (149 of 271) of SYMP patients, with 37 patients receiving antibiotics specifically to treat EPEC, 109 were given antibiotics for other reasons, and 3 received antibiotics specifically for EPEC plus other reasons. Reasons cited by providers for EPEC-specific treatment (35 adults and 5 children) were immunosuppression (including organ transplant) or cancer in 10 patients, duration of symptoms in 7, international travel in 7, severity of symptoms in 7, and presence of blood in stool in 3, whereas the reasons were not cited in the remaining 7. Other reasons for antibiotic use (103 adults and 9 children) included but were not limited to diverticulitis, transplant surgery, presence of other co-infecting enteric pathogens such as *C. difficile*, or extraintestinal infections such as urinary tract infection, bacteremia, or pneumonia. Oral hydration was recommended to
82.5% of patients prescribed antibiotics to specifically treat EPEC (Figure 16D). Ciprofloxacin (45.5% [20 of 40]), azithromycin (36.4% [16 of 40]), and levofloxacin (6.82% [3 of 40]) were used to specifically treat EPEC.

Figure 16: Hydration and Antibiotic Treatment for EPEC Infections. (A) Percentage of outpatient and inpatient SYMP(EPEC+) patients who received oral and IV hydration. ns, not significant. (B) Factors associated with oral vs IV hydration. (C) mVS for those who received oral or IV rehydration. Statistical analysis by Student t test. Data are presented as the mean ± standard error of the mean. (D) Percentage receiving antibiotic specifically for EPEC infection (EPEC ABX) or for other reasons (other ABX) grouped according to oral and IV hydration therapy. Statistical analysis by Fisher’s exact tests in A and D. (Figure source: Kralicek SE*, Sitaraman LM*, Kuprys PV, et al. Gastroenterology 2022 Nov;163(5):1321-1333)

Longitudinal Data of EPEC Infection

Many patients in this cohort had self-reported persistent or chronic diarrhea (Figure 4). Therefore, we questioned whether EPEC infection could be detected over extended periods of time. We identified 69 patients with >1 GIP test; of these, 17 had ≥2 EPEC-positive GIP results (ASYM, n = 4; SYMP, n = 13). To assess the longitudinal EPEC infection status with or without other co-infecting enteric pathogens, timelines showing GIP results, GI symptom status, and antibiotic treatment were created for each of these 17 patients (Figure 17A and C). Of note, only 3 patients (#1, 2, and 11 in Figure 12A) were aged <18 years, whereas the remaining patients ranged from 30 to 101 years, with a median age of 58 years (Figure 17A and C). The timelines revealed that EPEC as a sole or part of a mixed infection was present in some individuals for <1 month (#1–7, and 11; Figure 17A), 1 month to up to 1 year (#8–10, 12, 14, and 15; Figure 17A and C), or for 1 year to multiple years (#13, 16, and 17; Figure 17D).
Mixed infections were present in all but 3 patients (#1, 2, 16; Figure 17A and D). Of the 17 patients, 9 patients had 1 co-infecting pathogen, and 6 had multiple co-infecting pathogens (Figure 17B and D). *C. difficile* (n = 11), EAEC (n = 8), and norovirus (n = 5) were the most common single co-infecting pathogens (Figure 17B and D), in accordance with the larger cohort data (Figure 14). Despite the high prevalence of EPEC and non-EPEC mixed infections, patients had the greatest tendency to have GI symptoms when infected with EPEC as the sole pathogen (93.8%) (Figure 17E).

Antibiotic treatment at the time of each GIP test was assessed to determine whether treatment affected subsequent infection status (Figure 17A and C). EPEC-effective antibiotics were prescribed in 21 patients, 11 specifically for EPEC and 10 for other reasons (Figure 17A and C). EPEC persisted after EPEC-specific antibiotic treatment in 38% (8 of 21) of patients (#4, 5, 8, and 12–14; Figure 17A and C). In 62% (13 of 21), EPEC was not present in subsequent GIP (#10 and 12–17; Figure 17A and C) or was assumed to be eradicated because no further GIP analysis was performed (#1–3, 5, and 8; Figure 17A). Interestingly, 2 patients had an EPEC-negative result directly adjacent to EPEC-specific treatment, but subsequent tests identified EPEC (#15 and 16; Figure 17C).
Figure 17: Longitudinal Timelines of EPEC Infection. Longitudinal timelines for (A) <1 year and (C) ≥1 years of GIP results for patients who had ≥2 EPEC+ GIP results with age indicated. GIP result: EPEC sole infection (red dot); EPEC mixed infection (red dot/black center); negative for all pathogens (white dot); GIP-positive for another pathogen, but EPEC-negative (black dot). Asymptomatic (boxed); antibiotic treatment for EPEC (red arrow); other antibiotic treatment (black arrow). Data are presented as mean ± standard error of the mean. (B and D) Tables indicate the number (n) of sole or co-infections and the pathogens present for each patient. Single co-infections delineated by semicolon and ≥2 co-infections indicated by virgule (/). Camp, Campylobacter; Cyclo, Cyclospora; EIEC, enteroinvasive E coli; an, not applicable; Nuro, norovirus. (E) Percentage of patient cases with GI symptoms at the time of GIP; grouped by GIP result. Number positive for GI symptoms (inset); N = total cases. Statistical analysis by $\chi^2$ trend test. (Figure source: Kralicek SE*, Sitaraman LM*, Kuprys PV, et al. Gastroenterology 2022 Nov;163(5):1321-1333)

Summary

aEPEC infections are increasing worldwide, however, there is confusion regarding the clinical significance, especially in adults, due to paucity of data regarding clinical manifestations and strain heterogeneity. In addition, aEPEC is commonly detected in
symptomatic and asymptomatic individuals. aEPEC, in contrast to typical EPEC, lacks bundle-forming pili, altering its pathogenicity. The main objectives of this study were to determine if aEPEC causes disease in adults and to describe the symptoms and clinical manifestations associated with aEPEC infection. We also questioned if age or presence of co-infection contributes to variation in clinical manifestations in symptomatic individuals.

This study defined for the first time the clinical manifestations of sporadic aEPEC infection in United States children and adults. Performing this retrospective case-control study of 380 inpatients/outpatients of all ages I was able to determine that diarrhea, vomiting, abdominal pain, and fever were more prevalent in EPEC-positive cases than in EPEC-negative controls. Risk factors for EPEC infection were international travel, contact with a person with GI symptoms, and to a lesser extent presence of renal disease.

aEPEC is associated with a wide array of symptoms in adults, ranging from asymptomatic carriage to severe diarrhea. aEPEC infection caused mostly acute, mild diarrhea lasting for 6 to 13 days. However, some had severe diarrhea with 10 to 40 bowel movements per day or had persistent/chronic diarrhea. Diarrhea severity decreased with age. Fever, vomiting, and abnormal serum sodium levels were more common in children. Adults more often reported abdominal pain and longer duration of diarrhea. Symptomatic aEPEC infection was associated with leukocytosis in 24% of patients. Co-infecting pathogens did not alter diarrhea severity. Longitudinal data revealed that some are colonized for months to years or are repeatedly infected.

Limitations

A major limitation of this retrospective study is that the matched EPEC-negative control group does not represent the general population. GIP analyses were examined to determine
EPEC negativity, and the primary reason for GIP testing was diarrhea. Therefore, the likelihood that this cohort had diarrhea, other GI symptoms, or laboratory abnormalities is higher than the general population and may underestimate the association of symptoms with aEPEC infection.

Another limitation is the lack of consensus on how to determine diarrhea severity.\textsuperscript{200} The mVS was carefully chosen because of its validation and use in prospective and retrospective analyses of acute gastroenteritis in children in emergency department and outpatient settings.\textsuperscript{195, 196, 202} In addition, the measures used in mVS are the most commonly reported by physicians and are available in a retrospective analysis, unlike the adult scoring system described by Xiao et al.\textsuperscript{203} Indeed only 5\% (13 of 271) of our symptomatic patients were excluded because the mVS could not be calculated due to absence of one or more measure. Despite the apparent benefits of using the mVS system, it assigns equal weight to diarrhea lasting ≥6 days.\textsuperscript{195, 196} Therefore, the mVS may have underestimated diarrhea severity in adults in our study because adults tended to have longer duration of diarrhea.
CHAPTER 4
BACTERIAL LOAD OF EPEC-POSTIVE STOOLS IN ASYMPOTOMATIC AND
SYMPTOMATIC INDIVIDUALS

(Adapted from: Kralicek SE*, Sitaraman LM*, Kuprys PV, et al. Clinical manifestations and
stool load of atypical enteropathogenic Escherichia coli infections in United States children and

Introduction

Higher bacterial load has been associated with symptomatic status and diarrhea severity
during infection with aEPEC, EAEC, ETEC, Shigella, Campylobacter, and others.\textsuperscript{125,204-209} In
addition, greater Campylobacter load correlates with increased detection of stool calprotectin\textsuperscript{207}
indicating induced intestinal inflammation. A dose-dependent diarrheal response was observed in
tEPEC human volunteer studies.\textsuperscript{173} These data indicate that it is not just the presence of a
pathogen, but rather the quantity present which is predictive of disease and in some cases
severity. The few studies which examined aEPEC load focused on children in low-income
countries and immunosuppressed adults in the US.\textsuperscript{125,199} Barletta et. al. demonstrated a 10-fold
increase in median EPEC load between asymptomatic and symptomatic children and a 10-fold
increase in symptomatic children with EPEC as the sole pathogen compared to those co-
infected.\textsuperscript{125} Therefore, it is imperative to understand if EPEC load correlates with infection
symptoms or severity in adults and in individuals from a developed nation. I hypothesized that
EPEC load would be higher in symptomatic versus asymptomatic individuals regardless of age.
Based a previous study\textsuperscript{3}, I expected that EPEC load will be higher when EPEC is present with
another co-infecting pathogen (mixed) rather than sole infections. The objectives of this aim
were to determine if increased load is associated with symptoms or severity across different age groups or if co-infecting pathogens influence EPEC loads by comparing loads from 3 groups: healthy-asymptomatic (OB), hospitalized-asymptomatic (ASYM), and symptomatic (SYMP) individuals (Figure 18).

![Diagram of the Source and Number of Stools Used in EPEC Load Analysis.](image)

**Figure 18: Diagram of the Source and Number of Stools Used in EPEC Load Analysis.** (Figure adapted from: Kralicek SE*, Sitaraman LM*, Kuprys PV, et al. *Gastroenterology* 2022 Nov;163(5):1321-1333)

**Materials and Methods**

**Source of Stools**

GIP-EPEC-positive stools were obtained from the clinical microbiology laboratory at Loyola University Medical Center (LUMC) under Loyola University Chicago (LUC) IRB exempt status for retrospective analysis and correspond to stools from the hospitalized-ASYM and SYMP cohorts described in CHAPTER 3 (Figure 6). Hospitalized-ASYM and SYMP stools from LUMC were stored at 4°C in Cary-Blair transport media and obtained within 3–5 days of GIP analysis. Upon acquisition, stools were aliquoted into 1 mL samples, frozen, and stored at -80°C. Stools from healthy, asymptomatic individuals were acquired from OpenBiome (OB), a stool bank that collects stools for fecal microbiota transplants. OpenBiome donors undergo a rigorous screening process to ensure they were devoid of significant medical conditions, GI
symptoms, travel risks, and other health measures. Donor ages ranged from 18–49y; 33% were female. The FDA alerts which reported infections with EPEC and STEC following FMT prompted the stool bank OpenBiome to screen all available donor material using the Biofire GIP assay allowing for EPEC detection in these otherwise healthy stools. OB donor stools were received as either raw stool or glycerol-preserved stool (0.1g of stool resuspended in a glycerol-saline solution) and stored at -80°C.

**Relative EPEC Load Determination**

Relative EPEC loads (EPEC/total bacteria) were determined from stools of symptomatic and asymptomatic individuals at LUMC using methods previously described, except that genomic DNA was extracted from stool samples using the MagMAX Microbiome Ultra Nucleic Acid Isolation kit (Applied Biosystems, Foster City, CA), and quantitative PCR was performed using TBgreen Premix EXTaqII reagent (Takara Bio USA, San Jose, CA). Briefly, 100 ng of stool DNA in a total reaction volume of 20 μL with eaeA (0.5 μM) and 16S-rDNA universal (0.5 μM) primers. PCR for each sample was run in triplicate alongside triplicate standard curves and repeated 2-5 times. Eppendorf RealPlex software was employed for generation and analysis of standard curves to quantify picograms of EPEC (eaeA primers) and total bacterial (universal primers). To control for variability in stools, the universal primers were used as a normalization signal to detect total bacteria. EPEC (pg) divided by total bacteria (pg) equals relative EPEC load allowing for comparison between samples from ASYM-healthy(OB), ASYM-hospitalized (LUMC), and SYMP (LUMC) individuals.

**Statistical Analyses**

Statistical analysis and figure creation were performed with GraphPad Prism 10 (GraphPad Software, San Diego, CA) and Illustrator 2021 (Adobe Inc, San Jose, CA),
respectively. Statistical analyses used are indicated in figure legends. Data are presented where applicable as mean ± standard error of mean. Statistical significance was defined as $P < .05$, and “ns” indicates non-significant.

**Results**

Previous studies report that higher EPEC load in stool samples is associated with diarrhea.\textsuperscript{125,199} Therefore, I questioned if EPEC load correlates with symptoms or diarrhea severity in our cohort using stool samples from SYMP ($n=109$) and ASYM individuals ($n=44$), and from healthy, asymptomatic (healthy-ASYM) OpenBiome donors ($n=53$) (Figure 19A). EPEC represented $<0.01\%$ of total bacteria in most samples regardless of symptomatic status (Figure 19A). However, most samples with EPEC representing $<0.01\%$ of total bacteria were from healthy-ASYM(OB) individuals while samples from hospitalized ASYM patients had intermediate loads ($0.01–0.09\%$), and samples from SYMP individuals had the greatest percentage of loads with EPEC representing $\geq 0.1\%$ of total bacteria with many greater than 1\% (Figure 19A). Indeed, the average EPEC load was significantly higher in SYMP individuals compared to the healthy-ASYM donors (Figure 19B). Higher EPEC loads, however, did not correlate with severity or individual symptoms (data not shown). Co-infection status did not change the EPEC load in SYMP or ASYM individuals (Figure 19C) nor was there a significant difference in loads based on age (Figure 19D).
Summary

As with many different pathogens, I determined that EPEC load is higher in SYMP individuals compared to healthy-ASYM OB carriers. However, EPEC load did not correlate with individual symptoms, severity of diarrhea, or with age. Contrary to our previous finding, co-infection status did not affect EPEC load likely due to increased detection through optimization of the stool bacterial DNA purification and qPCR protocols. In addition, the increase in \( n \) in the
current EPEC load analysis also likely contributed to the more accurate representation of the effects of co-infections.

Although the average EPEC load was greatest in SYMP individuals, EPEC load is not as predictive of symptoms. Indeed, I found that some hospitalized and healthy-ASYM individuals had EPEC loads between 0.1% and 1.5% and that ~60% of symptomatic infections had EPEC loads of <0.1%, indicating that EPEC load may not directly correlate with clinical symptoms. These data indicate that there likely are additional bacterial or host factors that contribute to disease manifestation besides EPEC stool load.

Limitations

The major limitation in studying EPEC load was stool variability due to collection from asymptomatic (formed) and symptomatic (variable consistency) individuals and collection in different storage methods (transport media, saline-glycerol or raw). To account for variable stools, universal primers which detect total bacteria within the stool were employed as well as using the same amount of DNA input. In addition, the MagMAX Microbiome Ultra Nucleic Acid Isolation kit (Applied Biosystems) is optimized to remove stool and transport media impurities allowing for greater qPCR efficiency and lower detection levels.

Another limitation of this study was that I could not examine how EPEC load might change over the course of an infection as this was a retrospective analysis and data was not collected to correlate sample collection date to the day of illness. Barletta et. al. determined there was a tendency for higher bacterial loads during the first days of illness of Peruvian children. It would be interesting to understand how EPEC load changes over acute, persistent, and chronic infection in children and adults.
CHAPTER 5:
GENETIC VIRULENCE FACTORS ASSOCIATED WITH SYMPTOMATIC STATUS IN
ATYPICAL EPEC ISOLATES

Introduction

The pathogenic mechanisms of aEPEC are assumed to be similar to tEPEC, however, this is largely unsubstantiated warranting further investigation. tEPEC is non-invasive, lacks Shiga toxins. Pathogenesis is a multistage process involving attachment to host intestinal epithelial cells, formation of actin pedestals and attaching/effacing (A/E) lesions\textsuperscript{154}, and delivery of effector proteins via a T3SS (Figure 4 and 5). As expected with a multistage process of infection, each stage involves many different virulence proteins which are encoded on different genetic elements within tEPEC. aEPEC represent a large heterogeneous group of organisms often with virulence factors from other \textit{E. coli} pathovars and likely variable presence of known virulent genetic factors.\textsuperscript{9} Therefore, my first questions in understanding aEPEC pathogenesis were, does aEPEC have similar known genetic virulence factors and what is their homology to the reference strains tEPEC\_E2348/69 and EHEC\_Sakai? I hypothesized that aEPEC isolates harboring more genetic virulence factors, especially those related to adherence, and those with higher homology to reference EPEC/EHEC strains will have a greater tendency to be isolated from symptomatic individuals. The overall goal was to establish a baseline for distinguishing more virulent from less virulent strains of aEPEC.

The first stage of tEPEC pathogenesis depends on initial attachment via BFP encoded on EAF plasmids.\textsuperscript{87} In addition, other initial adhesion factors in tEPEC/EHEC (Table 5) encoded
throughout the genome contribute to host attachment but are not essential indicating a cooperative effort in providing optimal host cell attachment. Due to the lack of BFP in aEPEC, I speculated that aEPEC possess these or other adhesins that compensate and initiate attachment and downstream effects leading to diarrhea. Other *E. coli* pathovar specific adhesins (Table 5) have been found in aEPEC isolates and may contribute to cooperative host cell attachment. Overall, these adhesins are likely variable among aEPEC strains warranting full characterization of the adhesins present to begin to understand the initial attachment required for aEPEC pathogenesis.

**Table 5: Characterized Adhesin Genes**

<table>
<thead>
<tr>
<th>LEE associated:</th>
<th>eae, tir, espA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Others from:</td>
<td></td>
</tr>
<tr>
<td>aEPEC - lda</td>
<td></td>
</tr>
<tr>
<td>tEPEC/EHEC - toxB, lifA/efa1, espC, ecpA, lpfA, fimH, hcpA, elfA, paa, iha, agn43</td>
<td></td>
</tr>
<tr>
<td>DAEC - afaBC, daaC</td>
<td></td>
</tr>
<tr>
<td>UPEC - papC, sfaDE</td>
<td></td>
</tr>
</tbody>
</table>

Early attachment also depends on LEE pathogenicity island (PAI) encoded genes including the T3SS, EspA the needle protein, and the translocators EspB/D. Other LEE encoded proteins are intimin and Tir required for intimate attachment and pedestal formation, and effector proteins with downstream effects leading to changes in host cell physiology causing diarrhea. Reports indicate that T3SS components and most LEE PAI genes are highly conserved in aEPEC. However, it is unknown if sequence divergence among these proteins contributes to variable disease manifestations. tEPEC and EHEC have between 34-55 non-LEE effectors encoded on different integral elements and prophages throughout their genomes. These non-LEE effectors have variable functions – some disrupt host cell physiology or promote pedestal formation while some dampen the immune response, but in combination they promote
infection. Some aEPEC have been characterized for presence/absence of certain non-LEE effectors\textsuperscript{176,194} but their role in clinical manifestations has not been explored in detail.

The \textit{objectives of this aim} were to perform whole genome shotgun (WGS) sequencing and analysis on aEPEC isolates purified from stools of 3 different cohorts: (1) symptomatic (SYMP) individuals from LUMC, (2) LUMC hospitalized patients without diarrhea (ASYM), and (3) healthy, asymptomatic (OB) individuals obtained from the stool bank OpenBiome. Lineage determinants such as phylogroup, serogroup, and MLST designations were used to help define the evolution of aEPEC isolates and provide a historical context for our nascent isolates from children and adults in the US. I aimed to characterize genetic virulence factors of aEPEC isolates with a particular focus on adhesins listed in Table 5 and genes relevant to EPEC virulence including genes harbored in LEE and non-LEE PAIs, and EAF plasmids of reference strains tEPEC\_E2348/69 and EHEC\_Sakai to determine if virulence gene presence/absence or homology correlate with symptomatic status. I also used a pan-genome, un-biased approach to explore other virulence factors which may be more prevalent among SYMP isolates.

\textbf{Materials and Methods}

\textbf{EPEC Isolate Purification}

In order to identify \textit{eaeA}-positive \textit{E. coli}, flat, dry, pink and non-mucoid single colonies\textsuperscript{70} from SYMP, ASYM, or OB stools were patched onto MacConkey II (BD \#212306) agar plates and subjected to colony PCR using and \textit{eaeA}-specific primers as described\textsuperscript{3} Primary \textit{eaeA}-positive patches were re-streaked onto fresh MacConkey II agar and rescreened for \textit{eaeA} to ensure purity of isolated colonies, then grown overnight in Luria-Bertani (LB) broth (BD, 244620) to prepare glycerol stocks and stored at −80°C.
Whole Genome DNA Purification, Sequencing, Assembly, and Annotation

EPEC isolate genomic DNA was extracted from overnight LB cultures using DNeasy Blood and Tissue kit (69504, Qiagen, Germantown, MD) then submitted to the Loyola Genomic Facility for library preparation and whole genome shotgun sequencing as previously described. Briefly, Illumina Nextera Flex Library Preparation Kit was used for sequencing on Illumina MiSeq platform yielding 250bp paired-end reads with approximately 50-100x coverage. Paired-end reads were assembled and annotated using the BV-BRC Comprehensive genome analysis that employs the assembly method Unicycler v0.4.8 and RastK for annotation. Assembled contigs were then further submitted to the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (https://www.ncbi.nlm.nih.gov/genome/annotation_prok/). Paired-end reads were also submitted to the Enterobase pipeline which uses SPAdes for assembly and Prokka for annotation. Assemblies and annotations were examined and compared to enhance scientific rigor. Assembly with the BV-BRC Unicycler with PGAP annotation yielded the most complete assembly with fewest high quality contigs and most complete annotations. Complete annotations were assessed by manual curating known virulence genes of interest and assessing which annotation pipeline most successfully annotated the genes without giving hypothetical protein designations.

Final Isolate and Lineage Determination

Paired sequences from 130 eae-positive colonies were uploaded and assembled on Enterobase pipeline\textsuperscript{219,220} for \textit{in silico} phylogroup determination (there are 7 known \textit{E. coli} phylogenotypes – A, B1,B2,C,D,E,F\textsuperscript{73,221}, serotyping (H-antigen and O-antigen)\textsuperscript{71}, and Achtman multi-locus sequence typing (MLST)\textsuperscript{74}. If predictions were inconclusive from Enterobase, WGS reads were uploaded directly to SerotypeFinder 2.0\textsuperscript{71}, MLST 2.0.9 \textsuperscript{74}, or EzClermont \textsuperscript{72} to access
the latest web application updates. All strains were curated, analyzed for species determination and lineage determinants.

To determine if OB isolates originating from the same donor were duplicate strains, OB isolates underwent further *in silico* analysis. Percent coverage and small nucleotide polymorphisms (SNPs) were examined using the Variation Analysis tool available on the BV-BRC pipeline with parameters set for BWA-mem aligner and FreeBayes SNP caller. Percent coverage was calculated by examining the fraction of DNA bases aligned between each pair of genomes using the Variation analysis tool as follows and according to Fath et. al.:

\[
\frac{(x_{aln} + y_{aln})}{(x + y)} \times 100 = \% \text{ coverage};
\]

\(x\) and \(y\) = the lengths of genome X and Y in bases, respectively;

\(x_{aln}\) and \(y_{aln}\) = the number of aligned bases of genomes X and Y, respectively;

\(x_{aln} or y_{aln} = x or y - b_{x or y} ; b_{x or y} = \text{bases with} \leq 10x \text{ coverage.}\)

**Pangenome Analysis**

To explore the pangenome, PGAP annotations of 36 OB isolates were uploaded to PanExplorer processed using PanACoTa software at a threshold of 80% minimum identity; average nucleotide identity (ANI) processed using FastANI and the presence/absence matrix were analyzed for similarity between isolates. Core genes are defined as those present in 100% of genomes, accessory genes as those in at least 2 and up to all but 1 genome, and unique genes as those found in only 1 genome. The number of unique and accessory genes was compared between OB isolate genomes originating from the same donor and belonging to the same phylogroup, serotype, and MLST. The number of unique proteins between 2 isolates was calculated using:

\[
\frac{(n + u')}{(n + u' + a)} \times 100 = U;
\]
\[ n = \text{number of different accessory proteins between 2 compared isolates} \]
\[ u' = \text{number of unique proteins within 2 isolates determined in pangenome analysis} \]
\[ a = \text{number of accessory proteins similar between compared isolates} \]
\[ U = \text{percent unique proteins between compared isolates} \]

Pangenome analysis was performed on 112 aEPEC isolates as described for the 36 OB isolates. The number of unique and accessory genes was compared by symptomatic status. Chi-square tests were performed for each accessory gene comparing the number of isolates in each group, OB, ASYM, and SYMP, that contained that gene. To create a volcano plot, the (-log) transformation of the \( p \)-value from individual Chi-square tests was plotted on the \( y \)-axis against the percent change of the accessory gene present (percentage of SYMP isolates minus the percentage of OB isolates with that accessory gene) on the \( x \)-axis. Those with a negative percent change have a greater association with OB isolates and those with a positive percent change have a greater association with SYMP isolates.

**Core Genome and Individual Protein Phylogenetic Analyses**

Assembled genomes were submitted to Bacterial Gene Tree analysis on BV-BRC with parameters of 1000 core genes with max allowed deletions/substitutions 10/10. Amino acid sequences of proteins of interest were curated in BV-BRC protein feature groups then submitted to Gene/Protein Tree analysis with the following parameters: Alignment – trim ends of alignment threshold “0” and remove gappy sequences threshold “0”; Tree – RAxML model LG. Resulting core and individual gene trees were downloaded as a newick file and visualized and edited in FigTree v1.4.4 (https://github.com/rambaut/figtree/releases).
**In Silico Proteomic Analysis**

Protein homology comparisons of all aEPEC isolates to tEPEC strain E2348/69 and EHEC strain Sakai were performed using the Proteome comparison tool available on the BV-BRC pipeline\(^{218}\), manually curated, and heatmaps generated in GraphPad Prism 10 (GraphPad Software). Only those proteins with greater to or equal to 80% coverage were included in the *in silico* proteomic analysis of LEE, non-LEE, and adhesin virulence proteins. Cut-off values of 80% and 70% identity were used for LEE and non-LEE virulence proteins. Adhesins, autotransporter adhesins, and fimbrial protein annotations were searched for using Geneious Prime v2023.2, a bioinformatics software for sequence data analysis, and presence and absence curated and analyzed. The presence of genes was corroborated using BV-BRC Feature search on grouped genomes.

**Statistical Analyses**

Statistical analysis and figure creation were performed with GraphPad Prism 10 (GraphPad Software, San Diego, CA) and Illustrator 2021 (Adobe Inc, San Jose, CA), respectively. Statistical analyses used are indicated in figure legends. Data are presented where applicable as mean ± standard error of mean. Statistical significance was defined as \(P < .05\), and “ns” indicates non-significant.
Table 6: Strain and Clonal Determinant Analysis

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Donor</th>
<th>Day</th>
<th>Phylogroup</th>
<th>MLST&lt;sup&gt;a&lt;/sup&gt;</th>
<th>O-antigen</th>
<th>H-antigen</th>
<th>Cov. (%)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>name</th>
<th>SNP&lt;sup&gt;c&lt;/sup&gt; (A→B)</th>
<th>SNP&lt;sup&gt;c&lt;/sup&gt; (B→A)</th>
<th>SNP notes</th>
<th>ANI&lt;sup&gt;d&lt;/sup&gt; (%)</th>
<th>Acc.&lt;sup&gt;e&lt;/sup&gt; (%)</th>
<th>Acc.&lt;sup&gt;f&lt;/sup&gt; (%)</th>
<th>Clonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>OB84</td>
<td>A</td>
<td>211</td>
<td>A</td>
<td>10</td>
<td>O157</td>
<td>H16</td>
<td>98.09</td>
<td>A</td>
<td>123</td>
<td>184</td>
<td>Ag 43</td>
<td>99.986</td>
<td>60</td>
<td>8.05</td>
<td>no</td>
</tr>
<tr>
<td>OB86</td>
<td>A</td>
<td>216</td>
<td>A</td>
<td>10</td>
<td>O157</td>
<td>H16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB38</td>
<td>A</td>
<td>212</td>
<td>B1</td>
<td>381</td>
<td>O119</td>
<td>H9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB15</td>
<td>A</td>
<td>212</td>
<td>B1</td>
<td>381</td>
<td>O119</td>
<td>H9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB33</td>
<td>A</td>
<td>216</td>
<td>B1</td>
<td>381</td>
<td>O119</td>
<td>H9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB128</td>
<td>K</td>
<td>41</td>
<td>B2</td>
<td>2678</td>
<td>O137</td>
<td>H6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB129</td>
<td>K</td>
<td>41</td>
<td>B2</td>
<td>2678</td>
<td>O137</td>
<td>H6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB4</td>
<td>M</td>
<td>70</td>
<td>E</td>
<td>2792</td>
<td>O160</td>
<td>H23</td>
<td>99.99</td>
<td>A</td>
<td>124</td>
<td>95</td>
<td></td>
<td>99.997</td>
<td>149</td>
<td>11.64</td>
<td>no</td>
</tr>
<tr>
<td>OB64</td>
<td>M</td>
<td>70</td>
<td>E</td>
<td>2792</td>
<td>O160</td>
<td>H23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB145</td>
<td>P</td>
<td>7</td>
<td>B1</td>
<td>517</td>
<td>O131</td>
<td>H19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB146</td>
<td>P</td>
<td>16</td>
<td>B1</td>
<td>517</td>
<td>O131</td>
<td>H19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB147</td>
<td>P</td>
<td>16</td>
<td>B1</td>
<td>517</td>
<td>O131</td>
<td>H19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB8</td>
<td>P</td>
<td>22</td>
<td>B1</td>
<td>517</td>
<td>O131</td>
<td>H19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB40</td>
<td>P</td>
<td>16</td>
<td>B1</td>
<td>517</td>
<td>O131</td>
<td>H19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB148</td>
<td>P</td>
<td>22</td>
<td>B1</td>
<td>517</td>
<td>O131</td>
<td>H19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB149</td>
<td>P</td>
<td>22</td>
<td>B1</td>
<td>517</td>
<td>O131</td>
<td>H19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Achtmann 1 multi-locus sequencing type

<sup>b</sup> Fraction of DNA bases aligned between compared genomes, see Materials and Methods for detailed equation.

<sup>c</sup> Small nucleotide polymorphisms; high quality variants determined by BV-BRC Variation Analysis.

<sup>d</sup> Average nucleotide identity determined through FastANI on Panexplorer pipeline.

<sup>e</sup> Number of accessory proteins unique between 2 isolates

<sup>f</sup> Percentage of unique accessory proteins, see Materials and Methods for detailed equation.
Results

Determination of Final Isolates

Single presumed *E. coli* colonies from SYMP, ASYM, and OB stools plated on MacConkey II agar were screened via colony PCR for *eaeA*, the gene encoding intimin and that which is used to detect EPEC in the GIP assay. Previously, 60 *eae*-positive isolates from stools acquired from LUMC were purified by our laboratory and characterized for *in vitro* virulence phenotypes. An additional 70 *eae*-positive isolates were purified from stools obtained from LUMC and from OpenBiome. In total, 130 *eae*-positive isolates were subjected to WGS, assembly, and further analysis. MacConkey II agar allows for growth of gram-negative bacteria yielding different morphologies for different bacterial species. Therefore, to ensure that isolates were *E. coli*, paired-end sequence reads were uploaded to the Enterobase pipeline for automatic assembly, annotation, and lineage assessment revealing that the majority of *eae*-positive isolates were EPEC (n=118). The remaining isolates were *E. albertii* (n=4), *E. fergusonii* (n=1), and non-EPEC *E. coli* (n=4) and were therefore removed from further analysis (Figure 20A). To determine whether the 118 EPEC *eae*-positive isolates were atypical or typical EPEC, BFP presence was assessed using the EAF plasmid, pMAR2, from **e**TEPEC_E2348/69 as a reference and the in silico Proteomic analysis tool available on BV-BRC pipeline. BfpA was detected in 13 *eae*-positive isolates, SYMP (n=10), ASYM (n=1), and OB (n=2) (Figure 21). However, only one isolate, CE102, contained all 14 genes present in the BFP operon and was confirmed to express *bfpA* transcript while others were nearly undetectable. Therefore, isolate CE102 was deemed typical EPEC and removed from further analysis.
Figure 20: Final aEPEC Isolates. (A) Schematic of final aEPEC isolate determination. (B) Percent of isolates per age bracket grouped by symptomatic status and total distribution.

Figure 21: BFP and Per Operon Protein Homology. Protein homology within the BFP and Per operon proteins of EPEC isolates which were BfpA-positive by BV-BRC in silico proteomic analysis compared to pMAR2, the EAF plasmid of tEPEC-E2348/69. * indicates ASYM isolate.

Phylogenetic and Strain Lineage Determinant Analysis

In contrast to the SYMP and ASYM cohorts, each OB isolate did not originate from a unique donor, rather 13 donor stools yielded 36 OB isolates. Therefore, OB isolates underwent further analysis (SNP and Pangenome) to determine if isolates were clonal strains. OB isolates from the same donor with at least 3 of the 4 following criteria were considered to be clonal strains and removed from further analysis (total n=5 Table 6): ≥ 99.9% coverage, ≤ 100 high quality SNPs, ≥ 99.99% ANI, and the number of unique proteins between 2 isolates ≤ ~1%. 
Therefore, a total of 112 aEPEC isolates from the 3 cohorts as detailed in Figure 20A were analyzed for genetic virulence factors. Age distributions of individuals from which isolates originated were analyzed by symptomatic status and total distribution and were similar to the larger clinical cohort (Figure 20B). All OB isolates originated from individuals 18–49 years old due to FMT stool donor criteria.210

To understand their evolutionary relationship, historical lineage markers were assessed for 112 aEPEC isolates and association with symptomatic status were examined. Isolates originate from 6 of the 7 known E. coli phylogroups with B1 (39.3%), B2 (25%), and A (21.4%) predominating (Figure 22A). Although phylogroup A and B1 have more SYMP isolates than OB/ASYM combined, these differences are not statistically significant (Figure 22A and B) with near equal percentages of SYMP and OB/ASYM isolates in each phylogroup indicating that phylogroup does not predict symptomatic status.

![Figure 22: Phylogroup Does Not Predict Symptomatic Status for aEPEC Isolates. (A) Percentage of aEPEC isolates in each phylogroup with OB/ASYM (no fill) and SYMP (filled) pie slices. (B) Number of isolates in each phylogroup used for Chi-Square test.](image)

I next questioned whether core genome or other historical lineage markers such as serotype or Achtman MLST were predictive of symptomatic status. A phylogenetic tree based on 1000 core genes of the 112 aEPEC isolates along with 45 reference strains from different E. coli
phylogroups and pathovars was created. Of note, all OB isolates originate from donors living in the Boston, MA area whereas ASYM and SYMP isolates originate from LUMC inpatient and outpatient facilities serving the Chicagoland area. Despite the distinct geographical origins, core genome phylogenetics clustered according to phylogroup with OB, ASYM, and SYMP isolates distributed across the phylogenetic tree (Figure 23). The highly branched structure of the tree indicated great diversity among the isolates. In addition, lineage determinants underscore the great diversity of the aEPEC isolates with 50 known and 1 unknown Achtman MLSTs present and diverse serotypes present with 47 different O-antigens with 7 isolates from an unknown group and 26 different H-antigens. MLST and serotype clustered with similar isolates based on core genome (Figure 23), but overall, the high diversity in MLSTs and serotypes does not allow for prediction of symptomatic status. Of note, SYMP and ASYM isolates are numbered according to the order in which we received the stools, i.e. those with a lower CE number designation were acquired earlier and those with a higher CE designation were acquired later from the time span of 2016-2019. When examining the clustering and making special note of the CE number designation, isolates did not cluster according to temporal acquisition of stools as evidenced by low and high CE designations clustering together (Figure 23- red arrows) indicating these represent sporadic infections and are not indicative of outbreak infections.
Figure 23: Core Genome Phylogenetic Tree of aEPEC Isolates. RAxML phylogenetic tree based on 1000 core genes of aEPEC isolates and 45 reference strains with phylogroup indicated (A – orange; B1 – purple; B2 – red; D – green; E – blue; F – teal). Reference strains are colored according to their assigned phylogroup, SYMP (black), OB (light purple) and ASYM (light purple**). Red arrows denote examples of temporally distinct SYMP or ASYM (CE) isolates clustering together.

**Locus of Enterocyte Effacement (LEE) In Silico Proteomic Analysis**

To determine the known virulence factors present in the aEPEC isolates, the proteins encoded on the locus of enterocyte effacement (LEE) were examined using the *in silico* Proteomic analysis tool from BV-BRC218 and compared to both tEPEC E2348/69 and
EHEC_Sakai. Percent homology was manually curated and compiled in a heatmap with proteins ordered according to function and by symptomatic status with the range of percent homologies indicated (Figure 24). Overall, no clear clustering pattern is observed with most regulators, T3SS components, and chaperones having high homology to the reference strains regardless of symptomatic status. Unexpectedly, OB isolates appear to have greater homology in most proteins; greater diversity was observed in the effectors and translocators of SYMP isolates compared to OB isolates (Figure 24).

**Figure 24: Heatmap of LEE Protein Homology of aEPEC Isolates to tEPEC/EHEC Reference Strains.** Heatmap indicating percent identity as indicated in key to reference strains tEPEC_E2348/69 and EHEC_Sakai and is ordered by symptomatic status and proteins ordered by function. **R:** transcriptional regulators; **T3:** T3SS components; **C:** chaperones; **T:** translocon proteins; **E:** effectors.

To determine which LEE proteins have the greatest differences based on symptomatic status, the number of isolates with ≥80% homology were determined and compared between SYMP, ASYM, and OB isolates (Figure 25). Only 31% of OB isolates had EspA with ≥80%
homology, significantly fewer than ASYM and SYMP isolates. In contrast, a significantly greater number of SYMP isolates had reduced homology in the translocator, EspB, and the effectors Tir, EspH, and EspF (Figure 25). Although not significant, a greater number of SYMP isolates had more diversity in EscL, EspG, Map, EspZ, and CesF with ~10-20% differences between SYMP and OB isolates (Figure 25). Intimin was equally diverse with ~80% homology in all isolates (Figure 25). Contrary to my original hypothesis, OB isolates were more homologous to the reference stains whereas SYMP were less homologous.

![Differences in LEE Homology of aEPEC Isolates](image)

**Figure 25: Differences in LEE Homology of aEPEC Isolates.** The percentage of aEPEC isolates with ≥80% protein identity compared between OB, ASYM, and SYMP aEPEC isolates. Chi-square test based on number of isolates: * P < 0.05, ** P < 0.01, and **** P < 0.0001.

The evolutionary relationship of LEE lineages has been previously described in aEPEC isolates from seven study sites in Africa and Asia as part of the GEMS study. Ingle et. al. described three major LEE lineages with an additional 30 subtypes based on the individual phylogeny of seven LEE genes (eae, tir, espA, espB, espD, espH, and espZ). These are similar
to the most diverse proteins found in this study (Tir, EspA, EspB, EspH, and EspF) when compared by symptomatic status (Figure 25). However, Ingle et.al. did not report the symptomatic status of individuals harboring aEPEC isolates in their study, therefore it is unknown if certain LEE lineages or subtypes are associated with symptoms or severity of clinical manifestations of aEPEC infection. To determine if the LEE of aEPEC isolates of this study from the US were similar to previously described LEE lineages and subtypes and to examine symptomatic associations, individual protein trees were created for intimin, Tir, EspA,B,D, EspH, and EspZ along with reference subtypes described by Ingle et. al. Previous LEE subtyping was significantly associated with intimin subtypes\textsuperscript{175,176}; therefore, I initially analyzed intimin.

![Figure 26](image)

**Figure 26: Intimin Subtypes Are Not Associated with Symptomatic Status.** (A) Phylogenetic protein tree of intimin with major nodes indicated. (B) Reference intimin subtypes present in each intimin node as indicated. (C) Percent of isolates in each intimin node associated with OB, ASYM or SYMP status. Chi-square test based on number of isolates ($n$). *ns*: non-significant.
Our aEPEC isolates clustered into 6 major branches or nodes on the intimin phylogenetic tree (Figure 26A-red dots) with 22/31 known reference subtypes represented (Figure 26B). Together, these data indicate significant diversity among intimin of our aEPEC isolates. Although intimin nodes 2-5 had greater than 50% SYMP isolates, no significant association existed between intimin nodes and symptomatic status (Figure 26C) likely due to the diversity of subtypes present and insignificant n to detect differences (Figure 26B).

Next, the association with symptomatic status with Tir phylogeny was examined. Four major branches or nodes were identified for Tir (Figure 27A-boxed numbers and black dots). Tir nodes were significantly associated with symptomatic status with Tir Nodes 1 (14/18, 77.8%) and 2 (20/26, 76.9%) having the greatest number of SYMP isolates and significantly fewer associated with Tir Nodes 3 (10/16, 62.5%) and 4 (24/52, 46.2%) (Figure 27B). Sub-trees within Tir nodes could be ascribed to previously described Tir subtypes (Figure 27A -underlined). Interestingly, Tir subtypes correlated with specific intimin subtypes as indicated (Figure 27A) suggesting co-evolution of Tir-intimin pairs.
Figure 27: Tir Nodes are Significantly Associated with Symptomatic Status. (A) Phylogenetic protein tree of Tir with major nodes identified (boxed numbers and black dots), sub-trees ascribed to previously known Tir subtypes (underlined) and corresponding intimin subtypes denoted. (B) Percent of isolates in each Tir node associated with OB, ASYM or SYMP status. Chi-square test based on number of isolates (n).

Phylogeny of the remaining proteins (EspA, EspB, EspD, EspH, and EspZ) used to determine LEE subtypes were examined and major branches determined as indicated (Figure 28). Isolates found in the major branches of EspA,B,D, EspH, and EspZ correspond to the isolates found within the 4 major Tir Nodes (Figure 28- insets) indicating co-evolution of Tir and EspA,B,D, EspH, and EspZ. Indeed, a correlation matrix of protein homology indicated a positive correlation between all examined proteins except EspA, which was negatively correlated to all other protein homologies (Figure 29A). Tir homology displayed the strongest (Figure 29A) and most significant (Figure 29B) correlation with the other LEE proteins examined indicating a strong co-evolutionary relationship of Tir with the other effectors, translocators, intimin, and the chaperone, CesF. Furthermore, clear patterns in LEE protein homology clustering were evident when isolates were ordered according to the Tir protein phylogenetic tree (Figure 30).
Figure 28: Phylogenetic Protein Trees of EspA, EspB, EspD, EspH, and EspZ Are Associated with Tir Nodes. Major branches are indicated for each protein phylogenetic tree. Inset displays the percentage of individual branches in each Tir Node. Chi-square test for significance based on number in each node. All are significant **** $P < 0.0001$. 
Figure 29: LEE-Encoded Protein Homology Correlation Matrix. (A) The homology to the reference strains tEPEC/EHEC of each LEE-encoded protein was correlated to the homology of other LEE-encoded proteins as indicated. Matrix indicates Pearson R correlation values for each individual pair of protein homologies examined. X denotes non-significant. (B) Significance matrix for each of the Pearson correlation tests run in A. **** \( P < 0.0001 \), *** \( P < 0.002 \), ** \( P < 0.01 \), * \( P < 0.05 \).
Figure 30: Heatmap of LEE Protein Homology of aEPEC Isolates to Reference Strains tEPEC/EHEC Clustered According to Tir Nodes. Heatmap indicating percent identity as indicated in key to reference strains tEPEC_E2348/69 and EHEC_Sakai ordered by Tir Node and phylogeny order.

To further characterize the Tir Nodes, subtypes of the 7 proteins were curated and LEE lineages and subtypes determined (Table 7). Forty-two (37.5%) aEPEC isolates had novel LEE subtypes with Tir Nodes 1-3 having the greatest percentage of new LEE subtypes (Table 7). Tir Node 4 had the majority of previously described LEE subtypes (Table 7). These data suggest that specific Tir Nodes and subtypes and the corresponding LEE subtypes likely contribute to variance in disease manifestations of aEPEC infection.

Table 7: LEE Lineage and Sequence Type Determination of aEPEC Isolates

<table>
<thead>
<tr>
<th>Strain</th>
<th>Tir Node</th>
<th>Tir</th>
<th>Intimin</th>
<th>EspA</th>
<th>EspB</th>
<th>EspD</th>
<th>EspH</th>
<th>EspZ</th>
<th>LEE lineage</th>
<th>LEE SEQ type</th>
</tr>
</thead>
<tbody>
<tr>
<td>121</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>11</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>245</td>
<td>1</td>
<td>1</td>
<td>omicron</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>novel</td>
</tr>
<tr>
<td>180</td>
<td>1</td>
<td>1</td>
<td>omicron</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>novel</td>
</tr>
<tr>
<td>179</td>
<td>1</td>
<td>1</td>
<td>omicron</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>novel</td>
</tr>
<tr>
<td>241</td>
<td>1</td>
<td>1</td>
<td>omicron</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>novel</td>
</tr>
<tr>
<td>218</td>
<td>1</td>
<td>1</td>
<td>omicron</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>novel</td>
</tr>
<tr>
<td>126**</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>35</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>232</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>110</td>
<td>1</td>
<td>5</td>
<td>ypsilon</td>
<td>11</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>8</td>
<td>novel</td>
<td>novel</td>
</tr>
<tr>
<td>119</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>8</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>118</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>8</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>85</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>8</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>297**</td>
<td>1</td>
<td>5</td>
<td>9</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>14</td>
<td>3</td>
<td>13/14-novel</td>
</tr>
<tr>
<td>34</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>8</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>280</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>8</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>OB126</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>8</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>OB84</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>8</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>214**</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>201</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>166</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>350</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>OB84</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-----</td>
<td>---</td>
</tr>
<tr>
<td>OB86</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>376</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>82**</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>347</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>318</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>103</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>444**</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>10</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>186</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>10</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>99</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>10</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>133</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>288</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>10</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>98</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>10</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>152</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>10</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>336**</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>10</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>304</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>435</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>10</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>234</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>10</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>285</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>268</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>363</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>10</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>254</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>165</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>171</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>97**</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>445</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>OB51</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>ob6</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>23</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>OB63.1</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>239</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>OB127</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>243</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>174</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>279</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>86**</td>
<td>3</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>14</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>OB41</td>
<td>4</td>
<td>7</td>
<td>10</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>8</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>267</td>
<td>4</td>
<td>7</td>
<td>10</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>14</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>365**</td>
<td>4</td>
<td>7</td>
<td>12</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>14</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>37</td>
<td>4</td>
<td>7</td>
<td>12</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>14</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>46</td>
<td>4</td>
<td>7</td>
<td>11</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>13</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>211</td>
<td>4</td>
<td>7</td>
<td>11</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>13</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>237**</td>
<td>4</td>
<td>7</td>
<td>11</td>
<td>8</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>13</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
</tr>
<tr>
<td>224</td>
<td>4</td>
<td>11</td>
<td>18</td>
<td>9</td>
<td>5</td>
<td>8</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>4</td>
<td>10</td>
<td>18</td>
<td>9</td>
<td>5</td>
<td>8</td>
<td>7</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>334</td>
<td>4</td>
<td>10</td>
<td>18</td>
<td>9</td>
<td>5</td>
<td>8</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>193**</td>
<td>4</td>
<td>10</td>
<td>16</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>379</td>
<td>4</td>
<td>10</td>
<td>16</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>210</td>
<td>4</td>
<td>10</td>
<td>17</td>
<td>9</td>
<td>5</td>
<td>8</td>
<td>7</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB124</td>
<td>4</td>
<td>10</td>
<td>19</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB128</td>
<td>4</td>
<td>10</td>
<td>19</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB154</td>
<td>4</td>
<td>10</td>
<td>19</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB129</td>
<td>4</td>
<td>10</td>
<td>19</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB137</td>
<td>4</td>
<td>10</td>
<td>19</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>261</td>
<td>4</td>
<td>10</td>
<td>19</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>295</td>
<td>4</td>
<td>10</td>
<td>19</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>227</td>
<td>4</td>
<td>10</td>
<td>19</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>223</td>
<td>4</td>
<td>10</td>
<td>19</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB7</td>
<td>4</td>
<td>10</td>
<td>19</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB21</td>
<td>4</td>
<td>10</td>
<td>19</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB166.3</td>
<td>4</td>
<td>10</td>
<td>19</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>147</td>
<td>4</td>
<td>10</td>
<td>19</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB25</td>
<td>4</td>
<td>10</td>
<td>19</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB136.1</td>
<td>4</td>
<td>10</td>
<td>19</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>4</td>
<td>8</td>
<td>13</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>238</td>
<td>4</td>
<td>8</td>
<td>13</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>4</td>
<td>8</td>
<td>13</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>116**</td>
<td>4</td>
<td>8</td>
<td>13</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>246</td>
<td>4</td>
<td>8</td>
<td>13</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>131</td>
<td>4</td>
<td>8</td>
<td>13</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>170</td>
<td>4</td>
<td>8</td>
<td>13</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>374</td>
<td>4</td>
<td>8</td>
<td>13</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>8</td>
<td>13</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB2</td>
<td>4</td>
<td>8</td>
<td>13</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB13</td>
<td>4</td>
<td>8</td>
<td>14</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB56</td>
<td>4</td>
<td>8</td>
<td>14</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB20</td>
<td>4</td>
<td>8</td>
<td>14</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>4</td>
<td>8</td>
<td>14</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB15</td>
<td>4</td>
<td>8</td>
<td>14</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB38</td>
<td>4</td>
<td>8</td>
<td>14</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB33</td>
<td>4</td>
<td>8</td>
<td>14</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>270**</td>
<td>4</td>
<td>9</td>
<td>15</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB131</td>
<td>4</td>
<td>9</td>
<td>15</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB145</td>
<td>4</td>
<td>9</td>
<td>15</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB148</td>
<td>4</td>
<td>9</td>
<td>15</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OB36.2</td>
<td>4</td>
<td>9</td>
<td>15</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Having determined that Tir Nodes 1 and 2 had greater numbers of SYMP isolates, I examined if Tir Node contributed to relative EPEC load, or disease manifestations among SYMP isolates including diarrhea duration, number of bowel movements in 24 hrs, presence of fever, and diarrhea severity as assessed by the modified Vesikari score (mVS). Relative EPEC loads were compared between Tir Nodes (Figure 31A). Tir Node 2 had significantly higher average EPEC load (3.67% ±1.95% SEM) compared to Tir Node 3 (1.63% ± 1.36% SEM, P = 0.464) and Tir Node 4 (1.12% ± 0.63% SEM, P = 0.0029) (Figure 31A). Tir Node 1 had the lowest average EPEC load (0.553% ± 0.33% SEM), however, no significant differences with the other Tir Nodes were detected likely due to insufficient n in Tir Node 1 (n=7) (Figure 31A). Differences between Tir Nodes were also examined for other disease manifestations among SYMP isolates. Presence of emesis was significantly associated with Tir Node 1 isolates (n=10/14, 71.4%, Fisher’s exact test, P = 0.0386) in contrast Tir Nodes 2-4 isolates had significantly fewer cases of emesis (n=7/20, 35%; n=2/30, 20%; n=7/24, 29.2%, respectively) (Figure 31B). No significant differences in diarrhea duration, #BM/24hrs, or mVS were detected between isolates from the different Tir Nodes (data not shown). These data indicate that Tir Node status alone does not dictate symptom variation or severity of aEPEC infection.
Figure 31: Disease Manifestations of aEPEC Infection Associated with Tir Nodes. (A) Relative EPEC loads compared between Tir Nodes. Kruskal-Wallis with uncorrected Dunn’s test. (B) Percentage of SYMP isolates which originate from individuals who experienced emesis during aEPEC infection and compared by Tir Node status. Fisher’s exact test. Number (n) of those positive for emesis displayed.

Non-LEE In Silico Proteomic Analysis

In addition to LEE virulence factors, between 34-59 known non-LEE effectors have been described in tEPEC and EHEC.\textsuperscript{212-217} In silico proteomic analysis was performed on all 59 proteins, curated, and the number present in each isolate compared by symptomatic status and Tir Node (Figure 32). The number of non-LEE effectors was significantly higher in SYMP isolates (16.4 ± 1.07 SEM) compared to OB isolates (11.3 ± 1.33 SEM) with ASYM isolates (15.2 ± 1.50 SEM) having intermediate numbers of non-LEE effectors present (Figure 32A). In addition, Tir Node 2 (21.4 ± 1.09 SEM) had the highest number of non-LEE effectors present, Tir Node 4 (10.3 ± 0.96 SEM) had the least, and Tir Node 3 and 1 had intermediate numbers (18.4 ± 1.15 SEM and 15.5 ± 2.44 SEM), respectively) (Figure 32B). These data indicate that non-LEE effectors likely contribute to variable manifestations of aEPEC infection.
Figure 32: Non-LEE Virulence Factors Are More Prevalent in SYMP and Tir Node 2 aEPEC Isolates. Number of non-LEE effectors in each group based on (A) symptomatic status and (B) Tir Nodes. 1-way ANOVA; **** $P < 0.0001$, *** $P < 0.002$, * $P < 0.05$.

To determine which specific non-LEE effectors were associated with symptomatic status and Tir Nodes, we first examined the 34 non-LEE effectors and homologues common to both tEPEC_E2348/69 and EHEC_Sakai listed in Figure 33 and designated as tEPEC non-LEE effectors. Symptomatic isolates were more likely to house NleB, NleE, LifA, and EspL2 from the integrative element 6 (IE6) in addition to NleG from prophage 4 (PP4), its homologues NleG5-1 and NleG6-1, and EspJ from prophage 6 (PP2) (Figure 33 and Table 8). As seen on the protein identity heatmap of tEPEC non-LEE effectors, 20 tEPEC non-LEE effectors were significantly associated with Tir Nodes and in most cases, they were more prominent in Tir Node 2 or Tir Node 3 or both (Figure 33). Tir Node 4 had fewer non-LEE effectors, especially in those isolates clustered in Tir subtype 10 (CE150, CE334, CE193*, CE379, CE210, OB124, OB128, OB154, OB129, OB137, CE261, CE295, CE227, CE223, OB7, CE221, OB166.3, CE147, OB25, OB136.1) (Figure 33).
Figure 33: tEPEC Non-LEE Effectors of aEPEC Isolates. Percent identity heatmap of non-LEE effectors from EPEC/EHEC ordered by Tir Node clustering (x-axis) and by significance according to Tir Node analysis (y-axis). Red boxes indicate those which are significantly associated with SYMP status. Those delineated by line and asterisks are significantly associated with Tir Node. Fisher’s exact test; **** P < 0.0001, *** P < 0.002, ** P < 0.01, * P < 0.05. Details of significance based on symptomatic status are in Table 8.

Table 8: Non-LEE Effectors and Adhesins Associated with Symptomatic Statusa

<table>
<thead>
<tr>
<th>Type</th>
<th>OB n</th>
<th>ASYM n</th>
<th>SYMP n</th>
<th>Sig.</th>
<th>P-value</th>
<th>Source of analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>aEPEC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lda fimbrial</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>ns</td>
<td>&gt;0.9999 BV-BRC</td>
</tr>
<tr>
<td>tEPEC/EHEC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EspI_PP2</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>NleB_IE6</td>
<td>3</td>
<td>9.68</td>
<td>3</td>
<td>23.08</td>
<td>*</td>
<td>0.0247 Figure 33</td>
</tr>
<tr>
<td>NleE_IE6</td>
<td>4</td>
<td>12.90</td>
<td>3</td>
<td>23.08</td>
<td>*#</td>
<td>0.0295 Figure 33</td>
</tr>
<tr>
<td>EspL2_IE6</td>
<td>3</td>
<td>9.68</td>
<td>3</td>
<td>23.08</td>
<td>*</td>
<td>0.0247 Figure 33</td>
</tr>
<tr>
<td>NleG_PP4</td>
<td>0</td>
<td>0.00</td>
<td>1</td>
<td>7.69</td>
<td>**</td>
<td>0.0052 Figure 33</td>
</tr>
<tr>
<td>NleG5-1</td>
<td>0</td>
<td>0.00</td>
<td>1</td>
<td>7.69</td>
<td>***</td>
<td>0.0007 Figure 33</td>
</tr>
</tbody>
</table>

a: Details of significance based on symptomatic status are in Table 8.
### NleG6-1 effector EHEC

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th>32.35</th>
<th></th>
<th>0.001</th>
<th>Figure 33</th>
</tr>
</thead>
<tbody>
<tr>
<td>toxB</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>ns &gt;0.9999</td>
</tr>
<tr>
<td>lifA/efa1 IE2</td>
<td>3</td>
<td>9.68</td>
<td>3</td>
<td>23.08</td>
<td>15</td>
<td>22.06</td>
<td>ns 0.3034</td>
</tr>
<tr>
<td>lifA/efa1 IE6</td>
<td>0</td>
<td>0.00</td>
<td>2</td>
<td>15.38</td>
<td>22</td>
<td>32.35</td>
<td>***# 0.0003</td>
</tr>
<tr>
<td>efal/lifA_EHEC</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>4</td>
<td>5.88</td>
<td>ns 0.4315</td>
</tr>
<tr>
<td>lifA/efa1 any</td>
<td>3</td>
<td>9.68</td>
<td>4</td>
<td>30.77</td>
<td>27</td>
<td>39.71</td>
<td>** 0.0068</td>
</tr>
<tr>
<td>espC</td>
<td>11</td>
<td>35.48</td>
<td>7</td>
<td>53.85</td>
<td>20</td>
<td>44.12</td>
<td>ns 0.5004</td>
</tr>
</tbody>
</table>

### espC toxin

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Figure 33</th>
</tr>
</thead>
<tbody>
<tr>
<td>agn43</td>
<td>9</td>
<td>29.03</td>
<td>6</td>
<td>46.15</td>
<td>31</td>
<td>45.59</td>
<td>ns 0.2679</td>
</tr>
<tr>
<td>ecpA</td>
<td>28</td>
<td>90.32</td>
<td>11</td>
<td>84.62</td>
<td>64</td>
<td>94.12</td>
<td>ns 0.3369</td>
</tr>
<tr>
<td>lpfA</td>
<td>1</td>
<td>3.23</td>
<td>1</td>
<td>7.69</td>
<td>4</td>
<td>5.88</td>
<td>ns 0.8417</td>
</tr>
<tr>
<td>lpf major subunit LpfA</td>
<td>6</td>
<td>19.35</td>
<td>4</td>
<td>30.77</td>
<td>15</td>
<td>22.06</td>
<td>ns 0.7096</td>
</tr>
<tr>
<td>lpf major subunit LpfA-O113</td>
<td>9</td>
<td>29.03</td>
<td>7</td>
<td>53.85</td>
<td>17</td>
<td>25.00</td>
<td>ns 0.1306</td>
</tr>
<tr>
<td>lpfA1</td>
<td>4</td>
<td>12.90</td>
<td>2</td>
<td>15.38</td>
<td>7</td>
<td>10.29</td>
<td>ns 0.75</td>
</tr>
<tr>
<td>lpfA2</td>
<td>2</td>
<td>6.45</td>
<td>1</td>
<td>7.69</td>
<td>7</td>
<td>10.29</td>
<td>ns 0.887</td>
</tr>
<tr>
<td>elfA</td>
<td>8</td>
<td>25.81</td>
<td>4</td>
<td>30.77</td>
<td>22</td>
<td>32.35</td>
<td>ns 0.8306</td>
</tr>
<tr>
<td>paa</td>
<td>5</td>
<td>16.13</td>
<td>5</td>
<td>38.46</td>
<td>26</td>
<td>38.24</td>
<td>* 0.0355</td>
</tr>
<tr>
<td>iha</td>
<td>3</td>
<td>9.68</td>
<td>0</td>
<td>0.00</td>
<td>7</td>
<td>10.29</td>
<td>ns 0.7834</td>
</tr>
<tr>
<td>hcpA</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>ns &gt;0.9999</td>
</tr>
<tr>
<td>fliC</td>
<td>31</td>
<td>100</td>
<td>13</td>
<td>100</td>
<td>69</td>
<td>100</td>
<td>ns &gt;0.9999</td>
</tr>
<tr>
<td>csgE</td>
<td>31</td>
<td>100</td>
<td>13</td>
<td>100</td>
<td>68</td>
<td>100</td>
<td>ns &gt;0.9999</td>
</tr>
<tr>
<td>fimH</td>
<td>31</td>
<td>100</td>
<td>13</td>
<td>100</td>
<td>67</td>
<td>100</td>
<td>ns &gt;0.9999</td>
</tr>
</tbody>
</table>

### DAEC

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Figure 33</th>
</tr>
</thead>
<tbody>
<tr>
<td>afaBC (afaA)</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>1</td>
<td>1.47</td>
<td>ns &gt;0.9999</td>
</tr>
<tr>
<td>daaC(daaA)</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>1</td>
<td>1.47</td>
<td>ns &gt;0.9999</td>
</tr>
</tbody>
</table>

### UPEC

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Figure 33</th>
</tr>
</thead>
<tbody>
<tr>
<td>papC</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>3</td>
<td>4.41</td>
<td>ns 0.6902</td>
</tr>
<tr>
<td>sfaDE (sfaE)</td>
<td>0</td>
<td>0.00</td>
<td>2</td>
<td>15.38</td>
<td>7</td>
<td>10.29</td>
<td>ns 0.0733</td>
</tr>
</tbody>
</table>

\* Bolded are significantly different based on SYMP status with the greatest percentage of positive isolates shaded in blue or equally prevalent in grey. Parenthesis indicate gene detected.

\* Fisher's exact test comparing OB, ASYM, and SYMP; # comparison between OB vs. SYMP.
Next, non-LEE effectors originally identified only in *C. rodentium* (EspS,V)\(^{214,215}\), aEPEC_E110019 (EspT)\(^{213}\), tEPEC_B171(Cif)\(^{217}\), and EHEC (all others)\(^{216}\) as listed in Figure 34 were examined and designated as other non-LEE effectors. No significant differences were found based on symptomatic status (data not shown). However, significant associations with Tir Nodes were again identified (Figure 34). Homologues of EspR, EspX, and EspY were more prevalent in Tir Nodes 1-3 with less in Tir Node 4 especially in those isolates clustering with Tir subtype 10 (Figure 34). These data reiterate the increased presence of virulence factors in Tir Nodes 1-3 and the least in Tir Node 4.

**Figure 34: Other Non-LEE Effectors of aEPEC Isolates Associated with Tir Node, but Not SYMP Status.** Percent identity heatmap of other effectors from EHEC, *C. rodentium*, tEPEC-B171, and aEPEC-E110019. Significant values indicated and in bold are associated with Tir Node, but not SYMP status. Fisher’s Exact test; **** \( P < 0.0001 \), *** \( P < 0.002 \), * \( P < 0.05 \).

Having identified the number and types of non-LEE effectors present, I asked whether the number of all or individual categories of non-LEE effectors were correlated with clinical
manifestations of aEPEC infection. I examined continuous variables of clinical manifestations such as diarrhea duration, number (#) of bowel movements (BMs) in a 24hr period, and diarrhea severity assessed with the modified Vesikari score (mVS) (Figure 35). Diarrhea severity increased with the number of tEPEC non-LEE effectors present with a weak, but significantly positive correlation detected (R = 0.23, P = 0.015) (Figure 35).

Figure 35: Correlation Matrix of Non-LEE Effectors Present Versus Clinical Manifestation of aEPEC Infection. Pearson correlation R value displayed for each combination. * P = 0.015. tEPEC/EHEC – non-LEE effectors and homologues common to both tEPEC-E2348/69 and EHEC-Sakai (n=34); others – non-LEE effectors not found in tEPEC-E2348/69 (n=25); combined – all 59 non-LEE effectors.

Analysis of Known Adhesins

aEPEC likely have acquired additional adhesins or certain combinations of adhesins to compensate for the lack of BFP allowing for attachment to host intestinal epithelial cells. To determine the number of adhesins and the type of adhesins present in aEPEC isolates, annotated whole genome sequences were analyzed for genes annotated as adhesins, autotransporter (AT) adhesins, and fimbrial proteins using Geneious software and corroborated with BV-BRC Feature search, and then assessed for association with symptomatic status and distribution in Tir Nodes. Initially focused on the adhesin genes listed in Table 5 of characterized known adhesins then expanded the analysis to all detected.
First, I asked if associations were present between genes annotated as adhesins and symptomatic status. Paa (from Table 5) was the only adhesin that positively associated with SYMP isolates \((n=26 \ (38.2\%))\), \(p=0.0097\) Fisher’s Exact Test) compared to OB isolates \((n=5 \ (16.1\%))\); ASYM isolates \((n=5 \ (38.5\%))\) were similar to SYMP. In contrast, the CfaE/CblD family pilus tip adhesin was associated with OB isolates \((n=4 \ (12.9\%)\) \(p=0.0097\) vs. \(n=0 \ (0\%\)) in SYMP and ASYM; Fisher’s Exact Test). Fifteen additional adhesin genes (listed in Figure 36) were annotated in the aEPEC isolates but none had associations with symptomatic status including \(iha\), previously described in aEPEC isolates (from Table 5).\(^9\) Similarly, only 2 adhesins were associated with Tir Nodes. Surface-adhesin E family protein \((saFE)\) was positively associated with Tir Node 1 isolates \((n=5 \ (27.8\%))\), \(p=0.0141\), Fisher’s Exact Test) compared to Tir Nodes 2-4 \((n=2 \ (7.69\%); \ n=0 \ (0\%); \ n=2 \ (3.85\%),\) respectively) (Figure 36). \(sfaDE\) are part of the S Fimbriae associated with adherence of UPEC and associated with some aEPEC isolates (from Table 5).\(^9,224\) In contrast, the DUF823 domain-containing adhesin was associated with Tir Node 4 isolates \((n=25 \ (48.1\%) \ p<0.0001;\) Fisher’s Exact Test) compared to Tir Node 1-3 \((n=4 \ (22.2\%); \ n=3 \ (0\%); \ n=0 \ (11.5\%),\) respectively) (Figure 36).

Of note, several isolates have genes that have been associated with specific adherence patterns of different \(E. \ coli\) pathovars. Two SYMP isolates in Tir Node 4, CE37 and CE224, house several proteins associated with diffuse adherence \((afaA, \ daaA, \ daaF, \ draB, \ draD)\)\(^84\) (Figure 36 and Table 8). Interestingly, OB137 harbors genes associated with aggregative adherence \((aggA,B,C,D)\).\(^80\) Two isolates (CE239, CE435) contain the gene \(air\) whose protein product is involved as an accessory adhesin and aggregin of EAEC\(^225\) (Figure 36).
Figure 36: Adhesins of aEPEC Isolates. Presence/absence heatmap of adhesin genes detected in aEPEC isolates. Absent = 0; present = 1. Bolded adhesin genes are significantly associated with Tir Nodes.

Fourteen autotransporter (AT) adhesins (listed in Figure 37) were annotated in aEPEC isolates. Only one was different based on symptomatic status, ehaB, and it had a greater association with OB isolates \( n=27 (87.1\%) \) \( p=0.003 \); Fisher’s Exact Test) than SYMP or ASYM isolates \( n=36 (52.9\%); n=8 (61.5\%) \), respectively). Of note, agn43, an AT adhesin previously described in aEPEC \(^9\) (Table 5), was not associated with symptomatic isolates of this study. In contrast, 8/14 AT adhesin genes were associated with Tir Nodes (Figure 37). Tir Node 1 had the greatest percentage of isolates with AT adhesin genes with two genes \((agn43, ehaJ)\) predominately in Tir Node 1 alone (Figure 37 and Table – blue shading). In addition, two genes were equally high in Tir Nodes 1-3 \((ypjA)\) and 1/2 (AT family) (Figure 37 and Table 9 – grey shading). Tir Node 3 had two AT adhesin genes \((veeJ, iatC)\) most predominate. Although Tir Node 4 had greater detection of ehaB compared to the other Nodes (Figure 37 and Table 9), Tir
Node 4 was significantly lacking in AT adhesins mostly due to the absence in Tir subtype 10 (Figure 37).

**Figure 37: Autotransporter Adhesins of aEPEC Isolates.** Presence/absence heatmap of autotransporter adhesin genes detected in aEPEC isolates. Absent = 0; present = 1. Bolded autotransporter adhesin genes are significantly associated with Tir Nodes.

![Autotransporter Adhesins of aEPEC Isolates](image)

**Table 9: Autotransporter Adhesin Genes Significantly Associated with Tir Node**

<table>
<thead>
<tr>
<th></th>
<th>Tir Node 1</th>
<th>Tir Node 2</th>
<th>Tir Node 3</th>
<th>Tir Node 4</th>
<th>Sig. b</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>agn43</td>
<td>11 61.1</td>
<td>12 46.2</td>
<td>7 43.8</td>
<td>16 30.8</td>
<td>#</td>
<td>0.0207</td>
</tr>
<tr>
<td>yojA</td>
<td>9 50.0</td>
<td>15 57.7</td>
<td>9 56.3</td>
<td>5 9.6</td>
<td>****</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>yeeJ</td>
<td>4 22.2</td>
<td>10 38.5</td>
<td>11 68.8</td>
<td>11 21.2</td>
<td>**</td>
<td>0.004</td>
</tr>
<tr>
<td>AT family</td>
<td>9 50.0</td>
<td>14 53.8</td>
<td>5 31.3</td>
<td>9 17.3</td>
<td>**</td>
<td>0.0035</td>
</tr>
<tr>
<td>ehaB</td>
<td>13 72.2</td>
<td>5 19.2</td>
<td>7 43.8</td>
<td>46 88.5</td>
<td>****</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>ehaJ</td>
<td>4 22.2</td>
<td>0 0</td>
<td>0 0</td>
<td>8 15.4</td>
<td>*</td>
<td>0.0203</td>
</tr>
<tr>
<td>YadA family AT</td>
<td>7 38.9</td>
<td>3 11.5</td>
<td>0 0</td>
<td>18 34.6</td>
<td>**</td>
<td>0.0032</td>
</tr>
<tr>
<td>iatC</td>
<td>0 0</td>
<td>0 0</td>
<td>2 12.5</td>
<td>0 0</td>
<td>*</td>
<td>0.0193</td>
</tr>
</tbody>
</table>

a Those with greatest percentage of positive isolates shaded in blue and with equally high percentage in grey

b All Fisher's Exact test except agn43 # Chi-square test for trend

Eighty-eight *fimbrial protein genes* were annotated in the 112 aEPEC isolates (Figure 38 and 39). However, none were significantly associated with SYMP status, including those previously found in aEPEC (*ecpA, lpf, elfA*) (from Table 5). The Fim operon (*fimA-I,Z*) and
other broad categories of fimbrial protein genes (Figure 38) and other adhesins previously found in aEPEC isolates (\(fliC\) _flagella and \(csgE\) _curli\))\(^9\) (from Table 5) were prevalent in all aEPEC isolates, thus no significance was detected. However, isolates of Tir subtype 10 lacked \(fimZ\) and the broad category of “fimbrial family protein”, contrary to most other isolates and supports the notion of decreased known virulence factors in these specific isolates. Some previously described fimbrial proteins found in aEPEC were either not detected (\(lda\) _aEPEC, \(hcpA\) _pili of EHEC) or only present in a few (\(papC\) _P fimbriae of UPEC) (from Table 5).\(^9\) Only one, the CS1fimbrial protein, was significantly associated with OB isolates \((n=4\ 12.9\%\ p=0.0097\ vs. n=0\ 0\%\) in SYMP and ASYM; Fisher’s Exact Test).

**Figure 38: Fim Operon (Type I Fimbriae) and Other Fimbrial Protein Genes of aEPEC Isolates.** Presence/absence heatmap of fimbrial protein genes detected in aEPEC isolates. Absent = 0; present = 1. None are significantly associated with SYMP status or Tir Nodes.

In contrast, nearly half \((n=36)\) of the fimbrial protein genes were significantly associated with Tir Node (Figure 39 and Table 10). In most instances, Tir Node 2 had the greatest percentage of isolates with the specified fimbrial protein gene \((n=22)\) (Figure 39 and Table 10 –
blue shading). In 9 instances, Tir Node 2 had an equally high percentage of isolates with fimbrial protein genes compared to Tir Node 3 or 1 or both (Table 10- grey shading). Two fimbrial protein genes predominated in Tir Node 3 (ycfV and lpfA2), while Tir Node 1 had the greatest percentage of isolates with lpfE and the gene designation as an outer membrane usher protein (Table 10). Importantly, Tir Node 4 always had the least number of fimbrial protein genes (Table 10) and this was predominately due to those aEPEC isolates in Tir subtype 10 (CE150, CE334, CE193*, CE379, CE210, OB124, OB128, OB154, OB129, OB137, CE261, CE295, CE227, CE223, OB7, CE221, OB166.3, CE147, OB25, OB136.1) and 2 isolates in Tir subtype 8 (CE29, CE238) (Figure 39).
Figure 39: Fimbrial Protein Genes of aEPEC Isolates. Presence/absence heatmap of fimbrial protein genes detected in aEPEC isolates. Absent = 0; present = 1. Bolded fimbrial proteins are significantly associated with Tir Nodes with details in Table 10.

Table 10: Fimbrial Protein Genes Significantly Associated with Tir Node\(^ b\)

<table>
<thead>
<tr>
<th>Protein</th>
<th>Tir Node 1 ( n ) (%)</th>
<th>Tir Node 2 ( n ) (%)</th>
<th>Tir Node 3 ( n ) (%)</th>
<th>Tir Node 4 ( n ) (%)</th>
<th>Sig.</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ydeT</td>
<td>4 22.2</td>
<td>10 38.5</td>
<td>3 18.8</td>
<td>2 3.8</td>
<td>***</td>
<td>0.0006</td>
</tr>
<tr>
<td>elfA</td>
<td>9 50.0</td>
<td>13 50.0</td>
<td>9 56.3</td>
<td>3 5.8</td>
<td>****</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>elfC</td>
<td>14 77.8</td>
<td>25 96.2</td>
<td>16 100.0</td>
<td>26 50.0</td>
<td>****</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>elfG</td>
<td>14 77.8</td>
<td>24 92.3</td>
<td>16 100.0</td>
<td>26 50.0</td>
<td>****</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>sfmD</td>
<td>13 72.2</td>
<td>25 96.2</td>
<td>15 93.8</td>
<td>25 48.1</td>
<td>****</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>sfmF</td>
<td>14 77.8</td>
<td>25 96.2</td>
<td>15 93.8</td>
<td>27 51.9</td>
<td>****</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>yadC</td>
<td>2 11.1</td>
<td>9 34.6</td>
<td>4 25.0</td>
<td>2 3.8</td>
<td>**</td>
<td>0.0013</td>
</tr>
<tr>
<td>yadK</td>
<td>2 11.1</td>
<td>8 30.8</td>
<td>4 25.0</td>
<td>1 1.9</td>
<td>****</td>
<td>0.0006</td>
</tr>
<tr>
<td>yadL</td>
<td>9 50.0</td>
<td>21 80.8</td>
<td>11 68.8</td>
<td>48 92.3</td>
<td>***</td>
<td>0.001</td>
</tr>
<tr>
<td>yadM</td>
<td>5 27.8</td>
<td>21 80.8</td>
<td>11 68.8</td>
<td>23 44.2</td>
<td>**</td>
<td>0.001</td>
</tr>
<tr>
<td>htrE</td>
<td>6 33.3</td>
<td>22 84.6</td>
<td>11 68.8</td>
<td>43 82.7</td>
<td>***</td>
<td>0.0009</td>
</tr>
<tr>
<td>yadN</td>
<td>6 33.3</td>
<td>22 84.6</td>
<td>11 68.8</td>
<td>24 46.2</td>
<td>**</td>
<td>0.001</td>
</tr>
<tr>
<td>yadV</td>
<td>5 27.8</td>
<td>22 84.6</td>
<td>11 68.8</td>
<td>24 46.2</td>
<td>***</td>
<td>0.0004</td>
</tr>
<tr>
<td>ybgD</td>
<td>3 16.7</td>
<td>8 30.8</td>
<td>3 18.8</td>
<td>1 1.9</td>
<td>**</td>
<td>0.0012</td>
</tr>
<tr>
<td>ybgO</td>
<td>14 77.8</td>
<td>26 100.0</td>
<td>15 93.8</td>
<td>24 46.2</td>
<td>****</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>ybgQ</td>
<td>9 50.0</td>
<td>24 92.3</td>
<td>10 62.5</td>
<td>23 44.2</td>
<td>***</td>
<td>0.0002</td>
</tr>
<tr>
<td>ycbF</td>
<td>14 77.8</td>
<td>24 92.3</td>
<td>16 100.0</td>
<td>26 50.0</td>
<td>****</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>ycbU</td>
<td>14 77.8</td>
<td>25 96.2</td>
<td>16 100.0</td>
<td>26 50.0</td>
<td>****</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>ycbV</td>
<td>14 77.8</td>
<td>25 96.2</td>
<td>16 100.0</td>
<td>26 50.0</td>
<td>****</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>yfcP</td>
<td>3 16.7</td>
<td>8 30.8</td>
<td>4 25.0</td>
<td>1 1.9</td>
<td>***</td>
<td>0.0006</td>
</tr>
<tr>
<td>yfcQ</td>
<td>3 16.7</td>
<td>8 30.8</td>
<td>4 25.0</td>
<td>1 1.9</td>
<td>***</td>
<td>0.0006</td>
</tr>
<tr>
<td>yfcR</td>
<td>3 16.7</td>
<td>8 30.8</td>
<td>4 25.0</td>
<td>1 1.9</td>
<td>***</td>
<td>0.0006</td>
</tr>
<tr>
<td>yfcS</td>
<td>6 33.3</td>
<td>19 73.1</td>
<td>7 43.8</td>
<td>23 44.2</td>
<td>*</td>
<td>0.0368</td>
</tr>
<tr>
<td>yfcU</td>
<td>9 50.0</td>
<td>21 80.8</td>
<td>11 68.8</td>
<td>25 48.1</td>
<td>*</td>
<td>0.0277</td>
</tr>
<tr>
<td>yfcV</td>
<td>14 77.8</td>
<td>20 76.9</td>
<td>16 100.0</td>
<td>26 50.0</td>
<td>****</td>
<td>0.0003</td>
</tr>
<tr>
<td>yhcA</td>
<td>6 33.3</td>
<td>8 30.8</td>
<td>4 25.0</td>
<td>2 3.8</td>
<td>***</td>
<td>0.001</td>
</tr>
<tr>
<td>yhcD</td>
<td>6 33.3</td>
<td>8 30.8</td>
<td>4 25.0</td>
<td>2 3.8</td>
<td>***</td>
<td>0.001</td>
</tr>
<tr>
<td>yraH</td>
<td>14 77.8</td>
<td>21 80.8</td>
<td>11 68.8</td>
<td>21 40.4</td>
<td>**</td>
<td>0.0012</td>
</tr>
<tr>
<td>yraJ</td>
<td>13 72.2</td>
<td>20 76.9</td>
<td>11 68.8</td>
<td>21 40.4</td>
<td>**</td>
<td>0.0054</td>
</tr>
<tr>
<td>yraK</td>
<td>13 72.2</td>
<td>21 80.8</td>
<td>11 68.8</td>
<td>22 42.3</td>
<td>**</td>
<td>0.0045</td>
</tr>
<tr>
<td>yqiI</td>
<td>7 38.9</td>
<td>20 76.9</td>
<td>11 68.8</td>
<td>25 48.1</td>
<td>*</td>
<td>0.0262</td>
</tr>
<tr>
<td>lpfA1</td>
<td>5 27.8</td>
<td>2 7.7</td>
<td>5 31.3</td>
<td>1 1.9</td>
<td>***</td>
<td>0.0006</td>
</tr>
<tr>
<td>lpfA2</td>
<td>3 16.7</td>
<td>2 7.7</td>
<td>5 31.3</td>
<td>0 0</td>
<td>***</td>
<td>0.0003</td>
</tr>
<tr>
<td>lpfE</td>
<td>11 61.1</td>
<td>12 46.2</td>
<td>8 50.0</td>
<td>14 26.9</td>
<td>*</td>
<td>0.0422</td>
</tr>
<tr>
<td>yfcO</td>
<td>3 16.7</td>
<td>9 34.6</td>
<td>4 25.0</td>
<td>1 1.9</td>
<td>***</td>
<td>0.0003</td>
</tr>
</tbody>
</table>
Those with the greatest percentage of positive isolates shaded in blue and equally high percentage in grey. Those bolded have been associated with adherence in previous studies (TABLE 5).

\textsuperscript{b} Fisher's Exact test

In summary, 17 adhesins, 14 AT adhesins, 74 fimbrial protein genes, and 14 from the Fim operon and other broad fimbrial gene designations were examined for associations with symptomatic status and Tir Nodes. Individual gene associations were complex and hard to interpret when the contribution of each individual adhesin to adherence is unknown. Also, more than one adhesin is likely expressed and functional at one time, thus proficient adherence of aEPEC is likely the combined effort of multiple adhesins as has been previously described for tEPEC, EHEC, and aEPEC. Therefore, I asked if the number of adhesins were more abundant in isolates based on symptomatic status or by Tir Node. The number of total adhesins and individual categories of adhesins were similar based on symptomatic status (Figure 40A) in contrast to significant differences in Tir Nodes (Figure 40B). The total number of adhesins was higher in Tir Node 2 (77.7 ± 1.60 SEM) compared to Tir Node 4 (66.7 ± 1.73 SEM) and Tir Node 3 and Tir Node 1 had intermediate numbers (73.8 ± 2.73 SEM and 71.1 ± 2.85 SEM) (Figure 40B). These differences are likely due to the differences observed in different categories of adhesins; fimbrial protein genes were more abundant in Tir Node 2 (37.0 ± 1.52 SEM) compared to Tir Node 4 (24.5 ± 1.48 SEM) and AT adhesins were more abundant in Tir Node 2 (7.96 ± 0.74 SEM) compared to both Tir Node 3/4 (5.50 ± 0.34 SEM; 5.19 ± 0.40 SEM) (Figure 40B) indicating Tir Node 2 isolates likely have the greatest adherence potential followed by those in Tir Node 2/3 with Tir Node 4 isolates having the least potential.
Figure 40: Number of Adhesin Genes Present in aEPEC Isolates. (A-B) Total number of adhesins and number in specific categories as indicated and grouped according to (A) symptomatic status and (B) Tir Nodes 1–4. 1-way ANOVA. ns: no significance, *P < 0.05, ***P < 0.001, ****P < 0.0001.

Having identified the number and types of adhesins present, I asked whether the number of all or individual categories of adhesins were correlated with clinical manifestation of aEPEC infection. I examined continuous variables of clinical manifestations such as diarrhea duration, number (#) of bowel movements (BM) in a 24hr period, and diarrhea severity assessed with the modified Vesikari score (mVS) (Figure 41). Only the number of BMs/24hrs had a weakly positive correlation to the number of AT adhesins present (Figure 41).
Figure 41: Correlation Matrix of Adhesins Present Versus Clinical Manifestation of aEPEC Infection. Pearson correlation R value displayed for each combination. * $P = 0.03483$.

Overall, the adhesin *paa* positively associates with SYMP isolates. In addition, significant differences in the prevalence of adhesins and fimbrial protein genes in different Tir Nodes are present with Tir Node 1 isolates possessing the most adhesin genes and Tir Node 3 isolates possessing the most fimbrial protein genes. The number of AT adhesins weakly correlate with # BMS/24hr. These data indicate that known adhesins do not predict symptomatic status or symptoms or severity of diarrhea caused by these aEPEC isolates, however, they likely contribute and may give a competitive advantage during infection to those that are in specific Tir Nodes.

**Pangenome and Accessory Gene Analysis**

*E. coli* are a diverse set of organisms with great genetic plasticity often acquiring genes, virulence factors, and mobile genetic elements through horizontal gene transfer with bacteria it encounters from varied niches ranging from the environment to the intestinal tract of warm-blooded animals and humans. Due to the changing genomes of *E. coli*, virulence factor lists are likely to be incomplete and ambiguities exist as some pathogenic *E. coli* pathotypes share similar virulence factors. Therefore, a comprehensive list of virulence factors is hard to ascertain. To take an unbiased approach and examine differences by symptomatic status, a
pangenome approach was undertaken using the PGAP annotations of the 112 aEPEC isolates and the PanExplorer pipeline.223

Analysis of the pangenome of the 112 aEPEC isolates revealed that genes in the core genome represented only 9.2% (n=1693) (Figure 42A). The majority of genes were accessory genes (54.3%, n=10,032) representing those found in 2-111 genomes (Figure 42A) with SYMP isolates harboring a significantly higher number of accessory genes (2684 ± 23 SEM) compared to OB isolates (2599 ± 31 SEM) (Figure 42B). There was also a high percentage of genes unique to only 1 isolate (36.5%, n=16,572). Again, SYMP isolates (74.3 ± 8.2 SEM) harbored significantly higher number of unique genes compared to OB isolates (30.9 ± 11.1 SEM) (Figure 42C).

**Figure 42: Pangenome of aEPEC Isolates.** (A) Distribution of core, strain-specific, and accessory genes found in 112 aEPEC isolates. Number of (B) accessory genes and (C) unique genes present in each of the defined groups, OB, ASYM, and SYMP. (B) Student’s t-test and (C) 1-way ANOVA; *P < 0.05, **P < 0.01.

To further explore the pangenome of the 112 aEPEC isolates, presence or absence of accessory genes in each isolate was examined and compared by symptomatic status (Figure 43A). Of the 16,072 accessory genes, 233 were significantly associated with OB isolates (Figure 43A-blue dots) and 230 were associated with SYMP isolates (Figure 43A-red dots). To understand the function of the accessory proteins most associated with SYMP or OB isolates,
cluster of orthologous genes (COG) function designations were examined for each accessory gene (Figure 43B). The majority of assigned COG functions had greater numbers in OB isolates, except for those assigned to the COG function “chromatin structure and dynamics” (Figure 43B). Interestingly, SYMP isolates contained the most accessory genes that could not be assigned a COG function indicating the unknown nature of additional genes recognized by the pangenome analysis (Figure 43B). Further in-depth characterization of accessory and unique genes specific to SYMP and OB isolates is warranted to understand what roles metabolism, cellular processing and signaling, and information storage and processing may play in the pathogenic mechanisms of aEPEC infection.

Figure 43: Accessory Genes of the Pangenome of aEPEC Isolates. (A) Volcano plot displaying those that are significantly associated with OB isolates (blue dots) and those associated with SYMP isolates (red dots). The (-log) transformation of the p-value from individual Chi-square tests was plotted on the y-axis versus the percent change in the number of SYMP isolates minus OB isolates on the x-axis. (B) Number of annotations assigned to each COG function of the most significantly different accessory genes present in SYMP and OB isolates.
Summary

Genetic analysis revealed that the 112 aEPEC isolates originate from diverse phylogroups, serogroups, and MLST designations. The 1000-core genome of the aEPEC isolates clustered broadly by phylogroup and more specifically by serogroup and MLST but not based on symptomatic status, temporal acquisition, or by geographic location indicating the isolates originate from independent sporadic infections over time. Examination of virulent genetic factors focused on proteins encoded on the LEE pathogenicity island, non-LEE effectors, and adhesins. Importantly, I identified 4 Tir Nodes which are associated with symptomatic status. LEE protein homology patterns emerged when isolates were clustered according to Tir phylogeny. Contrary to my original hypothesis, SYMP aEPEC isolates overall had greater diversity in protein homology to the LEE of tEPEC/EHEC compared to OB isolates with significant differences observed in Tir, the translocator, EspB and the effectors EspH and EspF. In contrast, OB isolates overall were more homologous to the LEE of tEPEC/EHEC, however, significant diversity was observed in the needle protein, EspA.

LEE subtyping based on intimin, Tir, EspA,B,D, EspH, and EspZ\textsuperscript{176} revealed nearly 40% of our aEPEC isolates from the US have novel LEE subtypes compared to aEPEC isolates from Asia and Africa; the most novel isolates belong to Tir Nodes 1-3. The non-LEE effector repertoire was diverse but overall, isolates in Tir Node 2 had the greatest number, Tir Nodes 1 and 3 had intermediate levels, and Tir Node 4 had the least, especially in those aEPEC isolates designated as Tir subtype 10 of Tir Node 4. The same pattern was also observed for adhesin and fimbrial proteins. Two adhesins previously shown to have a greater association with aEPEC isolates from symptomatic individuals were also found in my study to be associated with
symptomatic status, LifA (efa1/lifA)\textsuperscript{227} which had the greatest abundance in Tir Node 2, and paa\textsuperscript{191, 228, 229} which was equally distributed throughout the Tir Nodes.

Through investigation into whether genetic virulence factors correlate with the clinical manifestations of aEPEC infection, I determined that Tir Node 2 isolates originate from stools with higher relative EPEC loads, and Tir Node 1 isolates originate from individuals who had more instances of emesis. In addition, weak but positive correlations between diarrhea severity and tEPEC non-LEE effectors, and the # BMs/24hrs and AT adhesins existed. Together these data suggest that Tir Nodes and the corresponding LEE subtypes, and the presence of certain non-LEE effectors and adhesins contribute to disease manifestations of aEPEC infection. One factor does not dictate symptoms or severity of aEPEC infection. Rather it is likely a combination of factors specific to individual isolates or groups of isolates that results in a strain being more or less virulent. Further analysis is needed to understand the collective effects of these known genetic virulence factors as well as the importance of the vast accessory and unique gene profiles of SYMP isolates detected in the pangenome analysis.

**Limitations**

One major limitation in determining the genetic virulence factors most likely associated with symptomatic status is the quality of the whole genome sequence, assembly, and annotations. I circumvented this limitation by ensuring high quality DNA went into the sequencing reaction and by comparing the assembly and annotations provided by 3 different pipelines, BV-BRC, Enterobase and PGAP. I found that BV-BRC pipeline which uses Unicycler (an optimizer for SPADES assembly), gave the best assemblies with the lowest number of contigs, longest total genome lengths, and greatest N50 values. The BV-BRC assembled contigs were then submitted to PGAP, which gave the most complete and up to date annotations as determined by comparing
known LEE and non-LEE gene annotations. Therefore, with the high-quality assemblies and annotations of our 112 aEPEC isolates I was able to perform a rigorous genetic analysis of known virulence factors.

Another limitation of the genetic analysis is the overall abundance in genetic diversity of our aEPEC isolates and *E. coli* in general. It is difficult to ascertain virulence traits of symptomatic isolates when each isolate is so unique and the possibility of acquiring one or more genes that could change a strain from avirulent to highly virulent is high in *E. coli* due to its promiscuity in acquiring genetic factors through horizontal gene transfer. In addition, it is likely that a combination of genetic virulence factors cause a strain to be more or less virulent. Higher order statistical analyses and machine learning approaches would be well suited to aid in understanding the contributory or the layers of genetic factors that affect disease manifestation of aEPEC infection. These approaches, however, are beyond the scope of this project.

Another limitation was the number of isolates I had for each symptomatic status group. Often significance could not be assigned to the ASYM group because of its low *n* of 13 isolates. Although we were fortunate to have acquired these, a prospective study aimed at acquiring greater numbers of isolates from asymptomatic and symptomatic individuals would help in understanding differences due to the genetic virulence potential of the aEPEC isolate versus the role that host factors contribute to disease manifestations of aEPEC infection.

Lastly, this study was not an exhaustive examination of all genetic virulence factors which may contribute to the pathogenicity of aEPEC. Further analysis is needed to explore the presence/and absence of toxins, biofilm and regulatory genes, other secretion systems, and genes associated with metabolism, niche fitness, survivability, and defense mechanisms.
CHAPTER 6
CHARACTERIZATION OF THE ABILITY OF ATYPICAL EPEC ISOLATES TO CAUSE IN VITRO VIRULENCE PHENOTYPES

Introduction

tEPEC pathogenesis depends on attachment via BFP that allows for microcolony formation and localized adherence (LA).\textsuperscript{141} In contrast, aEPEC, which lack BFP, display several different adherence patterns described as localized adherence-like (LAL), diffuse (DA), or aggregative (AA).\textsuperscript{3, 9} In addition, tEPEC pathogenesis depends on proteins encoded on the LEE pathogenicity island, such as EspA and intimin which are important for tEPEC initial and intimate attachment, respectively.\textsuperscript{141} The LEE also encodes the T3SS which injects effector proteins in host cells to induce actin pedestals, attaching/effacing lesions (A/E)\textsuperscript{154} and physiological changes in intestinal epithelial cells that contribute to diarrhea.\textsuperscript{6} For example, tEPEC effectors interfere with ion transporters/exchangers by decreasing SGLT-1 activity\textsuperscript{164} and inhibiting the Cl\textsuperscript{-}/HCO\textsubscript{3}\textsuperscript{-} exchanger, DRA.\textsuperscript{163} tEPEC also disrupt intestinal epithelial polarity and TJ structure and barrier function perturbing paracellular permeability.\textsuperscript{162, 165-168, 170}

The first step of tEPEC pathogenesis is attachment to host intestinal epithelial cells via BFP but tEPEC also depend on BFP for other downstream pathogenic mechanisms. Tir and EspF translocation are delayed when tEPEC lack BfpA and BfpF causing attenuated TJ disruption.\textsuperscript{143} This demonstrates that adherence directly influences delivery of effector proteins that cause disease-associated phenotypes. The contribution of aEPEC attachment mechanisms to virulence is unclear since BFP is lacking. While BFP, EspA, and intimin are the key tEPEC adhesins, I
speculated that aEPEC possess other adhesins or virulence factors that compensate for the lack of BFP to initiate attachment and downstream effects that lead to diarrhea. Therefore, detection of genetic virulence factors, as performed in CHAPTER 5, needs to be coupled to phenotypic virulence factor analyses to understand aEPEC pathogenesis. I hypothesized that variation in aEPEC attachment influences delivery of T3SS effectors that induce pedestal formation and other downstream effects leading to diarrhea. Similar to previous data, I determined in CHAPTER 5 that the adhesin genes, *lifA/efa1* and *paa*, were positively correlated with diarrhea. However, little information is available regarding the role of these adhesins in attachment and virulence. The objectives of this aim were to characterize the in vitro virulence phenotypes of 112 aEPEC isolates and correlate these phenotypes with genetic virulence factors and clinical manifestation of aEPEC infection (Figure 44). I determined adherence pattern and attachment levels of aEPEC isolates to the human colonic cell line, SKCO-15, to examine initial attachment mechanisms. In addition, pedestal formation was assessed and quantitated to examine T3SS functionality and intimate attachment of bacteria to host cells. In vitro phenotypes were correlated with the presence of specific adhesin genes, *lifA* and *paa*, as well as other adhesins and fimbrial protein genes, non-LEE effectors, Tir and EspA protein homology, and clinical manifestations of aEPEC infection. I speculated that aEPEC isolates displaying defined adherence (LA, LAL, AA, or DA) versus undefined adherence (UND) would induce more robust disease-associated virulence phenotypes as delivery of effectors into host cells would be enhanced. By coupling genetic and phylogenetic analyses to disease manifestations we will begin to understand aEPEC pathogenesis.
Figure 44: Schematic of Methods and Objectives to Determine \textit{In Vitro} Virulence Phenotypes of aEPEC Isolates.

Materials and Methods

Bacterial Culture and Infection of Host Intestinal Epithelial Cells

Bacterial cultures were grown as previously described\textsuperscript{13, 162, 170, 172} to activate the LEE operon and T3SS. Infection controls included tEPEC-E2348/69 (positive), HB101 (negative), and tEPEC-E2348/69ΔbfpA (positive after extended infection times). Primary cultures of clinical EPEC isolates were prepared in 5 mL of LB broth and grown at 37°C with shaking at 250 rpm for ~18 h. Then 133.3 μL of primary culture was inoculated into 4 mL of bacterial infection media consisting of serum- and antibiotic-free 1:1 (v/v) mixture of DMEM:Ham’s F12 (Gibco, 21700075) containing 0.5% mannose (Sigma, M6020) and grown to mid-log phase to reach ~5x10^8 CFU/mL as previously described.\textsuperscript{3, 162, 166, 170, 172, 230} Bacteria were centrifuged at 1000 rcf for 10 min, the pellet resuspended in 5 mL fresh 37°C bacterial infection media and added to epithelial cells at a multiplicity of infection (MOI) of ~50:1 for 2.5 - 5hrs as indicated.

Epithelial Cell Culture

Human SKCO-15 colonic epithelial cells were used between passages 30–45 and grown in a 5% CO2 incubator at 37°C in Dulbecco’s Modified Eagle Media (DMEM) (Gibco, 31600034) supplemented with 20.8 mM HEPES (US Biological, H2010), 19.4 mM D-dextrose
(Fisher, D16-500), 10% FBS (Gibco, 16140071), and 1% penicillin/streptomycin (Gibco, 15140163). Eighteen hours before bacterial infections, cell culture medium was changed to bacterial infection media. Monolayers were plated in 24-well culture plates on coverslips to 80-90% confluency for Giemsa and actin staining and in 12-well plates to 90-100% confluency for adherence assays.

**Determination of Adherence Patterns**

OB, ASYM, and SYMP aEPEC isolates were characterized for adherence patterns. After infection, monolayers were washed 3x in 1xPBS, fixed in 100% cold methanol for 5 minutes, stained with Giemsa (Sigma, 1092040500) at 1:20 in 1x PBS for 20 minutes at room temperature, washed, and coverslips affixed to slides using Permount mounting medium (Fisher, SP15-100). Images were acquired using phase-contrast light microscopy on Leica DMI4000 (Metamorph software) and processed using Adobe Photoshop. Adherence patterns were characterized as LA, LAL, AA, DA or UND.

**Quantitative Adherence Assay**

Adherence was quantitated as previously described.3 After infection, monolayers were washed, incubated for 5-10 minutes at 37°C in 200 uL of 1 % Triton X-100 in 1x PBS, and cells scraped into 300uL of 1xPBS (total volume 500 uL). Dislodged cells and bacteria were vigorously vortexed for 10-30 seconds and serially diluted in 1xPBS, plated on LB agar, incubated overnight at 37°C, and bacterial colonies enumerated. Input bacteria at T = 0 were also serially diluted and enumerated. Percent adherent bacteria equals CFUs of attached bacteria after infection divided by CFUs of input bacteria. Adherence levels of aEPEC isolates were compared across all infection times by dividing percent adherent bacteria by the average percent of
adherent ΔbfpA resulting in fold adherence above ΔbfpA. Adherence assays were performed in duplicate with at least 3 biological replicates.

**Pedestal Formation**

Functionality of the T3SS was assessed by examination of pedestal formation by aEPEC isolates as previously described. After infection and washing, monolayers were fixed in 3.7% paraformaldehyde (PFA) for 20 minutes, washed 2x in 1xPBS, permeabilized in 0.1% Triton X-100 for 15 minutes, blocked for 2 hours to overnight in blocking solution (Invitrogen, 000–105) and incubated with 1 U of BODIPY-558/568 Phalloidin (Invitrogen, B3475) to detect F-actin for 2-5 hours at room temperature in humidified chamber protected from light. Nuclei were stained with Hoechst 33342 (Invitrogen, H3570) and coverslips mounted using ProLong Gold Antifade reagent (Invitrogen, P36934). Images were acquired using either a Leica DMI4000 (MetaMorph software) fluorescence microscope, or a Leica TCS SPE DMI 4000B (LAS X software) confocal microscope and images processed using ImageJ and Adobe Photoshop. Pedestal formation was assessed as a binary yes or no answer as well as quantified. To quantify pedestals, 3-15 field of views were analyzed from 2-5 biological replicates. ImageJ was used to quantify host cells per field of view by automatic nuclei counting and pedestals were manually counted using the counting tool in ImageJ. Pedestal quantification was reported as the number of pedestals divided by the number of host cells.

**Statistical Analyses**

Statistical analysis and figure creation were performed with GraphPad Prism 10 (GraphPad Software, San Diego, CA) and Illustrator 2021 (Adobe Inc, San Jose, CA), respectively. Statistical analyses used are indicated in figure legends. Data are presented where
applicable as mean ± standard error of mean. Statistical significance was defined as $P < .05$, and “ns” indicates non-significant.

**Results**

**Adherence Patterns of aEPEC Isolates**

One of the earliest distinguishing factors of tEPEC was its ability to form dense microcolony clusters of bacteria on epithelial cells termed localized adherence (LA).\(^{141}\) We previously examined the adherence pattern for aEPEC ($n=56$) isolates and determined that the adherence patterns were diverse ranging from an UND pattern to more defined patterns such as diffuse (DA), aggregative (AA), or localized-adherence-like (LAL).\(^3\) However, adherence patterns were not previously analyzed for their association with symptomatic status or the genetic virulence factors present. Therefore, adherence pattern and adherence levels were examined in an additional fifty-six ($25$ ASYM or SYMP $n=25$ and OB $n=31$) aEPEC isolates. I questioned if associations were present between adherence pattern or level and symptomatic status, Tir Nodes, or the number of non-LEE effectors or adhesins/fimbrial protein genes present. Colonic-like intestinal epithelial cells (SKCO-15) were infected and adherence patterns and levels examined. Five aEPEC isolates lifted cells at the extended infection time of 5 hrs (CE234, CE245, CE254, CE285, OB137). Therefore, adherence patterns and levels of these isolates were assessed at 2.5 hrs post-infection. The strain tEPEC_E2348/69 lacking $bfpA$ ($\Delta bfpA$) was used as a control at both timepoints to assess baseline adherence of a strain which lacks BFP. Adherence of all other isolates was compared to $\Delta bfpA$ allowing for comparison of adherence between isolates across the different infection timepoints.

The adherence patterns of the 112 aEPEC isolates were diverse with LAL, DA, AA, and UND patterns detected; LA pattern was not observed for any aEPEC isolate (Figure 45A). An
UND pattern was displayed by the majority of OB isolates (80%) (Figure 45B) with only a few having defined patterns (n=6/31, 20%) (Figure 45C). In contrast, the majority of SYMP isolates displayed a defined pattern (77.9%) (Figure 45C) with DA (42.6%) pattern predominating followed by LAL (22.1%) and AA (13.2%) (Figure 45B). ASYM isolates had nearly equal percentages of defined versus undefined patterns (53.8% defined) (Figure 45B and C).

Adherence patterns were also examined by Tir Node. Defined patterns had the greatest association to Tir Nodes 1-3 (77.8%, 73.1%, and 68.8%, respectively) (Figure 45D) with DA (50%) and AA (16.7%) patterns the most abundant defined patterns in Tir Node 1 and DA (38.5%) and LAL (23.1%) most abundant defined patterns in Tir Node 2 (Figure 45E). Tir Node 4 had the fewest number with a defined pattern (43.1%) (Figure 45D) and UND pattern was most prevalent (56.9%) (Figure 45E). These data indicate that a defined adherence pattern is associated with symptomatic status of aEPEC isolates and Tir Node 4 is mostly associated with UND pattern.
**Figure 45: Adherence Patterns of aEPEC Isolates.** (A) Representative images of Giemsa stained SKCO-15 intestinal epithelial cells displaying adherence patterns of aEPEC isolates: LAL – localized-adherence-like, DA – diffuse, AA – aggregative, UND – undefined. Percent of isolates (B) displaying each adherence pattern and (C) those with defined adherence grouped according to symptomatic status. Percent of isolates (D) with a defined pattern and (E) displaying each adherence pattern grouped according to Tir Nodes. (B-E) Fisher’s exact tests and Chi-square test with *P values* as indicated.

**Adherence Levels of aEPEC Isolates**

Adherence level was also assessed and compared by symptomatic status and Tir Node.

SYMP isolates (1.766 ± 0.131 SEM) had significantly higher levels of attachment compared to OB isolates (0.999 ± 0.125 SEM) with ASYM isolates at an intermediate level (1.267 ± 0.205 SEM) (Figure 46A). Adherence was also significantly higher in Tir Node 1 isolates (2.071± 0.286 SEM) compared to those in Tir Node 4 (1.206 ± 0.105 SEM) (Figure 46B). Tir Node 2 (1.803 ± 0.219 SEM) and Node 3 (1.290 ± 0.243 SEM) isolates had intermediate levels (Figure 46B).

**Figure 46: Fold Adherence Level of aEPEC isolates.** (A-B) Fold adherence above ΔbfpA (set to 1 as indicated by black dashed line) grouped according to (A) symptomatic status and (B) Tir Nodes. Adherence level 2-fold higher than ΔbfpA is indicated by red dashed line. Adherence of isolates at or above dashed red line are significantly higher than ΔbfpA. Infection times as indicated: 2.5 hrs (orange squares) and 5 hrs (black triangles or circles). 1-way ANOVA ** *P < 0.01.*
Pedestal Formation by aEPEC Isolates

Intimate attachment of tEPEC is facilitated by intimin-Tir interactions. In addition, Tir is an effector that recruits actin-rich clusters under attached bacteria termed pedestals. Pedestal formation, therefore, is indicative of functional T3SS effector delivery as Tir alters actin arrangement in the cell. Therefore, I questioned if aEPEC isolates from SYMP individuals or from different Tir Nodes would have a greater propensity to form pedestals. Wild-type tEPEC-E2348/69 causes robust pedestal formation in tightly packed microcolonies (Figure 47A). Pedestals were evident after infection with ΔbfpA, however, they were dispersed and often had long actin tails (Figure 47A). Similarly, aEPEC isolates formed dispersed pedestals sometimes with actin tails (Figure 47Aii and 47Aiii, respectively). Attached bacteria without pedestals were also observed (Figure 47Ai). Regardless of the type of pedestals formed, more than half of all isolates (62.5%) formed pedestals. Although a greater percentage of isolates from the SYMP group formed pedestals (75%), the percentage was not significantly higher than those from the OB or ASYM groups (Figure 47B). Quantitation of pedestal formation was performed and the number of pedestals per host cell was analyzed (Figure 47C). SYMP isolates formed significantly more pedestals/host cell (0.652 ± 0.149 SEM) than OB isolates (0.049 ± 0.012 SEM) (Figure 47C). I also found that Tir Node 2 isolates formed significantly greater numbers of pedestals (0.633 ± 0.226 SEM) than Tir Node 4 isolates (0.322 ± 0.135 SEM) (Figure 48A). Surprisingly, even though Tir Node 1 isolates formed the greatest number of pedestals/host cell (0.723 ± 0.254 SEM) similar to Tir Node 2 isolates, and Tir Node 3 had the lowest numbers detected (0.246 ± 0.106 SEM), neither Tir Node 1 or 3 was significantly different than the other Nodes (Figure 48A). A wide range in values for the number of pedestals/host cell existed in all
Tir Nodes; this variability and smaller $n$ likely accounts for the lack of significance between Tir Node 1 and 3 and the others.

**Figure 47: Pedestal Formation by aEPEC Isolates.** (A) Representative images of SKCO-15 monolayers infected with aEPEC isolates and stained for F-actin with BODIPY-Phalloidin (red); bacterial and host cell nuclei were stained with Hoescht (blue). (i) Some isolates do not form pedestals although attached bacteria are apparent (white arrow), (ii) while others form robust pedestals, and (iii) some form elongated pedestals similar to ΔbfpA. Scale bar = 10 µm. Images adapted from: Carlino, MJ*, Kralicek SE*, Santiago SA, et. al. Gut Microbes. 2020 Nov 9;12(1):1-21. (B) Percent of isolates that form pedestals grouped according to symptomatic status with the number ($n$) of positive isolates indicated. Chi-square test for trend, $P = 0.0808$. (C) The number of pedestals/host cell was calculated and compared between OB, ASYM, and SYMP isolates. Kruskal-Wallis with Dunn’s multiple comparisons test, ** $P < 0.01$. 
Figure 48: Pedestal Formation of aEPEC Isolates in Different Tir Nodes. (A) The number of pedestals/host cell was calculated and compared between the different Tir Nodes. Kruskal-Wallis with Dunn’s multiple comparisons test, * $P < 0.05$. (B) Percent of isolates that form pedestals grouped according to Tir subtype with Tir Node and symptomatic status indicated. Chi-square test, *** $P = 0.0004$.

To examine pedestal formation among the Tir Nodes in more detail, the ability to form pedestals was examined per Tir subtype (Figure 48B). When examining Tir Node 1, subtypes 1 and 2 had significantly higher percentage of isolates that form pedestals compared to most other subtypes (83.3% and 100%, respectively) (Figure 48B). In contrast, only 22% (n=2/9) of Tir subtype 5_Node 1 isolates were able to form pedestals (Figure 48B). This difference likely accounts for the lack of significance detected in the pedestal quantitation in Figure 48A. Similar to the pedestal quantification, both Tir Node 2 subtypes, 3 and 4 had significantly higher percentage of isolates which formed pedestals (83.3% and 92.9%, respectively) (Figure 48B). SYMP isolates made up the majority of isolates in all Tir Node 1 and 2 subtypes that were able to form pedestals (81.8%) (Figure 48B-dashed bars). Tir subtypes 6-11 of Tir Nodes 3/4 had significantly fewer isolates that form pedestals and the percentages were more equally distributed between OB/ASYM and SYMP isolates with no pedestal formation detected in Tir subtype 9_Node 4 (Figure 48B). These data indicate that differences in Tir, especially when examined at the level of Tir subtypes, are associated with pedestal formation and with symptomatic status.

Correlation of In Vitro Virulence Phenotypes to Genetic Factors and Clinical Manifestations

My original hypothesis was that increased homology to tEPEC/EHEC within LEE proteins and greater detection of non-LEE and adhesin proteins, would result in greater adherence allowing for enhanced delivery of effectors and downstream effects such as pedestal formation and clinical manifestations. I determined that the adhesins lifA and paa were
associated with symptomatic status (CHAPTER 5). Therefore, I next asked if they were associated with *in vitro* virulence phenotypes, especially those related to adherence. A significantly higher percentage of isolates which had *lifA* _IE6_ (integrative element 6) were associated with DA adherence pattern and those which had *lifA* _IE2_ were significantly associated with AA adherence (Figure 49A). In contrast in *paa* isolates, AA, LAL and UND patterns were equally evident (Figure 49A). Similarly, there was a trend towards higher adherence of isolates which contained *lifA* from IE2 or IE6 (median 1.490 IQR 1.053 – 2.093) compared to those without either of these genes (median 1.045 IQR 0.6775 – 1.713) (Figure 49B). This trend was not detected in those which contained *paa* (median 1.335 IQR 0.892 – 2.313) compared to those which did not (median 1.067 IQR 0.682 – 1.810) (Figure 49C). Also, the presence of *lifA* _IE2/6_ tended to be associated with pedestal formation (76.47% compared to 56.41%) but not significantly (Figure 49D). These data indicate that detection of *lifA* from IE2 and IE6 contribute to adherence properties and pedestal formation, however other factors also play a role in *in vitro* virulence phenotypes.
Figure 49: Adhesins, LifA and Paa, Are Weakly Associated with In Vitro Virulence Phenotypes. (A) Percent of isolates with the adhesins, lifA_IE6, lifA_IE2, or paa, grouped according to adherence pattern. Fisher’s exact tests with significance as indicated. (B-C) Fold adherence compared to ΔbfpA grouped according to those with (B) lifA_IE2 or lifA_IE6 and (C) paa present compared to all other isolates which do not have the specified gene. Mann-Whitney tests with median (solid line) and interquartile range (dotted lines) and significance indicated. (D) Percent of isolates positive for pedestal formation compared between those that have lifA_IE2/6 and those that do not. Fisher’s exact test with significance as indicated. Number (n) positive for pedestal formation.

I then asked if correlations existed between adherence level or pedestal formation and protein homology of LEE virulence factors or clinical manifestations. A correlation matrix was created with individual Pearson R correlation values for individual tests performed between each pair of variables, i.e. adherence level versus average pedestal formation, or other non-LEE effectors, etc., with significance indicated for each correlative test (Figure 50). There was a weak, but significant, positive correlation between adherence and the number of non-LEE effectors present from strains other than tEPEC. Similarly, the number of non-LEE effectors from tEPEC and other strains combined was also weakly correlated with adherence level. In
addition, there was a moderate, positive correlation between adherence and the number of fimbrial protein genes (Figure 50). Together, these data indicate that as the number of non-LEE effectors and fimbrial protein genes increased, adherence level also increased. In addition, as adherence level increased, the number of pedestals/host cell also increased indicated by the moderate, positive correlation (Figure 50). Both adherence level and average pedestals/host cell had a moderate, positive correlation to mVS indicating that increased \textit{in vitro} virulence phenotypes correlate with diarrhea severity (Figure 50). Interestingly, tEPEC/EHEC Tir homology decreased as adherence level and the number of pedestals/host cell increased indicated by the moderate and weak, respectively, negative correlations between Tir homology and these two virulent phenotypic measures (Figure 50). In contrast, EspA homology did not correlate with adherence level or the number of pedestals/host (Figure 50).

\textbf{Figure 50: Correlation of \textit{In Vitro} Virulence Phenotypes with Genetic Factors and Clinical Manifestations of aEPEC Infection.} Pearson correlation R value displayed for each combination. Correlative tests were examined between adherence level (Adh level) or number of pedestals/host cell (Pedestal avg) versus percent identity of Tir or EspA to tEPEC/EHEC, non-LEE effectors (tEPEC non-LEE effectors, non-LEE effectors found in other pathovars (other), or combined), adhesin groups (adhesins, fimbrial protein genes, Fim operon genes, AT adhesins or combined), and different clinical manifestations including mVS (modified Vesikari score). Significant values: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, and **** $P < 0.0001$. ^non-significant but trending towards it: $^\wedge P = 0.065$. 
Summary

It was previously unknown if in vitro phenotypes related to adherence and pedestal formation correlated to clinical symptomology of aEPEC infection. I determined that SYMP isolates were more likely to display defined patterns of adherence, greater attachment levels, and greater ability to form pedestals. In addition, Tir Nodes 1 and 2, which contain the highest percentage of SYMP isolates, were also the Tir Nodes with the greatest propensity to cause in vitro virulence phenotypes. Although more than half of all aEPEC isolates examined, including OB isolates, form pedestals and are “primed” to be virulent, the propensity to have greater host cell attachment based on higher adherence level and defined adherence pattern likely contributes to the number of pedestals formed. For example, those with the adhesin LifA tended to display defined adherence patterns, increased adherence levels, and greater pedestal formation. However, not all isolates with LifA had the same in vitro phenotypes and significance was not always achieved when compared to isolates without LifA. These data suggest that LifA individually does not predict virulence but rather a combination of highly specific virulence factors at the strain level is required. This is supported by the weaker correlation values for each individual test associating in vitro phenotypes with adhesins, LEE protein homology and clinical manifestations indicating a contribution from each factor, but no single factor being the sole determinant of virulence. Interestingly, 67% (n=6/9) of isolates from Tir subtype 5 of Tir Node 1 were from SYMP individuals but nearly 80% did not form pedestals suggesting a unique pathogenic mechanism. In conclusion, these data suggest that aEPEC pathogenesis is multifactorial with strain-specific variability dictating the level of aEPEC virulence. My data suggest that variation in LEE proteins, especially Tir, and increased numbers of virulence genes related to infection
(non-LEE effectors) and adherence contribute to aEPEC virulence. However, there are likely more virulence factor determinants to be discovered.

**Limitations**

One limitation to my study was that all *in vitro* experiments were performed in the colonic cancer cell line SKCO-15. Although we previously determined that EPEC isolates have a greater propensity to form pedestals on SKCO-15 versus HeLa cells, a cervical cancer cell line, it is possible that transformations have occurred within the cells which may make them more or less permissive for adherence and pedestal formation by EPEC. Future studies are needed to determine adherence to primary intestinal epithelial cells where multiple levels of interaction and feedback among intestinal cell types, and between host cells and bacteria occur. Human enteroids/colonoids would be ideal for these experiments as several host factors could be accounted for such as intestinal segment tropism, the propensity to infect one cell type versus another such as enterocytes, Paneth cells, goblet cells, and neuroendocrine cells, and host variability. Interestingly, one isolate, CE37 contained adhesins specifically associated with diffuse adherence (*afaA, daaA, daaF, drab, draD*) and two isolates, CE239 and CE435, contained the gene, *air*, associated with aggregative adherence (Figure 36), yet all three displayed undefined adherence patterns on SKCO-15 cells, reiterating the notion that intestinal tropism of these aEPEC isolates is likely a strong contributor to virulence and needs to be explored. Also, aEPEC isolates were grown in infection media containing 0.5% mannose as the established epithelial cell culture infection model for tEPEC. Mannose blocks attachment to host intestinal epithelial cells via Type 1 fimbriae. However, aEPEC, which lack BFP, may require Type 1 fimbriae for efficient attachment as has been determined for other aEPEC isolates. Therefore, future studies should include exploration of adherence in the absence of mannose.
Another limitation of this study is the number of aEPEC isolates in each group that was examined. The smaller $n$ in Tir Nodes 1-3 compared to Tir Node 4 as well as the smaller $n$ in ASYM isolates caused trends to be observed but significant differences may have been missed. Greater $n$ would also allow for more advanced statistical and machine learning techniques to be employed. Machine learning techniques would allow for detection of specific combinations of LEE and non-LEE effectors, adhesins, and fimbrial genes which have a greater propensity to cause in vitro virulence phenotypes and disease manifestations.

Lastly, how changes in protein expression or function which may alter virulence mechanisms were not explored. Nor did I analyze other known downstream events which lead to diarrhea such as ion transporters/exchangers and others. These analyses were beyond the scope of this project but will require further study as they likely contribute to the multifactorial nature of aEPEC pathogenesis.
CHAPTER 7

PRINCIPAL COMPONENT ANALYSIS OF GENETIC AND PHENOTYPIC VIRULENCE FACTORS AND CLINICAL MANIFESTATIONS OF ATYPICAL EPEC INFECTION

Introduction

Previous studies of aEPEC infection have focused on associations of genetic virulence factors with certain strains of aEPEC mostly originating from countries other than the US. Some studies have assessed the ability of certain aEPEC strains to cause in vitro virulent phenotypes such as adherence pattern and pedestal formation. However, very little is known about the associations between these virulent genetic factors and in vitro phenotypes and clinical manifestations of aEPEC infection. Therefore, it is largely unknown if virulent pathogenic factors associated with tEPEC are also important for aEPEC virulence.

I determined that less homology to reference strains in the LEE components, Tir, EspH, EspB, and EspF, increased homology of EspA, and increased prevalence of certain non-LEE effectors and adhesin/fimbrial genes are associated with symptomatic status (CHAPTER 5). In addition, I determined that aEPEC isolates originating from symptomatic individuals had greater adherence properties and greater ability to form pedestals (CHAPTER 6). Together, my data suggest that aEPEC pathogenesis is multifactorial. Therefore, I hypothesized that the number and combination of virulence factors housed by aEPEC isolates determine symptoms, severity of diarrhea, and other disease manifestations caused by aEPEC infection. To begin to understand which virulence factors are important for aEPEC pathogenesis, I used a multivariate statistical approach. Due to the co-linearity of several of the virulence factors, for example, reduced
homology in many of the LEE proteins, I used principal component analysis (PCA) as a
descriptive exploratory analysis. The objective of this aim was to identify aEPEC isolates with
certain genetic and phenotypic explanatory variables that correlate with asymptomatic versus
symptomatic individuals with variable disease manifestations and severity.

**Methods**

Principal component analysis (PCA) was performed using GraphPad Prism 10 (GraphPad
Software, San Diego, CA). Continuous explanatory variables (as listed in Figure 51A) that were
determined to be associated with symptomatic status or Tir Nodes in single-factor analyses
(CHAPTEERS 5-6) were used as input for PCA along with continuous clinical outcome variables
determined in CHAPTER 3 (diarrhea duration, #BMs/24hr, and modified Vesikari score (mVS)).
PCA was employed using the Prism standardized method which scales data to have a mean of 0
and SD of 1. The method for selecting principal components (PCs) was based on parallel
analysis with eigenvalues greater than those from the simulations at the specified percentile level
of 95% with random seed value set to auto. Comparisons between different PCAs were made
and different explanatory variables selected to determine which PCA encompassed the fewest
variables while capturing the greatest percentage of cumulative total variance and by examining
which PCA score plots clustered OB, ASYM, and SYMP aEPEC isolates with the greatest
distinction.

**Results**

To begin to explore which genetic virulence factors had the greatest association with *in
vitro* virulence phenotypes and clinical manifestations of aEPEC infection, PCA was employed
using continuous genetic variables determined in CHAPTER 5 to be associated with
symptomatic status or Tir Nodes including homology of the translocon proteins, EspA and EspB,
and effectors Tir, EspH, and EspF, and the number of non-LEE effectors and adhesins/fimbrial protein genes, (Figure 51A). In addition, continuous \textit{in vitro} virulent phenotypes, such as adherence level (adh level) and the number of pedestals formed/host cell (pedestal avg), were included as variables for PCA analysis along with continuous clinical manifestations including mVS, diarrhea duration, and the number of BMs/24hr (Figure 51A). PC analysis A (PCA-A) indicated that the selected PC1-A had a proportion of variance of 30.95%, with PC2 and PC3 contributing 13.25% and 10.67%, respectively, for a total cumulative variance of 54.87% (Figure 51A). In PCA, each principal component (PC) is a linear combination of every single original variable and PCA loading values indicate the contribution of each variable to the PC with the highest absolute values contributing the most. Tir and EspH homology as well as the number of tEPEC non-LEE effectors, total non-LEE effectors, and fimbrial protein genes had absolute PC loading values (aPCLV) >0.7 indicating near equal contributions to PC1-A (Figure 51A). However, aPCLVs for Tir and EspH homology were positive values while the others were negative (Figure 51A); this was also evident by opposite directions of the vector loadings for these variables (Figure 51B). The negative values and opposite vector loadings indicate a negative correlation with these values, meaning that as Tir and EspH homology decrease, the number of non-LEE effectors and fimbrial proteins increase, as was also evident from my univariate analyses (CHAPTER 5). Moderate contributors to PC1-A were EspB, other non-LEE effectors, AT adhesins, and total adhesins with aPCLV between 0.6 and 0.7 (Figure 51A). EspA and EspF homology, adherence level, and diarrhea severity (mVS) were weaker contributors to PC1-A (aPCLV = 0.4 – 0.5) (Figure 51A and B). Strong and moderate contributors to PC2 were the number of other non-LEE effectors, Fim operon genes, and total adhesins (aPCLV = 0.5 – 0.8) while pedestal average, total non-LEE effectors, and EspH and Tir homology were weaker
(aPCLV = 0.3 – 0.5) (Figure 51A and B). The greatest contributors to PC2-A were the #BM/24hr, mVS, diarrhea duration and number of adhesins from highest to lowest aPCLV (Figure 51A and B). PC scores for each aEPEC isolate were plotted by PC1 versus PC2 and labeled by symptomatic status (Figure 51C). Distinct clusters of OB, ASYM, and SYMP isolates were not evident; quadrant 1 (Q1) had the greatest mix of aEPEC isolates (Figure 51C). In addition, many of the variables in PC2 contributed equally or were redundant to those in PC1 (Figure 51A). Therefore, these variables were selected for removal as indicated in Figure 51A and PC analysis B was performed.

**Table 1**: PCA Loading Values Analysis A

<table>
<thead>
<tr>
<th>Type</th>
<th>Variable</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>Analysis B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic</td>
<td>Tir</td>
<td>0.79</td>
<td>-0.455</td>
<td>-0.113</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EspA</td>
<td>-0.438</td>
<td>-0.110</td>
<td>-0.187</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EspB</td>
<td>0.896</td>
<td>-0.454</td>
<td>-0.121</td>
<td>remove</td>
</tr>
<tr>
<td></td>
<td>EspP</td>
<td>0.780</td>
<td>-0.451</td>
<td>-0.145</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EspF</td>
<td>0.408</td>
<td>-0.080</td>
<td>-0.133</td>
<td>remove</td>
</tr>
<tr>
<td></td>
<td>1EPEC non-LEE</td>
<td>-0.772</td>
<td>-0.250</td>
<td>0.032</td>
<td>remove</td>
</tr>
<tr>
<td></td>
<td>Other non-LEE</td>
<td>-0.645</td>
<td>-0.583</td>
<td>0.070</td>
<td>remove</td>
</tr>
<tr>
<td></td>
<td>Total non-LEE</td>
<td>-0.868</td>
<td>-0.458</td>
<td>0.056</td>
<td></td>
</tr>
<tr>
<td></td>
<td>adhesins</td>
<td>0.089</td>
<td>-0.168</td>
<td>-0.452</td>
<td>remove</td>
</tr>
<tr>
<td></td>
<td>fimbrial</td>
<td>-0.712</td>
<td>0.063</td>
<td>0.311</td>
<td></td>
</tr>
<tr>
<td></td>
<td>fim</td>
<td>0.143</td>
<td>0.718</td>
<td>-0.206</td>
<td>remove</td>
</tr>
<tr>
<td></td>
<td>AT adhesins</td>
<td>-0.593</td>
<td>-0.286</td>
<td>0.041</td>
<td>remove</td>
</tr>
<tr>
<td></td>
<td>Total adhesins</td>
<td>-0.667</td>
<td>-0.534</td>
<td>0.049</td>
<td></td>
</tr>
<tr>
<td>In vitro</td>
<td>Adh Level</td>
<td>-0.433</td>
<td>0.252</td>
<td>-0.215</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Peptide avg</td>
<td>-0.197</td>
<td>0.308</td>
<td>-0.364</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Diarrhea duration</td>
<td>-0.194</td>
<td>0.019</td>
<td>-0.468</td>
<td></td>
</tr>
<tr>
<td>Clinical</td>
<td>#BM/24hr</td>
<td>-0.237</td>
<td>0.043</td>
<td>-0.732</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mVS</td>
<td>-0.430</td>
<td>0.248</td>
<td>-0.687</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 51**: Principal Component Analysis A Produced Mixed Clusters of aEPEC Isolates. (A) PCA loading values listed and colored according to the level of contribution to each PC. PC summary of analysis A indicating the individual and cumulative proportion of variance and the PCs automatically selected by GraphPad Prims software. (B) Loadings vector plot reflecting the strengths and direction of contributors in A. (C) PCA-A scores plot with each aEPEC isolate indicated and labeled by symptomatic status. Q1= quadrant 1, Q2 = quadrant 2, Q3 = quadrant 3, and Q4 = quadrant 4.
PC analysis B (PCA-B) was performed with a select set of continuous variables from PCA-A. The PC summary data indicated greater captured variance in PC1-B (33.65%) with PC2-B contributing 15.94% for a total cumulative contribution of 49.59% (Figure 52A) than that captured by PC1/2 from PCA-A (44.20%) (Figure 51A). Genetic factors made the greatest contribution to PC1-B; Tir and EspH homology were the greatest contributors, then the number of fimbrial genes and all non-LEE effectors followed by total adhesins and EspA homology based on the aPCLV listed in Figure 52A. In addition, adherence level and mVS contributed to PC1-B (Figure 52A and B). In contrast, clinical manifestations made the greatest contributions to PC2-B with weaker contributions made by pedestal average, total adhesins, and fimbrial genes (Figure 52A and B). Importantly, distinct clusters based on symptomatic status were evident when aEPEC isolate scores were plotted from PCA-B (Figure 52C) with Q1 containing 100% SYMP isolates, Q2 and Q3 with fewer (80.95% and 65%, respectively) and Q4 with significantly lower (13.51%) (Table 11) indicating these selected eleven variables contribute to nearly 50% of the variance of aEPEC isolates grouped by symptomatic status.
Figure 52: Principal Component Analysis B Clustered aEPEC Isolates by Symptomatic Status. (A) PCA loading values listed and colored according to the level of contribution to each PC. PC summary of analysis B indicating the individual and cumulative proportion of variance and the PCs automatically selected by GraphPad Prims software. (B) Loadings vector plot reflecting the strengths and direction of contributors in A. (C) PCA-B scores plot with each aEPEC isolate indicated and labeled by symptomatic status. Q1 = quadrant 1, Q2 = quadrant 2, Q3 = quadrant 3, and Q4 = quadrant 4.

PCA-B was then analyzed in further detail to understand the relationship of the individual variables in PC1 and PC2 and to determine which aEPEC isolates clustered according to these variables. The PC score plots from PCA-B were repeatedly reexamined with aEPEC isolates labeled according to each individual variable (Figures 53-56). Tir percent identity clustered along PC1 (x-axis) with the least homologous Tir in Q1/3 (< 70% identity) (blue oval) and the greatest homology evident in Q2 (90-100%) (red oval) (Figure 53A). Similarly, these same clusters
corresponded to Tir Node clustering with the majority of Tir Node 1 and Tir Node 2 isolates in Q1/3, the majority of Tir Node 4 isolates in Q2/4, and Tir Node 3 isolates spanning Q1/4 (Figure 53B). Unexpectedly, a subset of Tir Node 1 isolates clustered in Q2 and Tir subtype labeling indicated that these isolates correspond to Tir subtype 5 (Figure 53B and C). The majority of the other isolates that constituted Q2 were those of Tir subtype 10 (Figure 53C). The remaining Tir subtypes clustered according to Tir Node groups (Figure 53B and C).

EspH percent identity clustered similarly to Tir clustering with the least homologous proteins found in Q1/3 (~70% identity) and the most homologous isolates clustering in Q2/4 (90 – 100% identity) (Figure 54A). EspA percent identity clustered opposite of Tir and EspH with Q2/4 having isolates with the least homology (70-80% identity) and Q1/3 having the greatest (≥90%) (Figure 54B). Of note, the 2 isolates that clustered in Q2 with 100% identity in EspA are from Tir subtype 5 (Figure 53C and Figure 54B). The number of fimbrial proteins and non-LEE effectors in aEPEC isolates had similar clustering patterns with isolates having the least number of genes present (0-10) in Q4 and all others having mostly similar numbers present ranging from 20-50 genes (Figure 54C and D).

**Figure 53: Tir Clustering Based on Principal Component Analysis B.** (A-C) PCA-B scores plot labeled by (A) percent identity of Tir, (B) Tir Node, and (C) Tir subtype. Q1= quadrant 1, Q2 = quadrant 2, Q3 = quadrant 3, and Q4 = quadrant 4.
Figure 54: Clustering of EspH, EspA, Fimbrial Protein Genes, and Non-LEE Effectors Based on Principal Component Analysis B. (A-D) PCA-B scores plot labeled by (A-B) percent identity of (A) EspH and (B) EspA, and the number of genes present with designation as a (C) fimbrial protein gene and (D) non-LEE effector. (A-D) Q1= quadrant 1, Q2 = quadrant 2, Q3 = quadrant 3, and Q4 = quadrant 4. Black arrows designate Tir subtype 5_Node 1 isolates.

Clustering of *in vitro* virulence phenotypes by PCA-B was not as distinct as genetic virulence factors (Figure 55) corresponding to the moderate to weaker contributions to PC1 and PC2 (Figure 52A). Adherence patterns were distributed throughout Q1-Q4 (Figure 55A); however, more defined patterns were associated with Q1 compared to Q4 (Table 11). Adherence level was highest in those isolates clustering in Q1 and Q3, Q2 had intermediate attachment, and Q4 isolates had the lowest levels of adherence (Figure 55B and Table 11). The average number of pedestals/host cell were highest in Q1 and decreased in a stepwise manner in Q2-Q4 with Q4 having the lowest average number of pedestals formed (Figure 55C and Table 11).
Figure 55: *In vitro* Virulence Phenotypes Clustered According to Principal Component Analysis B. (A-C) PCA-B scores plot labeled by (A) adherence pattern (B) adherence level as the fold compared to ΔbfpA, and (C) average number of pedestals formed per host cell. (A-C) Q1 = quadrant 1, Q2 = quadrant 2, Q3 = quadrant 3, and Q4 = quadrant 4. Black arrows designate Tir subtype 5_Node 1 isolates.

Clinical manifestation measures were also examined by PC analysis B and labeled according to each symptom (diarrhea duration and #BMS/24hr) or by the modified Vesikari score (mVS) (Figure 56). Similar clustering patterns were evident for each clinical manifestation in which the longest diarrhea duration, greatest number of BMs/24hr, and highest mVS was associated with Q1 isolates and the measures decreased in a stepwise manner in Q2-Q4 with Q4 having the average lowest duration, #BMS/24hr and mVS (Figure 56A-C and Table 11).

Figure 56: Clustering of Clinical Manifestations of aEPEC Infection According to Principal Component Analysis B. (A-C) PCA-B scores plot labeled by (A) diarrhea duration (B) number of BMs in 24hr, and (C) modified Vesikari score (mVS). (A-C) Q1 = quadrant 1, Q2 = quadrant 2, Q3 = quadrant 3, and Q4 = quadrant 4. Black arrows designate Tir subtype 5_Node 1 isolates.
Table 11: Summary of Principal Component Analyses

<table>
<thead>
<tr>
<th>Factora</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYMP positive (%)</td>
<td>100.00</td>
<td>++++</td>
<td>80.95</td>
<td>+++</td>
</tr>
<tr>
<td>Diarrhea duration (days)</td>
<td>14.79</td>
<td>++++</td>
<td>5.05</td>
<td>+++</td>
</tr>
<tr>
<td>#BMS/24hr</td>
<td>6.04</td>
<td>++++</td>
<td>4.10</td>
<td>+++</td>
</tr>
<tr>
<td>mVS (scale 0-9)</td>
<td>5.96</td>
<td>++++</td>
<td>4.43</td>
<td>+++</td>
</tr>
<tr>
<td>Adherence patternd</td>
<td>1.36</td>
<td>+++</td>
<td>1.10</td>
<td>++</td>
</tr>
<tr>
<td>Adherence level (%)</td>
<td>1.88</td>
<td>++++</td>
<td>1.35</td>
<td>++</td>
</tr>
<tr>
<td>Pedestal avg (%)</td>
<td>0.77</td>
<td>++++</td>
<td>0.59</td>
<td>+++</td>
</tr>
<tr>
<td>Tir (%)</td>
<td>74.4</td>
<td>++</td>
<td>94.4</td>
<td>+++</td>
</tr>
<tr>
<td>EspH (%)</td>
<td>82.4</td>
<td>++</td>
<td>97.3</td>
<td>+++</td>
</tr>
<tr>
<td>EspF (%)</td>
<td>80.4</td>
<td>+</td>
<td>90.3</td>
<td>+++</td>
</tr>
<tr>
<td>EspB (%)</td>
<td>81.0</td>
<td>++</td>
<td>97.1</td>
<td>+++</td>
</tr>
<tr>
<td>EspA (%)</td>
<td>85.7</td>
<td>++++</td>
<td>82.6</td>
<td>++</td>
</tr>
<tr>
<td>Total non-LEE (#)</td>
<td>20.07</td>
<td>+++</td>
<td>7.52</td>
<td>+</td>
</tr>
<tr>
<td>Fimbrial (#)</td>
<td>35.00</td>
<td>++++</td>
<td>18.29</td>
<td>+</td>
</tr>
<tr>
<td>Total adhesins (#)</td>
<td>77.54</td>
<td>+++</td>
<td>58.33</td>
<td>+</td>
</tr>
<tr>
<td>tEPEC non-LEE (#)</td>
<td>11.82</td>
<td>++++</td>
<td>5.52</td>
<td>+</td>
</tr>
<tr>
<td>Other non-LEE (#)</td>
<td>8.25</td>
<td>++++</td>
<td>2.00</td>
<td>+</td>
</tr>
<tr>
<td>AT adhesins (#)</td>
<td>7.25</td>
<td>++++</td>
<td>4.33</td>
<td>+</td>
</tr>
<tr>
<td>Fim operon/others (#)</td>
<td>31.50</td>
<td>equal</td>
<td>31.90</td>
<td>equal</td>
</tr>
<tr>
<td>Adhesins (#)</td>
<td>3.79</td>
<td>equal</td>
<td>3.81</td>
<td>equal</td>
</tr>
</tbody>
</table>

- Factors bolded were those analyzed in PCA-A/B, non-bolded were only analyzed in PCA-A.
- Quadrants corresponding to Figures 52-56 with the average value and relative level listed for each factor.
- Symptomatic status is a categorical factor and was not included in either PCA-A or B.
- Adherence patterns were assigned numerical values (DA=1, LAL=2, AA=3, UND=0) with avg. ≥1 indicating more defined patterns and avg. < 1 indicating more undefined patterns. As this is a categorical factor, adherence pattern was not included in either PCA-A or B.
Summary

aEPEC pathogenesis is likely a multifactorial process involving networks of effectors with coordinated efforts that subvert host physiology causing downstream effects leading to diarrhea. To begin to understand which aEPEC isolates cluster according to genetic risk factors, in vitro virulence phenotypes, and clinical manifestations of aEPEC infection, I employed principal component analysis (PCA) to reduce the dimensionality of the variables that correlate with symptomatic status. PC analysis B with the contracted 11 variables explained nearly 50% of the variance with two PCs and clustered isolates into distinct groups based on symptomatic status (Figure 52). Further analyses of the clusters within the PC score plot quadrants indicated relevant shared factors among the clustered isolates as summarized in Table 11.

I determined that aEPEC isolates which clustered in quadrant 1 of PC analysis B (Q1_PC-B) were 100% from SYMP individuals and had the most robust manifestations of infection including longer duration of diarrhea, increased number of BMs, and had the highest diarrhea severity scores. In addition, Q1_PCA-B isolates had the most robust in vitro virulence phenotypes, including more defined adherence, increased adherence level, and formed the most pedestals. As I hypothesized, Q1_PCA-B isolates housed a robust number of adhesins/fimbrial protein genes and non-LEE effectors and had high homology in the needle-protein, EspA. In contrast and contrary to my original hypothesis, homology was greatly reduced in the translocon protein, EspB, and in the effectors, Tir, EspH, and EspF. In addition, likely contributing to higher mVS scores, 40% of Q1_PCA-B isolates originated from individuals who had emesis during the course of their infection with significantly higher prevalence than isolates from other quadrants (Q2 – 29%, Q3 – 20%, Q4 – 3%) (Fisher’s Exact test, \( P = 0.0009 \)). The presence of emesis did not correlate with co-infection status; 50% of isolates originated from those who had a sole
EPEC infection and vomiting. The number of isolates originating from those with vomiting (38%) is similar to the number detected from the clinical cohort (34%).

Interestingly, isolates clustered in Q4_PCA-B had nearly the exact opposite composition of factors as those found in Q1 and only 13.51% were associated with SYMP individuals. Q4_PCA-B isolates had the lowest clinical manifestation measures, lowest adherence properties, lowest average pedestal formation, the fewest number of adhesins/fimbrial protein genes and non-LEE effectors and had isolates with the lowest homology in EspA. However, Q4_PCA-B isolates had higher homology in effectors Tir, EspH, and EspF and the translocon protein EspB compared to Q1.

The aEPEC isolates that clustered in Q3_PCA-B followed a different trend than that observed for Q1 and Q4 isolates. Fewer SYMP isolates (65%) were present in Q3_PCA-B compared to Q1. Internal clustering within Q3_PCA-B was apparent; SYMP isolates tended to cluster higher along the y-axis between PC2 value -1 – 0 while ASYM/OB isolates mainly clustered near PC2 value -2 (Figure 52C). Despite the sub-clustering in Q3, similarities were observed within the variables. Q3_PCA-B isolates had robust numbers of adhesins/fimbrial protein genes and non-LEE effectors and high homology in EspA which correlated with high adherence levels, similar to those in Q1. However, Q3 isolates had the lowest homology in Tir, EspH, EspF, and EspB and had only mild disease manifestations suggesting that greater adherence alone is not sufficient to cause robust disease and that specific variants of Tir and other effectors are needed as well. Other factors not examined in this study likely contribute to sub-clustering in Q3 warranting further investigation into isolate specific variability and how these may contribute to different pathogenic mechanisms.
The aEPEC isolates that clustered in Q2_PCA-B had the greatest variability among themselves and also did not follow the trends observed with Q1 and Q4 isolates. Instead, isolates of Q2_PCA-B had the greatest homology in Tir, EspH, EspF, and EspB while having low homology to EspA, similar to the pattern observed for Q4 isolates. In addition, Q2_PCA-B isolates had reduced numbers of adhesins/fimbrial protein genes and non-LEE effectors and had correspondingly moderate adherence properties. However, average pedestal formation was higher in Q2 isolates, as well as having more robust clinical manifestations present with 80.95% of Q2 isolates originating from SYMP individuals. Q2 isolates mainly originated from Tir subtype10_Node 4 with four from Tir subtype 5_Node 1. Upon further examination of Tir subtype 5_Node 1 isolates, 50% (n=2/4) had 100% homology within EspA, however, 75% (n=3/4) did not form pedestals. Despite the lack of pedestals in three Tir subtype 5_Node 1 isolates, 2 out of these 3 were from SYMP individuals. In addition, two of the Tir subtype 10 isolates of Q2_PCA-B had some of the most robust pedestal formation of all isolates. Although clustering based on symptomatic status was apparent in PCA-B, the variability in Q2_PCA-B underscores the notion that diversity of specific isolates may lead to different in vitro virulence phenotypes and clinical manifestations. In depth characterization and comparison of Q2_PCA-B isolates is warranted to understand other factors that likely contribute to variability of virulent phenotypes and clinical manifestations.

Limitations

One limitation of this study, as discussed in CHAPTER 6, is that in vitro virulence phenotypes were assessed in cultured SKCO-15 cells and phenotypes may become more robust when examined on the preferred intestinal segment. Therefore, when intestinal tropism is defined, greater correlation and clustering may become more evident.
Another limitation to this study is that in using PCA, the method inherently reduces the dimensionality of the data and PCs are determined by those variables which capture the most variance.\textsuperscript{232, 233} Therefore, outliers and variables that may cause small effects are often missed or excluded.\textsuperscript{233} This was evident in my data by the variability in \textit{in vitro} virulence phenotypes of isolates in Q2 and Q3 of PCA-B. In addition, PCA-B captured 50\% of the variance indicating other virulence factors are needed to fully define aEPEC pathogenesis. Also, I analyzed the number of non-LEE effectors and adhesins. However, the individual functions and collective effects of these virulence factors are likely what determine pathogenicity. PCA is a good exploratory tool, however, it does not predict the percent contribution of each virulence factor.\textsuperscript{233} Linear regression models or machine learning techniques with additional isolates would be better suited to predict the contributions that novel, specific, or collective virulence factors make to aEPEC pathogenesis. However, these methods are beyond the scope of this project.

Lastly, \textit{in vitro} systems do not always correlate with clinical manifestations as they do not account for host factors which may contribute to variable disease manifestations, such as the intestinal microbiota, intestinal tropism, host immune response, and others. Therefore, more advanced \textit{ex vivo} and animal models are needed for effective predictive modeling to assess the complex, multifactorial pathogenesis of aEPEC isolates.
CHAPTER 8

DISCUSSION

Clinical Manifestations of aEPEC Infections


EPEC has long been considered an infection of infants in developing countries. However, aEPEC is now one of the most commonly detected enteric pathogens in children and adults worldwide.³,⁴ Debate surrounds aEPEC pathogenicity, especially in adults; therefore, one goal of this study was to determine whether aEPEC causes significant diarrhea and determine if age contributed to differences in clinical manifestations of infection. Most patients studied were adults between 18 and 64 years (55%) and ≥65 years (32%). Genetic analysis of 82 isolates from stools of the EPEC-positive ASYM and SYMP cohorts revealed that 99% lacked the BFP operon, indicating that aEPEC infections outnumber tEPEC infections in my study. In addition, phylogenetic analysis based on 1000 core genes of all 112 aEPEC isolates from OB, ASYM, and SYMP individuals revealed that isolates did not cluster on geographic location, temporal acquisition, or by symptomatic status. Therefore, to my knowledge, this is the first study characterizing the symptoms and severity of sporadic aEPEC infection with a focus on adults from a developed nation.

EPEC-Positivity and Higher Stool Loads Are Associated with Symptomatic Individuals

I determined that GIP-EPEC positivity was strongly associated with diarrhea and moderately associated with vomiting, abdominal pain, and fever, despite the high prevalence of
ASYM individuals (~29%) in the EPEC-positive cohort and high numbers of GI SYMP individuals in the EPEC-negative cohort. It has been suggested that multiplex PCR detection assays may be too sensitive, and that quantitative PCR may be more useful to predict the probability of symptomatic infections and severity. A recent systematic review correlating gastrointestinal pathogen load with patient clinical presentations and outcomes reported overall higher pathogen load of *C. difficile*, norovirus, and *Shigella* was associated with patient disease or more severe healthcare outcomes.\(^{208}\) Furthermore, symptomatic EPEC infections are associated with higher EPEC loads in children and adults with cancer or who are immunosuppressed.\(^{125,199}\) Similarly, I found that EPEC loads were greatest in SYMP individuals. However, EPEC load is not as predictive of symptoms.\(^{125,208}\) Indeed, I found that some hospitalized and healthy-ASYM (OB) individuals had EPEC loads between 0.1% and 1.5% and that ~60% of symptomatic infections had EPEC loads of <0.1%, indicating that EPEC load may not directly correlate with clinical symptoms. Higher EPEC loads may be accounted for by co-morbidities, such as immunologic, metabolic, and diarrheal diseases, as they are often associated with intestinal microbiome dysbiosis with increased prevalence of *Enterobacteriacea*.\(^{197,198}\) However, no association was found in our study between IBD, diabetes, and liver or peptic ulcer diseases and GIP-EPEC positivity; renal disease was moderately associated. Although the systematic review described above reported higher loads associated with worse outcomes, patient clinical presentations and outcomes were often varied or contradictory with only 55% (12/22) of the studies reporting high pathogen load associated with symptomatic status or severity.\(^{208}\) The wide range of EPEC loads in ASYM and SYMP individuals of my study and varied and contradictory associations in other enteric pathogens, suggests that pathogen load is not a good predictor of symptomology or disease severity, especially for aEPEC.
Symptoms and Severity of aEPEC Infections

Diarrhea severity and most symptoms in children and adults aged 18 to 64 years were similar; some reported up to 40 BMs/d indicating severe diarrhea. In contrast to a recent finding that aEPEC infection caused diarrhea without most other symptoms in children, we found that fever, vomiting, and follow-up visits for diarrhea were more often associated with children than adults, contributing to children having the highest proportion of severe cases as determined by mVS. Fever and vomiting were present in some adults with EPEC infection; however, adults tended to have abdominal pain and longer diarrhea duration than children. On average, diarrhea lasted 11 to 18 days in adults and 7 days in children, with some reporting persistent and chronic diarrhea, similar to published reports. It is possible that the more robust host response seen in children results in more rapid clearance of aEPEC. Adults aged ≥65 years tended to have less severe diarrhea than children and had the greatest incidence of EPEC asymptomatic carriage. However, mVS and other diarrhea severity scoring systems do not consider, persistent or chronic diarrhea assigning equal weight to any diarrhea lasting 6-8 days or longer. An improved diarrhea severity scoring system is needed, especially for aEPEC infections, which cause longer diarrhea duration, to better highlight the clinical manifestations of this disease and to assess the long-term effects.

Severe diarrhea can lead to serum electrolyte imbalances. tEPEC causes watery diarrhea in murine models of infection and alters sodium and chloride absorption in cell culture models. However, we found that low serum sodium was the only electrolyte abnormality in pediatric patients and that low serum potassium in adults correlated with diarrhea severity even though 13% to 34% had electrolyte imbalances of any type. Similarly, a study of adults with acute diarrhea revealed that serum electrolytes and other serologic values were not significant
indicators of severe diarrhea. A limitation of this study was that electrolyte levels were not determined before or after the index date; thus, aEPEC causation could not be determined. Prospective studies are needed to determine whether aEPEC infection causes electrolyte and other laboratory abnormalities.

**Persistent/Chronic aEPEC Infection and Possible Long-Term Health Outcomes**

Most GI infections, including tEPEC, are self-limited, lasting 3 to 7 days; however, aEPEC infection has been reported to have longer duration. Persistent and chronic diarrhea was reported by 16% in this study. Analysis of GIP repeat tests revealed that some are infected with EPEC for extended periods of time, months to years, or repeatedly infected, even after antibiotic treatment. Further studies are needed to understand whether individuals are chronically infected with the same aEPEC strain or whether nascent infections arise. Acute infectious gastroenteritis can result in post-infectious functional GI disorders, with symptoms remaining for at least 8 years, can lead to IBD, or reactive arthritis. In addition, linear growth deficits in children aged <5 years from developing nations are associated with tEPEC and other bacterial causes of gastroenteritis. In addition, *E. coli* are part of the family *Enterobacteriaceae* which are often associated with increased abundance during disease states, such as during active UC or CD. In fact, one pathotype adherent-invasive *E. coli* (AIEC) are frequently isolated from biopsies of CD patients and due to long-term inflammation are linked to colorectal cancer. AIEC pathogenesis consists of three stages: adhesion, invasion, and replication within epithelial cells and macrophages, however, the pathogenic mechanisms have yet to be fully elucidated. In addition, emerging evidence indicates bacterial infections contribute to colorectal carcinogenesis through effectors called genotoxins that are found in several stains of *E. coli* and cause DNA damage by introducing single- and double-strand DNA
breaks in host IECs.\textsuperscript{241} Whether aEPEC infection causes any long-term health effects remains to be determined but warrants further investigation in light of long term outcomes of enteric infection and disease.

**Risk Factors for aEPEC Infection**

Risk factors strongly associated with EPEC detection were contact with another person with GI symptoms and international travel. Similarly, NORS outbreak study of enteric infection indicating that outbreaks were most commonly transmitted through direct person-to-person contact.\textsuperscript{102} Also similar to previous studies, we found that most aEPEC traveler’s diarrhea was associated with mixed infections predominated by EAEC and enterotoxigenic \textit{E coli}.\textsuperscript{99} However, \textasciitilde 20\% of international travel cases in this study had sole aEPEC infections, indicating that aEPEC can be a cause of traveler’s diarrhea.\textsuperscript{242} Some have suggested that aEPEC proliferates during other enteric infections but does not contribute to symptoms, leading some to question its virulence. However, we found that sole aEPEC infection caused similar severity of diarrhea as in mixed infections, with sole infections having longer duration of diarrhea than those with \( \geq 2 \) co-infecting pathogens. Similarly, co-infections of enteric pathogens did not increase risk of developing gastroenteritis in Dutch children aged 0-4 years.\textsuperscript{243} Additional studies are needed to understand whether the high prevalence of aEPEC detection with \textit{C. difficile} and EAEC detected in our study influences the severity of disease caused by these infections.

**Treatment Measures of aEPEC Infections**

After the introduction of BioFire GIP testing, the percentage of patients that were started on optimal medication for treatment of diarrhea increased and occurred sooner in their hospital course.\textsuperscript{127} Antibiotics have proved effective for treating EPEC infections in cancer or immunocompromised patients,\textsuperscript{199, 244, 245} but treatment regimens for sporadic, adult EPEC
infections have not been examined. Treatments recommended in this study ranged from minimal (oral hydration) to more extensive measures, including antibiotics and IV hydration, depending on symptom severity, underlying conditions, and co-infection status. Similarly, co-infection status likely played a role in the decision to prescribe antibiotics in adult, hospitalized patients from the UK with EPEC and other enteric pathogens. The rehydration therapy and antibiotics prescribed in our study were in accordance with Centers for Disease Control and Prevention recommendations for treatment of most diarrheagenic *E coli*. However, antibiotic use must be carefully considered in view of the prevalence of antibiotic resistance in diarrheagenic and non-diarrheagenic *E coli* and the risk for adverse reactions. In addition, our longitudinal data indicate that EPEC was detected in an adjacent test after EPEC-specific antibiotic treatment in ~38% of patients causing one to question if antibiotic treatment was effective. However, in some patients, 17 to 55 days separated antibiotic treatment and the adjacent EPEC-positive GIP results. Therefore, we could not discern from this data whether the secondary EPEC detection arose from a nascent infection or represents long-term colonization.

Based on the findings from our study, we suggest that until prospective case-control studies specifically determine the appropriate use of antibiotics for aEPEC infection, clinicians should take an individualized approach to treatment, with careful consideration of diarrhea severity and duration, presence of co-morbidities, and overall patient status. Typically, those with mild-moderate symptoms do not require antibiotic therapy because aEPEC infection is usually self-limited. Asymptomatic aEPEC carriers should not be treated with antibiotics due to unnecessary disruption of the gut microbiota and potential to generate antibiotic-resistant bacteria. If co-infections are present that necessitate antibiotic therapy, selection should focus on the co-infecting organisms, with consideration for the presence of aEPEC.
Due to the risk of antibiotics causing dysbiosis of the gut microbiota and the increased risk of antibiotic resistance, the use of probiotics to treat gastroenteritis has been explored by some in clinical trials. A meta-analysis of treatment of acute infectious diarrhea with the probiotics, *Lactobacillus rhamnosus* GG, *Saccharomyces boulardii*, or *Lactobacillus reuteri* indicates that probiotics likely made little or no difference in the duration of diarrhea.\(^{247}\) However, the mean diarrhea duration was shorter and stool consistency normalized more rapidly in children with acute gastroenteritis treated with Probiotical (*Streptococcus thermophilus*, *Lactobacillus rhamnosus*, *Lactobacillus acidophilus*, *Bifidobacterium lactis*, *Bifidobacterium infantis*, fructo-oligosaccharides).\(^{248}\) Therefore, future studies are needed to determine if treatment of infectious diarrhea with a consortium of prebiotics and probiotics rather than a single probiotic would result in greater efficacy to reduce duration and increase stool consistency. Mechanisms of action of the consortium of probiotics will also need to be investigated to determine if the probiotics are reversing diarrheal symptoms, reducing pathogen colonization, or both. Combination antibiotic and probiotic therapy may prove useful in more severe cases; however, further studies are needed. Overall, no serious adverse events were attributed to probiotic use\(^{247}\), and therefore, probiotics represent a low-risk treatment option for those suffering from acute infectious diarrhea.

**Conclusions from Clinical Manifestation Data of aEPEC Infections**

Our study of the clinical manifestations of EPEC infection suggests that aEPEC causes a wide array of symptoms in children and adults, ranging from asymptomatic carriage to severe diarrhea accompanied by fever, vomiting, and prolonged diarrhea and long-term detection in some individuals. We observed more asymptomatic carriage in adults ≥65 years. Diarrhea severity and most symptoms in children and adults aged 18 to 64 years were similar. However,
severe diarrhea occurred in some in all age groups; some reported up to 40 BMs/day and 16% had persistent/chronic diarrhea. Together, these data indicate that age does not contribute significantly to varying aEPEC disease manifestations. Our data suggest that simultaneous co-infection with another enteric pathogen and presence of co-morbidities also did not contribute to variability in clinical manifestations. Higher stool EPEC loads did correlate with symptomatic infection, but the range of detectable loads and the lack of association with diarrhea severity suggests that load is not a good predictor for disease severity. Although there are more host factors that may contribute to variable disease manifestations, our clinical data suggest that host factors play a minor role in contributing to variable disease manifestations of aEPEC infection. In view of EPEC genome heterogeneity\(^{192}\), I hypothesized that sub-lineages of aEPEC have greater virulence properties that account for variable disease manifestations. In support of this, although virulence factors were not examined, a recent study indicated that human isolates of aEPEC associated with diarrhea were derived from cattle and were genetically distinct from isolates originating from healthy individuals and other animals\(^{108}\) indicating distinct genetic lineages of aEPEC can account for disease manifestations. Therefore, another aim of this study was to characterize the genetic and \textit{in vitro} phenotypic bacterial virulence factors associated with symptomatic aEPEC infection to distinguish more virulent from less virulent strains.

**Bacterial Virulence Factors Associated with Clinical Manifestations of aEPEC Infection**

The pathogenic mechanisms of tEPEC infection are largely understood. In contrast, aEPEC pathogenic mechanisms are poorly understood and are confounded by high genomic heterogeneity among aEPEC isolates, detection in some asymptomatic individuals, and sometimes with other enteric pathogens. Also, it is assumed that aEPEC mechanisms are similar to tEPEC because of the detection of \textit{eae} and other LEE pathogenicity island genes in aEPEC
strains. Although some studies have examined aEPEC lineage and association of specific virulence genes with symptomology, none to date have examined the association of multiple genetic virulence factors in combination with in vitro phenotypes and how they may contribute to the variable disease manifestations of aEPEC infection. To the best of my knowledge, this is the first study indicating that aEPEC pathogenesis is multifactorial likely involving known virulence effectors and adhesins from tEPEC, EHEC, and other E.coli pathovars and suggests that variation in homology, the number of effectors and adhesins, and the presence of specific effectors account for symptomatic status and diarrhea severity.

**Phylogenetic Analysis of aEPEC Isolates**

Lineage determinants such as MLST, serotype, and phylogroup as well as more specific lineage designations, such as EPEC and LEE subtypes, help us understand evolutionary relatedness of aEPEC strains. Thus, one can assess whether EPEC detection results from a sporadic or outbreak infection, assess transmission mode, and can establish reservoirs for sources infection. A drawback to examining shared genes, lineage determinants often cannot distinguish between virulent, less virulent, and avirulent strains. However, in order to understand how our aEPEC isolates are related to previous strains, I assessed phylogroup, serotype, and Achtman MLST lineage determinants.

Using *in silico* phylogroup typing and phylogenetic analysis of 1000 core genes of our 112 aEPEC isolates, I determined that our isolates represented 6 out of the 7 known phylogroups with B1 the most prevalent (39.3%) followed by B2 (25%), A (21.43%) E (11.6%), E (1.79%) and F (0.89%). Phylogroup was not associated with symptomatic status. Phylogenetic analysis indicated major break points clustered by phylogroup while smaller branches clustered by MLST and serotype revealed high heterogeneity among the isolates indicated by the extensive
branching. Similar aEPEC phylogenetic analyses also indicate high heterogeneity by lineage determinants in aEPEC isolates.\textsuperscript{179,192,231} OB isolates originate from the Boston area while all ASYM and SYMP isolates originate from the Chicagoland area designated (CE). Despite this, isolates did not cluster on geographic location and represented sporadic infections, as evidenced by certain clonal clusters containing both OB and CE (phylogroup B2\textunderscore ST3303\textunderscore O87:H6 as an example) and CE designations displaying large numerical differences (CE34 vs CE280 of B2\textunderscore ST2086\textunderscore O33:H34), respectively.

I also assessed EPEC subtyping within my phylogenetic analysis by using EPEC subtype reference strains.\textsuperscript{179} Although phylogroup was not associated with symptomatic status, certain EPEC subtypes were more prevalent with OB/ASYM or SYMP isolates. For instance, phylogroup A had the highest percentage of SYMP isolates (75\%) due mostly to the large clonal group designated ST10 that also corresponds to EPEC10 subtype. Over half of the isolates from this subtype had a diarrhea severity score >5 with some as high as 9 mVS, representing some of the more moderate to severe cases in our study. In addition, the more severe cases in ST10\textunderscore EPEC10 were associated with individuals \(\leq 18\) years old. Interestingly EPEC10 was a novel lineage determined by analyzing isolates from the GEMS diarrheal children study; only 2 isolates were previously associated with EPEC10, but both were from non-lethal symptomatic (NSI) children.\textsuperscript{179} Future studies are needed to determine if the EPEC10 subtype is expanding in the US or if age is a contributing factor to virulence. In contrast, 2 novel clusters were detected in phylogroup B1. The first group had a greater frequency of SYMP isolates (62\%) and was more closely related to EHEC2, the defining pathotype of phylogroup B1. The second novel cluster was the most distal to EHEC2 with the majority OB/ASYMP isolates (80\%). Similarly, phylogroup B2 had a large cluster of OB/ASYM isolates (65\%). Although its closest relative was
EPEC4, most isolates represented novel EPEC subtypes that were distally related to tEEPC_E2348/69, the strain which originally defined B1 phylogroup and EPEC1 subtype. These data suggest that with continued evolution, concomitant loss of pathogenicity occurs. A recent phylogenetic analysis of over 6000 E. coli genomes indicates that all E. coli stem from 3 ancestral lineages: (1) phylogroup D, the most closely related to the ancestral core genome, (2) E/B1/A – diversification was driven by gain of carbohydrate utilization and metabolic flexibility, and (3) F/G/B2 – diversification due to loss of environmental-adaptive functions or changes in colonization-virulence. These adaptive trajectory functions were defined by SNPs and gain/loss of genes and account for the current 7 phylogroups. My phylogenetic analysis is similar to the evolutionary diversification proposed by this group and suggests that gain or loss of specific virulence determinants likely dictate disease manifestations and severity. The remainder of my study focused on identifying virulence determinants of aEPEC.

**Other Lineage Determinant Analyses**

aEPEC isolates had great diversity in serotype and MLST designations with 50 known and 1 unknown MLST designations, 47 known and 7 untypeable O-antigens (LPS), and 26 H-antigens (flagella). None associated with a specific symptomatic status group. Of the eleven classic O-serotypes stated by WHO to be associated with EPEC infection, I detected only 5 with very few in each, O26 (2.7%), O55 (2.7%), O119 (5.4%), O127 (5.4%), and O128 (1.8%). Similarly, some common EHEC/STEC-serotypes also detected in my study were O157 (4.5%) and O145 (1.8%). Along with the more common O119 and O127, frequently detected O-serotypes were those that are unknown (6.3%), O137 (5.4%) and O87 (5.4%). The remaining O-serotypes were detected at frequencies ranging from 0.9 – 4.5%. Due to the high diversity in MLST and O-serotypes the contribution to virulence is difficult to ascertain. However, a recent
study demonstrated that during EPEC infection Tir clustering, independent of actin polymerization, triggers rapid Ca\(^{2+}\) influx, which induces LPS internalization, followed by activation of caspase-4 and pyroptosis.\(^{250}\) LPS also induces TLR4 signaling activating MAPK, NF-κB, and other pro-inflammatory response. Therefore, it is interesting to speculate how different O-antigens and LPS may induce differential immune stimulatory outcomes from aEPEC infection.

Flagella contribute to multiple aspects of bacterial virulence, including motility, adhesion, biofilm formation, and immune modulation through H-antigen and activation of TLR-5.\(^{251}\) There are 53 H-antigens in *E. coli* with 43 encoded by genes in the *fliC* locus and the remaining encoded by *flk, fll*, and *flm*.\(^{252}\) Our aEPEC isolates where characterized by 26 different H-antigens all from the *fliC* locus. H6, also present in tEPEC_E2348/69, was the most prominent H-antigen of our isolates but was only detected in 14 isolates (12.5%). The remaining H antigens ranged from \(~0.9\% – 8\%\) prevalence. Some noteworthy H-antigens also detected in our strains were H8 (8.0%), H2 (7.1%), H7 (5.4%), H4 (4.5%), and H40 (3.6%). H8 antigen is associated with outbreaks of STEC and EPEC in the US and Japan.\(^{253}\) H2 from prototypical tEPEC_B171, H6 from tEPEC_E2348/69, and H7 from EHEC_Sakai/EDL933 and neonatal meningitis *E. coli* are involved in invasion and adherence to IECs and mucins.\(^{254}\) Interestingly, specific human microbiota metabolites upregulate *fliC* and other motility genes, enhancing EHEC motility and pathogenicity observed in a human colon-chip infection model.\(^{255}\) H4 flagella of the ST131 reference strain EC958_UPEC (also present in ST131 STEC O104:H4 that caused a large outbreak of diarrhea and HUS in Germany in 2011) contributed to motility, adherence to and invasion of bladder epithelial cells, and enhanced bacterial uptake by macrophages. This phenotype was also observed for H1 and H7 and although H1, H4 and H7 all induced IL-10
secretion from monocytes co-cultured with uroepithelial cells, H4 caused the greatest induction.\textsuperscript{254} Flagella of aEPEC 1711-4 associated with O51:H40 are required for efficient adhesion, invasion, and early but not late IL-8 production in intestinal epithelial cells \textit{in vitro}.\textsuperscript{256} Together, these data exemplify the important role of flagella in pathogenicity of EPEC and EHEC/STEC strains. However, with the abundance of diversity of H-antigens among our aEPEC isolates I could not determine their role in virulence. Future studies focused on individual contribution related to adherence, motility and immune activation are essential in determining aEPEC pathogenicity.

**Phylogeny of Intimin, Tir, and Other LEE-Associated Proteins**

Intimin is one of the most important EPEC virulence factors mediating intimate attachment to the host through its receptor Tir. Thus, \textit{eae} and lack of other pathovar genes is the definition of EPEC-positivity in stools. SNPs located in the C-terminal region of \textit{eae} has allowed for the distinction of at least 35 intimin subtypes with beta1, epsilon1, and theta subtypes being the most frequently detected in aEPEC isolates thus far.\textsuperscript{9,175} Our aEPEC isolates were similar to 21/35 known intimin subtypes; the most frequently detected included: theta (14%), beta1 (12%), and epsilon1 (11%) along with beta2 (13%), iota1 (9%), and epsilon2 (6%). The remaining 15 intimin subtypes were detected at frequencies of only 0.9-4.5% representing detection of only 1-5 isolates within any intimin subtype. With such great diversity, I did not detect an association between intimin subtype and symptomatic status. In fact, intimin had \( \geq 80\% \) homology to the reference strains tEPEC-E2348/69 and EHEC-Sakai regardless of symptomatic status indicating that intimin diversity alone does not dictate virulence in our aEPEC isolates. Interestingly, intimin-alpha1 from tEPEC\_E2348/69 displayed different intestinal tropism than intimin-gamma1 from EHEC\_Sakai and from O55:H7 EPEC indicating intimin contributes to tissue
specificity.\textsuperscript{257, 258} In addition, intimin is part of a larger family of invasin proteins often found in \textit{Yersinia} which are known to bind β1-integrin in the host membrane suggesting there may be binding sites for intimin on host IECs.\textsuperscript{259} EHEC intimin-gamma binds host β1-integrin with low affinity and nucleolin (expressed in nuclear and outer cellular membranes) in a Tir-independent manner.\textsuperscript{259} However, the contribution that intimin diversity as well as the role of Tir-independent binding has on virulence, especially in humans, remains to be determined.

Next, I examined Tir for its role in intimate attachment as well as the formation of pedestals and A/E lesions. In contrast to intimin, Tir homology to tEPEC/EHEC was significantly more diverse in SYMP isolates compared to OB isolates. Similarly, and in contrast to my original hypothesis, most LEE-encoded effectors and translocon proteins had less homology in SYMP isolates compared to OB isolates with significant differences detected for EspH, EspF, and EspB. In contrast, EspA had significantly greater homology in SYMP isolates compared to OB. LEE-encoded regulators, T3SS components and chaperones all had high conservation regardless of symptomatic status. I next explored Tir phylogeny using reference Tir subtypes as determined by Ingle et. al from the GEMS diarrheal study in children.\textsuperscript{176} Interestingly, I defined four major Tir Nodes which correlated with the greatest frequency of isolates from symptomatic individuals in Tir Node 1 (77.8%) and Tir Node 2 (76.9%), followed by Tir Node 3 (62.5%) and Tir Node 4 (46.2%). Previously defined Tir subtypes clustered within the newly defined Tir nodes, with Tir node 1 housing subtypes 1/2/5, Tir Node 2 housed subtypes 3/4, Tir Node 3 was exclusively subtype 6, and Tir Node 4 consisted of subtypes 7-11. Additional sub-lineages of Tir were evident within the Tir subtypes and interestingly, corresponded to the diverse intimin subtypes indicating co-evolution of Tir-intimin pairs. Similarly, homology of EspH, EspF, and EspB positively correlated and EspA negatively
correlated with Tir homology. In addition, protein phylogenetic sub-lineages of EspA, EspB, EspD, EspH, and EspZ correlated with Tir Nodes indicating co-evolution of LEE-encoded effectors and translocon proteins with Tir. LEE lineage subtyping resulted in ~40% of isolates representing novel LEE subtypes (Table 7) which clustered with Tir Node 1-3. These data suggest that specific Tir Nodes and subtypes and the corresponding LEE subtypes likely contribute to variance in disease manifestations of aEPEC infection. Therefore, divergence in Tir homology is likely a good indicator of aEPEC virulence with Tir Nodes 1 and 2 corresponding to greater genetic virulence factors. In support of this hypothesis, I also determined that Tir Node 1 and 2 had the greatest number of adhesins and non-LEE effectors, caused the most robust in vitro virulence phenotypes, and corresponded to higher EPEC loads and greater association of emesis in patients. No significant differences were detected in diarrhea duration, #BMs/24hrs, or mVS between isolates from the different Tir Nodes, however, more robust pedestal formation and increased number of non-LEE effector common to tEPEC/EHEC correlated with an increased diarrhea severity score. Similarly, increased number of autotransporter (AT) adhesins correlated with greater #BMs/24hrs. Together these data indicate that Tir homology is a major virulence determinant for aEPEC pathogenicity. However, aEPEC virulence likely depends on several virulence factor pathogenic mechanisms including effector and translocon protein homology and the number of adhesins and non-LEE effectors harbored by the bacteria. In fact, I determined that specific non-LEE effectors and adhesins were associated with symptomatic status. Therefore, it is likely that in certain Tir virulence backgrounds, the presence or absence of different effectors and adhesins explain variability in clinical manifestations.

Despite my evidence indicating that Tir homology and presence/absence of non-LEE effectors and adhesins contribute to virulence, there are likely more strain-specific factors that
contribute to aEPEC pathogenicity. Evidence of such specificity is displayed in certain Tir subtypes, namely Tir Node 1_subtypes1/5 and in Tir Node 4 subtype10. Tir Node 1_ST1 isolates all originated from individuals who had emesis and all but one was associated with intimin-omicron. These Tir Node 1_ST1 isolates represent a completely novel LEE subtype not seen in the study by Ingle et.al.\textsuperscript{176} Intimin_omicron is associated with an invasive strain of aEPEC.\textsuperscript{9} This subgroup likely represents a novel pathogenic mechanism worth further investigation.

Tir Node 1_ST5 also displays sub-lineage strain specificity. The majority of Tir Node_subtype5 isolates did not form pedestals (78%), yet 67% originate from symptomatic individuals. I further investigated their homology to EHEC and tEPEC and determined that these isolates had greater homology to EHEC Tir which lacks the critical tyrosine 474 found in tEPEC and instead uses TccP as the coupler to the N-WASP-Arp2/3 complex.\textsuperscript{154} Therefore, I explored this mechanism in more detail in our 112 aEPEC isolates. TccP/TccP2 was detected in 31\% \((n=35)\) of our aEPEC isolates and was not associated with symptomatic status. In contrast, a high prevalence of TccP/TccP2 was detected in a study of 90 aEPEC isolates with nearly half positive for TccP/TccP2.\textsuperscript{177} Of those TccP positive, the majority harbored Tir\textsubscript{Y474} suggesting this virulence factor was acquired by EPEC lineage strains. In contrast, 25\% were TccP-positive/Tir\textsubscript{Y474-negative} suggesting these strains descend from EHEC and lost Stx.\textsuperscript{177} Similarly, I determined that Tir Node 1_subtype5 as well as Tir Node1_subtype2 isolates lacked Tir\textsubscript{Y474} but only one isolate in TirNode1_subtype 2/5 was TccP positive indicating that the few strains able to form pedestals do so in a Tir\textsubscript{Y474}-TccP-independent mechanism suggestive of a novel pathogenic mechanism. Both tEPEC and EHEC can form pedestals by alternative mechanisms to Tir\textsubscript{Y474} or TccP.\textsuperscript{154} Considering the expansive repertoire of accessory genes associated with SYMP isolates as determine by my pangenome analysis, further investigation of virulence
factors that allow for pedestal-dependent and -independent mechanisms is needed in these Tir subtypes. These data reiterate the notion that specific virulence factors have been lost or gained and likely account for the variability in in vitro phenotypes and clinical manifestations.

Lastly, Tir Node 4_subtype10 isolates had the fewest genetic virulence factors, yet 53% were from SYMP isolates. Of those from SYMP individuals, most had robust in vitro phenotypes and moderately high diarrhea severity scores due to duration of diarrhea of up to 7-20 days and up to 10-12 BMs/24hr in some. I also determined that all Tir Node 4_subtype10 isolates originated from phylogroup B2 with a subset related to subtype EPEC4. Phylogroup B2 was originally defined by tEPEC_E2348/69, however, none of our isolates were clonally similar to E2348/69. Some had a more proximal relationship with tEPEC-E2348/69 but most were distant relatives that retained about 50/50 mix of symptomatic to asymptomatic suggesting that repeated gain or loss of virulence factors contributes to mixed virulence manifestations. Similarly, in two previous studies, EPEC4 isolates originated from children with or without diarrhea (1-symp and 1-asym)\textsuperscript{231} and from the GEMS study, 2 were lethal cases (LI) while 2 were symptomatic (NSI) and 2 asymptomatic (AI) (of note all 6 were bfpA\textsuperscript{+})\textsuperscript{179}. Future studies are needed to decipher what gains and losses have occurred to drive symptomatology or not in this mixed population of isolates.

In conclusion, my data indicate that aEPEC virulence stems from a multifactorial mechanism involving many virulence factors such as homology of Tir, effectors EspH/F, and translocon proteins EspA/B, as well as the number of non-LEE and adhesins present. These factors lead to more robust in vitro virulence phenotypes such as defined adherence patterns, increased adherence to host cells, and greater pedestal formation as well as more pronounced clinical manifestations. However, strain specificity likely not only occurs from gains and losses
of the virulence factors specified here, but also from unknown factors yet to be determined. My pangenome analysis revealed a greater number of accessory and isolate-specific genes present in isolates from symptomatic individuals indicating the abundance of diversity that is yet to be explored.

**aEPEC Infection Likely Results from a Multifactorial Pathogenic Mechanism**

In a study of *C. rodentium*, using sequential gene deletions of its 31 effectors, machine learning modeling, and experimental confirmation, it was determined that substantial contractions in effectors were tolerated by *C. rodentium* and that effector essentiality for infection was context dependent.\(^{159}\) This study indicated that rather than working individually, it is the collection of the effectors that forms robust networks that work in a coordinated manner to achieve virulence.\(^{159}\) Interestingly, different effector networks were shown to be virulent but triggered different immune responses.\(^{159}\) Furthermore, this study determined that Tir, EspZ, and NleA were essential effectors. Although common between *C. rodentium* and tEPEC_E2348/69, replacing the *C. rodentium* with E2348/69 effector network did not enable efficient colonization indicating effector networks are specific to the host. Thus, results from this mouse model are not directly applicable to aEPEC infection in humans. However, this study reveals much about the redundancy of effectors and how specific but different effector networks can colonize mice as efficiently yet cause different immune responses. One can speculate that different aEPEC effectors or adhesin networks or a combination would elicit different immune or other host responses resulting in variable clinical manifestations.

In fact, using PCA I determined that distinct clusters of isolates were evident, correlating with genetic and *in vitro* phenotypic virulence determinants and severity of disease. All Cluster 1 isolates originated from symptomatic individuals and had the most robust *in vitro* phenotypes.
and the highest diarrhea severity scores. Increased severity was due to reporting of the longest duration and greatest number of bowels movements as well as the greatest percentage with emesis. Emesis isolates tended to cluster along the border between Cluster 1 and Cluster 3 isolates. Genetic virulence factors of Cluster 1 included the greatest homology in EspA, the greatest number of non-LEE effectors and adhesins, but lower homology in Tir, other LEE effectors (EspH and EspF), and the translocon, EspB.

Cluster 4 isolates had nearly the exact opposite genetic and phenotypic virulence factors as Cluster 1 with ≥85% of isolates originating from asymptomatic individuals. Cluster 4 had the least in vitro phenotypes, lowest diarrhea severity scores, least EspA homology, low to intermediate numbers of adhesins and non-LEE effectors, and high to intermediate homology in Tir, EspH, EspF, and EspB. This pattern supports my original hypothesis that greater genetic and phenotypic virulence factors correlate with increased disease manifestation, except for the surprising finding that greater diversity in effectors, but not EspA, was more prominent in symptomatic individuals. Of note, all our SYMP isolates originated from individuals who had mild-moderate diarrhea severity with none from severe cases. Therefore, the pathogenic mechanisms of aEPEC that cause severe symptoms may be different.

Cluster 3 isolates were similar to Cluster 1, but with some key differences. Cluster 3 isolates had the greatest number of non-LEE effectors and adhesins and greatest homology to EspA, similar to that seen in Cluster 1 isolates, and correspondingly had high adherence levels. However, Cluster 3 had fewer isolates with defined adherence and pedestal formation and had less severe clinical manifestations. Interestingly, Cluster 3 had the least homology in Tir, EspH, EspF, and EspB suggesting that too much variance in these proteins may result in less virulence.
In addition, isolates in Cluster 2 had more variable genetic factors that did not fit well into the trend described above. The majority originated from Tir Node 4_subtype10 with a subpopulation originating from Tir Node 1_subtype5; all were phylogroup B2 which is defined by tEPEC-E2348/69. Isolates from Tir subtype10 had the least number adhesins and low homology to EspA corresponding to reduced adherence properties. Conversely, Tir subtype10 isolates retained high homology in Tir and other effectors but had the lowest number of non-LEE effectors corresponding to variable pedestal formation; 2 isolates had very robust pedestal formation while most had either very low or undetectable pedestals. Although clinical manifestations varied, Tir subtype10 isolates caused the second most robust severity of diarrhea. By examining the phylogenetic tree (Figure 23), Tir subtype10 isolates are descendants of tEPEC-E2348/89, suggesting   that overtime many virulence factors have been lost from the tEPEC strain, however in some, other virulence factors have been gained that are likely novel to the virulence factors determined in this study.

The four isolates of Cluster 2 that originated from Tir Node 1_subtype5 were characterized by high homology to Tir and EspH, low numbers of non-LEE effectors and adhesins, and variable homology to EspA. Phenotypically, most did not form pedestals, and had diverse adherence patterns and variable adherence levels. In addition, clinical manifestations also varied among these isolates indicating individual specificity. Interestingly, 3 out of 4 were from SYMP individuals who also had co-infecting pathogens present with EPEC and may account for the variability in the genetic and phenotypic factors as well as the variable clinical manifestations. In contrast, the ASYM isolate originated from an EPEC only case indicating that in this particular group of isolates, co-infecting pathogens may contribute to variable disease
manifestations. In summary, Cluster 2 isolates are more unique and clinical manifestations likely result from individual strain variance or co-infection status requiring further investigation.

Overall, I determined that the greater number of adherence factors and non-LEE effectors as well reduced homology in Tir and other LEE effectors but high homology in EspA, play a role in aEPEC virulence. Despite this compelling evidence suggesting an aEPEC pathogenic mechanism, this network of virulence factors accounted for ~50% of variance. Therefore, it is likely that specific LEE and non-LEE effectors and adhesins with their combined effects contribute to aEPEC virulence. It is important to understand the function of the specific virulence factors that had greater associations with symptomatic isolates and those from specific Tir Nodes in order to determine how combinations of effectors influence aEPEC virulence.

**Specific LEE Genetic Factors Associated with Symptomatic aEPEC Infection**

I determined that reduced homology in Tir, EspH, EspF, and EspB was associated with isolates from symptomatic individuals whereas reduced homology in EspA was associated with asymptomatic isolates especially those from Tir Node 4_subtype10. Most of these proteins have multiple domains and are multifunctional, therefore, depending on where SNPs or INDELs occur the change in protein activity could be profoundly different. Detailed protein sequence alignment and functional assays need to be employed to determine how these changes may be affecting aEPEC virulence.

**Tir.** Important domains of Tir include the flexible N-terminal intracellular homodimerization domains\textsuperscript{260}, two transmembrane domains of which TMD2 is more important for proper secretion while TMD1 is more critical for activity of the protein within host cells\textsuperscript{261}, the extracellular domain which consists of two \(\alpha\)-helical domains required for proper positioning of the Tir-intimin binding domain\textsuperscript{150}, and the disordered C-terminal domain with phosphorylation
sites involved in actin recruitment and polymerization required for pedestal formation. The Tir binding domain directly interacts with domain 3 of the lectin-like extracellular domain of intimin. In view of the many different intimin subtypes that correspond to specific Tir subtypes in my study, one questions if Tir-intimin pairs evolved together ensuring affinity is preserved or do certain Tir-intimin pairs diminish or increase binding affinity? Single point mutations within Tir binding domain have profound effects on Tir-intimin binding affinity; a 20-fold reduction in binding affinity occurred when EHEC Tir interacted with EPEC intimin indicating specificity in Tir-intimin interactions which could greatly alter virulence. Similarly, C-terminal mutations could have drastic effects promoting novel Tir binding partners, altering affinity for known bindings partners, or promoting Tir
\textsuperscript{Y474}-independent mechanisms, thus pedestal formation may be promoted or suppressed depending on the mutation.

**EspB and EspA.** EspA and EspB along with EspD form the translocation machinery complex connecting the bacteria to the host cell allowing for effectors to be directly injected into the host cell. EscF connected to the bacterial cell wall forms the base of the needle which binds EspA, the needle filament and translocation tube. EspA binds EspD then EspD binds EspB. EspD and EspB form the pore that spans the host cell membrane allowing for delivery of effectors. Interestingly, in my study, EspD was highly conserved among all isolates and greater virulence was associated with conserved homology in EspA filament but reduced homology in EspB. Mutations in EspD are likely not tolerated as EspD is the major pore forming protein of which multiple interactions are required for secretion including homo-oligomerization, oligomeric complexes with EspA and EspB, and insertion and interaction of its transmembrane helix hairpin into the host cell membrane resulting in pore formation. Similarly, EspA filament is a highly complex structure involving extensive protein-protein interactions among 5 protein
strands that make up 1 sub-unit, ~10 sub-units interact to form 1 turn of the filament, and the filament can extend up to 600nm, largely dependent on the amount of EspA expressed.\textsuperscript{264-267} Single point mutations in essential residues of EspA as well as EspA originating from different serotypes produced stubby, vestibular filaments or none at all.\textsuperscript{265, 266} Although high conservation of structure is needed, one can image that different lengths of EspA filaments, controlled by expression, chaperones, or the needle “ruler” EscP, may be beneficial for virulence allowing for penetration of the intestinal mucus layer.\textsuperscript{264, 267} Interestingly, EspA filaments are shorter in \textit{espD} mutants but not with \textit{espB} mutants and EspB maintains function even when mutated.\textsuperscript{152, 268} EspB also acts as an effector in the host where it binds to myosin and α-catenin, inhibiting its interaction with actin and contributing to microvilli effacement and EPEC phagocytosis.\textsuperscript{152, 268} Although it is unknown, EspB mutations may influence interaction with the host membrane or downstream effector functions warranting investigation into the specific mutations found in our aEPEC isolates.

\textbf{EspF and EspH.} As described in CHAPTER 2 “Typical EPEC pathogenesis”, EspF and EspH are multifunctional proteins that contribute to disruption of ion and solute transporters, TJ structure and function, mitochondrial integrity. Both EspF and EspH contribute to actin remodeling promoting pedestal formation. EspF also contributes to ER stress. In contrast, EspH reduces filopodia formation and is involved in suppressing NF-κB and MAPK signaling. With so many diverse functions it is hard to speculate as to how diversity in these effectors may contribute to aEPEC virulence. However, one speculation is that with non-LEE effector loss EspF and EspH may have mutated to become less disruptive to maintain the characteristic balance between pro-and anti-inflammatory mechanisms. Or they may have developed more robust mechanisms to promote pedestal formation in view of the reduced homology in Tir.
Detailed analysis of these effectors is warranted due to their diverse and vast functions contributing to EPEC virulence.

**EspZ and EspD.** Although a significant association was not detected, I found that EspZ tended to have greater diversity in symptomatic isolates, especially those located in Tir Node 2_subtype4 and some in Tir Node 1_subtype2. In addition, LEE lineage subtyping resulted in ~40% of isolates representing novel LEE subtypes (Table 7). This was largely due to EspZ subtyping not matching its cognate subtype partners. For instance, several of our isolates matched all 6 subtypes for *eae, tir, espA, espB, espD and espH*, which defined Tir subtype7 but they did not match the canonical *espZ_ST7*\(^{176}\), instead they had *espZ_ST10*, indicating a novel LEE subtype. Therfore, one questions how this variability in EspZ may affect aEPEC infection.

EspZ is an effector secreted early during infection and is inserted into the membrane near pedestals where it acts as a functional gatekeeper for other effectors restricting translocation into the host cell.\(^{156}\) Thus, EspZ reduces cytotoxicity in the cell by regulating effector translocation. EspZ interacts with the translocon protein EspD.\(^{156}\) In addition, EspZ gatekeeping function is dependent on its extracellular domain which is restrictive based on strain specificity, i.e. EspZ from EHEC O157:H7 could restrict translocation from either EPEC O127:H6 or EHEC O55:H7, however, EspZ from EPEC was only restive of itself.\(^{156}\) The authors speculated this phenotype resulted from interaction with a secreted protein and/or with a bacterial outer membrane protein, such as O-antigen. However, comparisons in homology or SNPs of EspH and EspD within the strains were not performed, therefore, variance in protein-protein interactions between EspZ-EspD cannot be ruled out. Interestingly, one could imagine a scenario where greater diversity or even SNPs in EspZ but not in EspD would alter EspZ’s restrictive activity allowing for greater effector translocation. This would allow aEPEC isolates to overcome adhesion shortcomings.
through greater effector delivery resulting in more robust *in vitro* phenotypes or clinical manifestations. Future studies of the EspZ-EspD interaction and translocation activity and how they impact aEPEC infection are warranted.

**Specific Non-LEE Genetic Factors Associated with Symptomatic aEPEC Infection**

In my study, I determined that the number of non-LEE effectors was greater in symptomatic compared to asymptomatic isolates. In addition, specific non-LEE effector genes were more prevalent in isolates from symptomatic individuals including: NleB, NleE, EspL2, and LifA (housed on IE6_tEPEC-E2348/69); EspJ of PP2_tEPEC-E2348; and NleG of PP4_tEPEC-E2348/69, NleG5-1, and NleG6-1 of EHEC-Sakai. Similarly, several studies have found *nleB, nleE, lifA/efa1* associated with diarrheal patients more often than asymptomatic individuals.\(^{178, 227, 269}\) Although EspJ and NleGs have been detected in aEPEC isolates\(^{176, 178}\), this is the first study to determine that they are associated with symptomatic status.

**EspL2.** My data reveal that EspL2 (33.8%) was more prevalent in aEPEC isolates originating from symptomatic individuals than asymptomatic (9.7%). EHEC EspL2 and annexin2 are recruited and directly interact in Tir-independent mechanisms at sites of infection under attached bacteria.\(^{270, 271}\) EspL2 promotes annexin2 actin bundling activity and alters membrane morphology.\(^{270}\) EspL2 and annexin2 work in concert with Tir-dependent actin polymerization to induce actin microfilament aggregation promoting condensed the 3D bacterial colony.\(^{270, 271}\) In the absence of EspL2 or annexin2, adhered bacteria form pedestals, but each pedestal is distinct from the other.\(^{270, 271}\) It will be interesting to determine what role EspL2 has in aEPEC infection and if its presence promotes pedestal formation or adherence pattern.

**EspJ.** I determined the EspJ was more prevalent in aEPEC isolates originating from symptomatic (51.5%) than from asymptomatic individuals (19.3%). EspJ_EPEC is a non-LEE
T3SS bacterial effector protein which post-translationally modifies and inactivates a range of host nonreceptor tyrosine kinases (NRTK) using novel amidase and ADP-ribosyl transferase activity.\textsuperscript{272, 273} EspJ preferentially ADP-ribosylates the Src, Abl, Csk, Tec, and Syk NRTK families, which are membrane tethered and involved in a myriad of functions including: phagocytosis, signal transduction downstream of Toll-like receptors (TLR), complement activation, and modulation of cell cycle, morphology, and motility through DNA- and actin-binding domains.\textsuperscript{272, 273} Interestingly, not only does EspJ modulated cell immune responses, it also plays a role in regulation of pedestal formation. Over-expression of EspJ reduces actin polymerization and pedestal formation after EPEC infection through its inhibition of Src and Abl NRTKs, the kinases involved in Tir phosphorylation required Nck–N-WASP recruitment triggering actin polymerization.\textsuperscript{272} Tir is one of the first effectors to enter the host cell, whereas EspJ is secreted at later infection timepoints. It is interesting to speculate the role EspJ may have in regulating pedestal formation or disassembly at late stages of infection. Although an \textit{espJ} mutant of \textit{C. rodentium} colonizes mice as efficiently as the parental WT strain\textsuperscript{273}, the role of EspJ in inhibiting phagocytosis and modulating host immune responses warrants further investigation into its effects during aEPEC infection of humans.

\textbf{NleE, NleB1, and NleG.} I determined that NleE (35.3\%) was more prevalent in aEPEC isolates originating from symptomatic than asymptomatic individuals (12.9\%). NleE is a cysteine methyltransferase which modifies the TAK1-binding proteins, TAB2 and TAB3, blocking their recruitment to TAK1, and resulting in inactive TAK1.\textsuperscript{274} Therefore, NleE through inhibition of TAK1 serine/threonine kinase activity, also inhibits MAPK and NF-κB activation resulting in suppression of downstream gene transcription involved in immune and inflammatory responses, cell proliferation, autophagy, tissue remodeling, and metabolic regulation.\textsuperscript{274, 275}
My data reveal that NleB1 (33.8%) was more prevalent in aEPEC isolates originating from symptomatic than asymptomatic individuals (9.7%). NleB1 is a glycotransferase which modifies FADD, RIPK1, and TRADD preventing DISC assembly and caspase-8 activation. Thus, NleB1 inhibits extrinsic apoptotic signaling via the FAS death receptor. In the *C. rodentium* mouse model, NleB contributes to colonization.

My data reveal that NleG from (22.1%) and NleG5-1 (27.9%) and NleG6-1 (32.4%) were more prevalent in aEPEC isolates originating from symptomatic individuals than asymptomatic (0%, 0%, 3.2%, respectively). NleG homologues constitute the largest family of T3SS effectors delivered by pathogenic *E. coli* with fourteen members present in EHEC_Sakai of which eight have been confirmed to be expressed and secreted. Over 20 NleG homologues can be found in different strains of pathogenic *E. coli* and *Salmonella*. In contrast, EPEC_E2348 has only one NleG effector located on PP4. NleG homologues have a conserved N-terminus responsible for interacting with host protein targets and a C-terminal U-box E3 ubiquitin ligase domain that engages with the host ubiquitination machinery targeting protein for degradation. NleG_PP4(EPEC) has been implicated through its interaction with the helicase DDX60 to suppress the dsRNA induction of Type I interferon response involved in activation of NLRP3 inflammasome and cell death. NleG5-1 localizes to the host cell nucleus and targets the MED15 subunit of the Mediator complex which functions as a transcription co-activator and global regulator of gene expression including TGF-β. NleG6-1 function and interaction partners are yet to be elucidated. Interestingly, the three NleG effectors (G1, G7, and G8) of *C. rodentium_DB5100* have distinct roles during infection in mice involved in bacterial survival, expression of diarrheal symptoms, and accelerated lethality, respectively. These data indicate that depending on the host protein that NleG targets for ubiquitin-mediated degradation vastly
different effects could occur during infection. Further exploration is needed of NleG functions and targets in the context of the specific effector repertoire present in individual aEPEC strains as this will likely affect distinct downstream processes involved in aEPEC pathogenesis.

**Summary.** The majority of non-LEE effectors associated with symptomatic individuals were from IE6 of EPEC (NleB, NleE, and EspL2) which also houses LifA/Efa1,\textsuperscript{212} an adhesin discussed in further detail in “Adhesins Associated with Symptomatic aEPEC Infection” suggesting the entire insertional element has either been acquired or retained over time. Interestingly, *espJ* along with a Cif pseudo gene and *nleH1* are housed on PP2,\textsuperscript{212} however, different distribution patterns were observed among the three factors indicating different evolutionary trajectory paths. Interestingly, EspL2 may enhance pedestal formation while EspJ at later infection timepoints may regulate or suppress pedestal formation. NleE, NleB1, and NleG and homologues are anti-inflammatory by suppressing MAPK, NF-κB, TGFβ, inflammasome and apoptosis. Similar to most non-LEE effectors, NleE, NleB1, and NleG attenuate inflammation to counteract the pro-inflammatory mechanisms elicited by LEE effectors and ultimately enhance colonization. EPEC infection is characterized by the balance of pro- and anti-inflammatory mechanisms to allow enough disruption of IEC to provide nutrients for the bacteria but attenuated to suppress clearance by the immune systems. Further studies are needed to understand how NleE, NleB1, and NleG contribute to symptomatic aEPEC infection including machine learning models to understand the complexity of how these and other factors enhance or suppress infection or colonization.
Specific Non-LEE Effectors Associated with Tir Nodes of aEPEC Isolates

I detected several non-LEE effectors that were more prominent in certain Tir Nodes. These included NleA, NleF, NleC, NleD, Cif, EspV EspR, EspX, and EspY and will be described in more detail in the following sections.

**NleA and NleF.** In EPEC-E2348/69 *nleA* and *nleF* along with *nleH2* are harbored on PP6. All isolates from Tir Node 2 harbored NleA (100%) with reduced detection in Tir Node 4 (46.2%), Tir Node 1 (33.3%), and Tir Node 3 (25%). Similarly, nearly all isolates from Tir Node 2 harbored NleF (92.3%) with detection in about half of all isolates in Tir Nodes 1,3,4 (50-56.3%). NleH2 was detected equally among Tir Nodes (27-38%) indicating differential acquisition or loss along PP6. Loss of NleH2 is tolerated in most indicating it is not an essential factor for disease manifestation. In contrast, NleA and NleF detection in all or almost all of Tir Node 2 indicate they are essential in this genetic background. NleA was one of the essential virulence factors determined in the *C. rodentium* virulence networks study and is a multifunctional protein involved in disruption of ion transporters, TJ structure and function, and aberrant vesicle trafficking. In contrast, NleF inhibits caspase-4, 8, and 9 resulting in inhibition of apoptotic and pyroptotic death.

**NleC and NleD.** NleC and NleD are both zinc metalloproteases that target NF-κB subunits and p38 and JNK, respectively, thus promoting blockade of signaling cascades that would normally cause pro-inflammatory responses. NleC and NleD are housed on PP4_EPEC-E2348/69 along with *nleB2* and *nleG*. NleC and NleD have a similar detection pattern with the greatest abundance in Tir Node 2 (69.2% and 57.7% respectively), Tir Node 4 (42.3% and 32.7%), Tir Node 1 (both 27.8%) and Tir Node 3 (6.25% and 0%). NleG, NleG5-1, and NleG6-1 were associated with symptomatic status but only in a small number of isolates (22-32%) and are
E3 ubiquitin ligases involved in degradation of host proteins inhibition inflammasomes and other pro-inflammatory responses (see previous section NleG). Similarly, 5/9 of the EHEC NleGs were significantly associated with Tir Node and of these 4/5 had the greatest abundance in Tir Node 2 with an average 48.3% positive and only 11.5% for NleG6-2. In contrast, NleB2 was detected abundantly in all isolates based on Tir Nodes (56.3% -88.5%). NleB2 is a Arg-glucose transferase and similarly inhibits extrinsic apoptotic signaling via the FAS death receptor (see previous section NleB1).^149

**Cif.** In a study of 5,049 E. coli isolates, cycle-inhibiting factor (cif) was found in 115 strains, and the presence of cif and LEE were highly correlated.^284 However, 66% of cif-positive isolates were negative for cif-related phenotypes, including tEPEC_E2348/69 and EHEC_Sakai, due to mutations or N’ or C’ terminal truncations. Cif induces cell cycle arrest through inhibition of the ubiquitin/proteasome degradation pathway leading to accumulation of p21/p27 and inhibition of CDK-cyclin complexes, whose activation is needed for G1/S and G2/M transitions.^217 In addition, at 2 days post-infection, Cif induces apoptosis characterized by activation of caspases, accumulation of cleaved caspase-3, and LDH release.^217 I compared our 112 isolates to cif from tEPEC_B171 determined to produce full length, secreted and functional Cif.^284 Nearly all isolates from Tir Node 3 (81%) harbored cif and 54% of those in Tir Node 2 with less prevalence in Tir Node 1 (33%) and even less in Node 4 (20%). Further detailed studies are needed to determine if full-length, functional Cif is secreted and if cell-cycle arrest and apoptotic death contribute to aEPEC infection.

**EspV.** I determined that EspV had greater prevalence in Tir Node 2 (46.2%) compared to Tir Node 1 (33.3%), Tir Node 3 (19.2%) and Tir Node 4 (12.5%). espV was first discovered as a pseudogene in EHEC_Sakai, with full-length genes subsequently found in C. rodentium and
However, very little is known about the function of EspV. Prevalence of \( espV \) in 230 EPEC and EHEC isolates was examined with 16% (37/230) \( espV \)-positive, of those that were \( espV \)-positive 22% were aEPEC, 8% tEPEC and 10% EHEC.\(^{214}\) Ectopic expression of EspV and translocation by tEPEC_E2348 (\( espV \)-negative) into HeLa cells caused dramatic morphological changes in the host cells characterized by nuclear condensation, cell rounding and formation of dendrite-like projections.\(^{214}\) In addition, Baumgartner et. al. determined that the presence of aEPEC correlated with laboratory, clinical and endoscopic disease activity in UC, as well as microbiota dysbiosis. EspV-positive aEPEC isolates were associated with UC patients with active disease (higher stool calprotectin levels), harbored more genetic virulence factors, and stimulated higher levels of IL-8 secretion from human primary colon epithelial cells (HCEC-1CT) compared to those isolated from Crohn’s disease patients which lacked EspV.\(^{285}\) This could be a contributing factor to toxicity of epithelial cells or could elicit a greater immune response after infection with those isolates that harbor EspV. Context dependent studies are needed to understand the effects of EspV during aEPEC infection and its role in pathogenesis.

**EspR, EspX, and EspY.** The effector EspR comes from a family of 5 proteins in EHEC_Sakai first identified using mass spectrometry of secreted proteins, however, to date there are no known functions or described homologues of these proteins.\(^{286}\) The effector EspX comes from family of proteins with 7 members in EHEC_Sakai with no known function, however, they by definition are SopA-like homologues.\(^{216}\) SopA from *Salmonella* SPI-1 is required for invasion into host IEC and through its E3 ligase activity modulates two E3 ubiquitin ligases, TRIM 56 and TRIM 65.\(^{52,287}\) Opposing reports describe SopA action on TRIM56/65.\(^{287}\) Therefore, the exact functions of SopA in promoting invasion and inflammation remain unclear and may affect multifaceted or diverse cellular processes such as intracellular signaling,
development, apoptosis, protein quality control, innate immunity, autophagy, and carcinogenesis. The SopA-like homologues, EspX1,4,5, were prominent in most Tir Nodes except Tir node 4 especially absent in Tir_subtype10 isolates, but Tir Node 3 had the highest average number of any EspX (72.5%) compared to Tir Node 2 (64.6%), Tir Node 1 (48.9%), and Tir Node 4 (30.8%). It will be interesting to determine if EspXs have E3 ligase activity and what role they play in aEPEC infection.

The effector EspY comes from a family of proteins with 5 members in EHEC_Sakai that contain an N-terminal WEX5F domain with homology to Salmonella SopD protein of SPI-1. SopD cooperatively with SopB initiates "ruffling" and rearrangements of cellular actin promoting invasion of the bacteria. EspY3 is translocated through the T3SS where it localizes under attached bacteria and induces elongation of polymerized actin pedestals generating a significant increase in the size of the pedestal area. Isolates from Tir Node 1, 2 and 3 had the greatest prevalences of EspY1-5 with Tir Node 1 (27.8% – 38.9%), Tir Node 2 (3.8 – 46.2%), and Tir Node 3 (0 – 56.3%) in stark contrast to Tir Node 4 (0 – 3.8%). Out of any EspY present, Tir Node 3 and Tir Node 1 had the highest average number (30% each) followed by Tir Node 2 (19.2%) and Tir Node 4 (1.5%).

Previous studies of aEPEC isolates have not detected espR/X/Y. In contrast, I determined a high prevalence of espRI, espX1/X4/X5, and espY1 in Tir Nodes 1-3 (40-100%) and decreased prevalence in Tir Node 4 (4-52%) due mostly to the absence in isolates of Tir subtype10. Most of the isolates that contained all espR/X/Y genes, whether from asymptotic or symptomatic individuals, belonged to phylogroup E which encompasses common EHEC O157:H7 strains such as Sakai, EDL933, and TW14539 indicating prevalence in these isolates may be associated with inherited traits rather than virulence. Still, espR/X/Y are more prevalent in Tir Nodes
associated with symptomatic status so it will be interesting to determine how or if these contribute to aEPEC virulence.

**Summary.** As with other previously described non-LEE effectors, Cif, NleC, NleD, NleF, NleB2, NleH2 and NleG inhibit MAPK signaling, apoptosis, and inflammasome formation suppressing pro-inflammatory processes. Most were prevalent in Tir Node 2, followed by Tir Nodes 2 and 3, and Tir Node 4 consistently had the least number of non-LEE effectors. In contrast, many disruptive and pro-inflammatory non-LEE effectors, such as NleA, EspV, EspR, EspX, and EspY, were similarly associated with Tir Node 2 with the least detected in Tir Node 4 isolates. These data reiterate the notion that aEPEC pathogenesis, similar to tEPEC, involves pro-inflammatory/pro-disruptive effectors counterbalanced by anti-inflammatory effectors. However, more studies are needed to understand how the specific combination of pro- and anti-inflammatory effectors in different genetic backgrounds affect virulence and clinical manifestations. In depth neural networking analysis would aid in the detection of genetic and virulence networks in different Tir Node and subtypes and how they contribute to aEPEC pathogenesis.

**Adhesins Associated with Symptomatic aEPEC Infection**

In a study of 71 aEPEC strains isolated from children with diarrhea (n=54) and healthy individuals (n=17) in Brazil and Australia, prevalence of several fimbrial genes was assessed through PCR. The most prevalent fimbrial genes found, in descending order, were hcpA (98.6%), ecpA (86%), fimA (76%), elfA (72%), and csgA (19.7%).\(^\text{188}\) Despite the high prevalence of ecpA detection, only 36.6% of strains produced *E. coli* common pilus (ECP) under the experimental conditions but strains from symptomatic cases tended to produce ECP over asymptomatic control strains, though not significantly.\(^\text{188}\) Also, strains produced several different pilus types
simultaneously during infection of HeLa cells\textsuperscript{188} indicating multiple adhesins likely contribute simultaneously to aEPEC adherence. In addition, studies of EPEC adherence indicate that multiple adhesins contribute to attachment but few are absolutely required when others are present indicating a multifactorial mechanism of adherence.\textsuperscript{9,149}

In my study, I detected a plethora of different adhesins including those designated as adhesins, autotransporter (AT) adhesins, fimbrial protein genes, and genes associated with the }\textit{fim} operon among other non-descript fimbrial gene designations. Overall, higher number of fimbrial protein genes correlated with greater \textit{in vitro} phenotypes, and the number of AT adhesins weakly correlated with #BM$s$/day. I also determined that two 2 adhesins, \textit{paa} and \textit{lifA/efa1}, were associated with symptomatic status but only present in \textasciitilde40\% of SYMP isolates. Despite the lower frequencies, adherence was higher in those isolates in which either LifA\_IE2 or IE6 was detected. In contrast, adherence was not significantly higher in isolates in which \textit{paa} was detected, likely due to other mechanisms of adherence that are present such as LifA or other fimbrial proteins. Similarly, studies have found that \textit{paa} and \textit{lifA/efa1} are associated with diarrheal cases of aEPEC infection.\textsuperscript{186,187,269} LifA/Efa1 contributes to EPEC and EHEC epithelial cell adherence \textit{in vitro}.\textsuperscript{146} Through mutations in glycosyltransferase and protease motifs of LifA/Efa1 it was determined that LifA/Efa1 is required for intestinal colonization of mice by the related A/E pathogen \textit{C. rodentium}.\textsuperscript{148} LifA also inhibits the proliferation of mitogen-activated lymphocytes and the synthesis of proinflammatory cytokines.\textsuperscript{290} More studies are needed of LifA/Efa1 to understand if its requirement \textit{in vivo} is related to reduction of inflammation or adherence and if LifA/Efa1 is a structural adhesin on the outer membrane of the bacteria. In contrast to LifA/Efa1, the adhesin Paa from porcine O45 \textit{E. coli} strain 86-1390 made direct contact with the epithelium of ileal explants and induced A/E lesions which were detected
earlier than onset of diarrhea in inoculated piglets.\textsuperscript{291} In addition, Paa protein was detected with TEM to be uniformly distributed over the bacterial surface indicating Paa is an adhesion involved in the early steps of infection.\textsuperscript{291} More studies are needed to understand if LifA and Paa have similar roles in adherence of our aEPEC isolates. In view of the multiple adhesins detected in my study and in previous studies, aEPEC adherence is a multifactorial mechanism involving many adhesins that contribute attachment and one single factor does not determine adherence. Better \textit{in vitro/ex vivo} models that recapitulate the native host intestine during infection are also needed to understand the role these adhesins have in aEPEC infection.

**Emerging Strains of Hybrid Pathogenic \textit{E. coli}**

A recent review highlighted the growing number of hybrid pathogenic \textit{E. coli} (HyPEC), described as a new combination of virulence factors among classic \textit{E. coli} pathotypes.\textsuperscript{93} At least 7 new HyPEC strains have been reported in the literature, notably 3 involved EPEC. Two reported hybrids of EPEC (both typical and atypical) and ETEC harboring intimin and LEE genes with ST and LT toxin genes. Four EPEC/ETEC hybrids were isolated from the GEMS childhood diarrheal study, one tEPEC/ETEC strain originated from a lethal outcome of diarrheal disease, two aEPEC/ETEC strains were from non-lethal symptomatic cases, and one aEPEC/ETEC was from an asymptomatic individual.\textsuperscript{292} Phylogenetic analysis revealed these strains were more genomically related to EPEC and had acquired plasmids which harbored ETEC toxin genes.\textsuperscript{93} An EPEC/ETEC strain isolated from India harbored intimin (\(\beta\) subtype), other LEE encoded genes, and the \textit{elt} operon encoding LT housed on a conjugative plasmid; this isolate came from a 6 month old child which presented with acute, watery diarrhea that contained traces of blood with 6–7 episodes/day, mild fever and cough.\textsuperscript{293} In addition, a hybrid of tEPEC
and STEC which harbored intimin, BFP, and Stx2 was isolated from pet birds in Brazil suggesting a zoonotic risk.\textsuperscript{294}

Similarly, I determined that the isolate OB137, from a healthy individual devoid of all gastrointestinal symptoms, housed the LEE with overall high homology to tEPEC/EHEC-LEE and housed genes associated with aggregative adherence ($aggA,B,C,D$). OB137 displayed robust attachment, a very clear aggregative pattern, lifted cells after prolonged infection times, and although pedestals were not observed, caused disturbances in the brush border of SKCO-15 cells (data not shown). OB137 also had an average number of adhesins/fimbrial protein genes but lacked nearly all non-LEE effectors. Is the lack of non-LEE effectors or clear pedestal formation enough to inhibit disease manifestations in a healthy individual? Would this be the case for someone with intestinal dysbiosis? Or are there host-microbiota factors that suppress disease manifestation by OB137? Interestingly, OB137 was isolated from a donor (Donor M) which also housed OB136.1 and OB2 simultaneously (isolates originated from stool samples collected within 0-1 day of each other). OB137, OB136.1 and OB2 clustered in Q4 of the PCA_B suggesting they are less virulent, however, they clustered in different locations in Q4 indicating different genetic backgrounds. Similarly, four other instances occurred in which OB isolates originated from the same donor with concurrent colonization (as indicated by isolates originating from stool samples collected within 0-8 days of each other). Interestingly, OB84/86 were similar genetically and phenotypically, and clustered in Q3, indicating an intermediate virulence potential, and was concurrently detected with OB38/15/33 which clustered in Q4 (least virulent). This raises the question of whether bacterial competition within a species prevents disease manifestation, i.e. are disease manifestations suppressed by OB136.1/OB2 inhibiting OB137 or OB38/15/33 inhibiting OB84/86? Shotgun metagenomic sequencing analysis studies of these and
future samples are needed to understand the concurrent relative abundance of these isolates within the stools. *In vitro* and *in vivo* bacterial competition assays would help elucidate fitness characteristics of these isolates and how they influence disease manifestation of aEPEC infection.

**Possible Bacterial and Host Factors Which May Contribute to aEPEC Virulence**

Many different bacterial factors could influence disease manifestation of aEPEC infection as discussed throughout but were beyond the scope of this study. Highlighted here are a few that need further attention. The effects of aEPEC on ion transporters/exchangers needs to be examined due to the known mechanisms employed by EPEC, EHEC and other *E. coli* pathovars to disrupt electrolyte secretion and absorption as reviewed in previous sections of this study.

In addition, regulatory mechanisms involved in LEE and non-LEE effector and adhesin expression are needed as different expression levels of intimin, Tir, EspA, EspB, and EspD were evident in different aEPEC stains and different when incubated with cells. Other factors which could influence aEPEC pathogenicity are the ability to form biofilms or produce toxins. In my assessment of adhesins, the AT adhesin genes *ehaB/J* involved in biofilm formation were detected and *ehaB* had greater prevalence in Tir Node 4 isolates but was not associated with symptomatic status. It is interesting to speculate that biofilm formation may be protective against host defense mechanisms rendering Tir Node 4 isolates less virulent or “hidden” from the host. In addition, *toxB* also involved in biofilm formation was not detected in any strain. At least two other biofilm genes (*saa, sab*) as well as NleB/E/H are known to play a role in EHEC biofilm formation warranting further investigation into these and others.

Intestinal tropism of aEPEC isolates likely plays a considerable role in virulence as extrinsic signals from the environment and microbiota dramatically changes along the GI tract.
and the pathophysiology and severity of diarrhea vary depending on the segment of bowel affected and dictates strategies for treatment.\textsuperscript{27, 31} tEPEC infect the small bowel\textsuperscript{149}, and EHEC preferentially infect Peyer’s patches and colonic epithelia\textsuperscript{295}, however, the preferred attachment site for aEPEC is unknown and may differ for various isolates. In a study of enteroaggregative \textit{E. coli} (EAEC) infection and adhesion, different human enteroid/colonoid lines were infected with EAEC and unique patterns of intestinal segment- and donor-specific adherence were found. Interestingly, adherence pattern and level were regulated by different mechanisms.\textsuperscript{296} This study exemplifies the need for studying tropism along different intestinal segments as well as from different individuals when considering tropism and pathogenic mechanisms of aEPEC.

Colonization resistance suppresses growth of invading pathogens or pathobionts through many direct and indirect interactions, including competition for nutrients, production of bacteriocins and other antimicrobials by the gut microbiota, and promotion of host immune defenses.\textsuperscript{297} For example, a recent study determined that commensal \textit{E. coli} strains produce SCFA\textsuperscript{64} and growth of certain commensal microbiota is suppressed in the presence of SCFA and amino acids\textsuperscript{60} indicating competition for niche between commensal bacteria. SCFA promote EHEC expression of \textit{iha} (encoding a non-fimbrial adhesin) and stimulate production and functionality of flagella, indicating EHEC promote adherence in the presence of certain bacterial metabolites.\textsuperscript{295} In addition, during disease states, there is increased competition for niche and nutrients. For instance, the CD-associated \textit{E. coli} strain LF82 shifts metabolism away from simple sugars present in a healthy gut and instead uses L-serine when inflammation is present giving LF82 competitive fitness over commensal bacteria.\textsuperscript{298} Phylogroup B2 \textit{E. coli} strains often prevalent in IBD patients have distinct metabolic genes versus the more transient phylogroups A and B1 indicating the role of metabolism contributing to intestinal mucosa colonization.\textsuperscript{299} Even
among strains of *E. coli* in B2 phylogroup, diversity exists between the metabolome of the probiotic strain EcN and pathogenic *E. coli.*

In addition, many studies have examined how pathogens, particularly EHEC, have overcome colonization resistance using niche factors which promote virulence, such as microbiota cross-feeding, or factors involved in metabolism or oxidative stress response. For instance, the human gut commensal *Bacteroides thetaiotaomicron* can exacerbate EHEC infection by increasing metabolites involved in gluconeogenesis which are sensed by the virulence transcription factor Cra and upregulate LEE proteins and Stx toxin expression. Under these conditions, EHEC was also shown to provide a distinct growth advantage to *B. thetaiotaomicron.* Similarly, EHEC strains take advantage of a nitrogen source in the bovine intestine that is not consumed by the resident microbiota suggestive of an ecological niche favoring EHEC persistence. Also, in a mouse model of infection and competition, EHEC used sugars not consumed by the commensal strain MG1655 suggesting EHEC can overcome nutrient competition with resident microbiota. EHEC also gains energetic advantage over resident microbiota by altering its metabolism to utilize microbiota-derived vitamins such as biotin and down-regulates its own *de novo* synthesis.

In the presence of gut microbiota metabolites, stress response genes were upregulated in EHEC, such as oxidative and envelope stress response genes, *soxS* and *cpxP* and *spy*, and stress-induced genes of unknown function *bhsA* and *yhcN.* Although type VI secretion systems (T6SS) are typically involved in direct bacterial-bacterial killing by injection of effectors, in EHEC, the T6SS secretes a catalase promoting survival of EHEC in macrophages by neutralizing reactive oxygen species. EPEC strain B171 harbors 18 T6SS orthologs, but their function is unknown. The pathogenic mechanism of *Plesiomonas shigelloides,* a bacterial enteric
pathogen that causes moderate to dysenteric diarrhea often due to consumption of contaminated seafood, is likely due to secretion of putative virulence factors and a cytotoxin via a T6SS.\textsuperscript{44,307} Interestingly 50-60\% of our aEPEC isolates house at least two T6SS genes warranting further detailed genetic and functional analysis of the T6SS.

The transmissible locus of stress tolerance (tLST) confers resistance to heat, pressure, chlorine and other oxidizing agents in \textit{E. coli} and is harbored mostly in strains belonging to phylogroup A and to a lesser extent phylogroup B and C.\textsuperscript{308,309} Although rare, tLST is associated with different MLSTs, such as STc10 (associated with different EPEC strains) and ST301 (associated with hybrid DEC-ExPEC strains).\textsuperscript{309} Interestingly, 13 of our aEPEC isolates were designated STc10\_phyloA with 2 from Tir Node 1\_ST1 (CE245,180), 8 from Tir Node 2\_ST3 (CE347,318,103, 166,376,444*,OB84,OB86), and 3 from Tir Node 3\_ST6 (CE165,13,243). In addition, one isolate (CE126*) was designated ST301\_phyloA from Tir Node 1\_ST2. Most isolates from STc10 clustered in quadrant 3 of PCA-B (62\%) with some clustering in Q1 or Q2. Further investigation is warranted to determine if these isolates harbor tLST as resistance to chlorine may contribute to persistence in nosocomial related infections.\textsuperscript{308}

In conclusion, EHEC is clearly able to colonize the human gut even in the presence of microbiota due to competitive growth or stress response systems among other factors. Bacterial niche and fitness characterization of our aEPEC isolates were beyond the scope of this study. Further detailed genomic analysis and functional assays of our aEPEC isolates are needed to determine if they harbor similar metabolic, colonization, and niche factors such as those found in EHEC or other pathobionts. One can speculate that if these EHEC colonization factors which promote virulence were mutated or absent, aEPEC may transiently pass through the gut but may not be able to cause infection resulting in asymptomatic detection. Conversely, acquired diverse
metabolism or niche factors may cause imbalances in homeostasis possibly resulting in an inflammatory environment which could promote aEPEC virulence. These data indicate the necessity of exploring the metabolic capabilities of individual strains of aEPEC and their specific effects during infection.

Summary

In conclusion, I determined that individuals who were GIP-EPEC-positive had greater likelihood of developing diarrhea, vomiting, abdominal pain, and fever. Most patients experienced mild to moderate diarrhea, but symptoms varied greatly with some individuals having up to 10-40 bowel movements/day and some had persistent/chronic diarrhea. Age and co-infection status were not major contributors to diarrheal symptoms. Fever and vomiting were more prominent in younger individuals whereas abdominal pain was more evident in middle aged individuals. Interestingly, those aged 65 years or older tended to have more asymptomatic carriage and less severe diarrhea as assessed by the modified Vesikari score. Out of all EPEC-positive cases, asymptomatic carriage was detected at a relatively high rate of 29%. Our study represented the first in-depth characterization of the clinical manifestations of aEPEC infection in a primarily adult population from the US and provides insight into the range and variability of symptoms associated with this pathogen. aEPEC pathogenicity in adults has been a topic of long-standing debate but our study indicates that aEPEC is associated with a wide array of symptoms in adults, ranging from asymptomatic carriage to severe diarrhea. Clinicians should not ignore EPEC-positive results but rather base treatment decisions on the severity of symptoms.

I also determined that bacterial load was higher in symptomatic individuals but did not predict symptomatic status or diarrhea severity indicating that genetic virulence factors likely contribute to disease manifestations. We purified and characterized 112 isolates originating from
our clinical cohort and from stools of healthy donors with no preexisting conditions to characterize their genetic and in vitro phenotypic virulence factors. I determined that our clinical study was reflective of sporadic aEPEC infections in children and adults.

Due to the genetic and phenotypic heterogeneity of aEPEC strains, pathogenic mechanisms of aEPEC infection have yet to be elucidated. However, by coupling genetic and in vitro phenotypic virulence factor analysis of aEPEC isolates with their corresponding clinical manifestations, I was able to discern two possible emerging aEPEC pathogenic mechanisms. The first potential aEPEC pathogenic mechanism is based on the majority of our isolates and is most similar to tEPEC pathogenesis in that it involves pedestal formation and robust adherence due to a greater number of virulence factors likely compensating for the lack of BFP. In these isolates, aEPEC virulence was associated with reduced homology to reference strains tEPEC_E2348/69 and EHEC_Sakai in the effectors Tir, EspH, and EspF, and the translocon protein EspB as well as increased homology to the needle-protein EspA and increased numbers of adhesins and non-LEE effectors. Interestingly, certain Tir phylogenetic groups designated as Tir Nodes contained greater numbers of isolates from symptomatic individuals. This genetic profile corresponded to greater adherence properties and increased pedestal formation and was associated with more robust disease manifestations such as higher number of bowel movements per day, longer duration of diarrhea, and greater diarrhea severity scores. Isolates from asymptomatic individuals lacked most of these virulence factors indicating a clear distinction between avirulent and virulent strains. To understand this mechanism in detail, future studies should explore the specific effects of EspA and EspB mutations and if these changes affect needle formation, initial attachment, and/or effector delivery. In addition, changes in Tir need to be explored to understand the effects on translocation, insertion into the host membrane, Tir-intimin interactions during intimate
adherence, and downstream effector functions such as pedestal formation and signal transduction. Functional analysis of EspB, EspF, and EspH should also be explored for known contributions to pedestal formation and actin dynamics, disruption of ion and solute transporters, TJ structure and function, mitochondrial integrity, ER stress, and suppression of NF-κB and MAPK signaling. My data indicate a multi-factorial pathogenic mechanism; however, future studies should employ linear regression or machine learning modeling with greater numbers of isolates to discern the contribution that each specific genetic virulence factor makes to aEPEC pathogenicity.

A subset of our isolates did not fit the multifactorial pathogenic mechanism based on greater virulence factors and reduced homology in certain effectors but rather represent a second group that likely have unique mechanisms based on individual isolate characteristics. In general, these isolates have high homology to reference strains tEPEC_E2348/69 and EHEC_Sakai in Tir, EspB, EspH, and EspF, but have reduced EspA homology and the lowest numbers of adhesins and non-LEE effectors. Correspondingly, these isolates had low adherence properties and varied pedestal formation, but all originated from symptomatic individuals with fairly robust disease manifestations. This subset of unique aEPEC isolates originates from phylogroup B2 defined by tEPEC-E2348/69 and exemplifies the loss and gain of virulence factors affecting isolate pathogenicity. These data indicate that many aEPEC isolates are “primed” to be virulent as they house the entire locus of enterocyte effacement pathogenicity island containing the type 3 secretion system. The variability in clinical manifestations is likely dependent on the presence/absence of specific genetic virulence factors. Future studies are needed to assess intestinal segment tropism of aEPEC isolates and their effects on ion/solute transporters to fully understand the pathogenic mechanisms that cause diarrheal disease. Other studies should include
use of a humanized murine model to assess the effects of microbiota and microbial metabolites on aEPEC pathogenic mechanisms.

In conclusion, aEPEC are a diverse group of bacteria that are associated with a range of variable clinical manifestations. Therefore, aEPEC detection should not be dismissed when considering the etiological agent of diarrheal disease in children and adults. Although my study has begun to elucidate potential pathogenic mechanisms of aEPEC, further investigation is sorely needed to understand this diverse group of organisms, to develop better diagnostic tools, and to develop better treatment plans for those infected with aEPEC.
REFERENCE LIST


VITA

The author, Sarah Kralicek, was born in Lombard, IL on August 21st, 1979, to John and Charlotte Knopf. She attended Benedictine University in Lisle, Illinois where she earned a Bachelor of Science, cum laude, in Molecular Biology in May 2002. After graduation, Sarah worked full time in the laboratory of David Hecht, MD while pursuing her Master of Science degree at Loyola University Chicago where she was mentored by Gayatri Vedantam, PhD. Sarah’s thesis work involved the biochemical characterization of the Bacteroides fragilis conjugation-specific proteins BmpH and BctA. Upon graduation with honors, Sarah continued to work at Loyola as Lab Supervisor with Dr. David Hecht and later with Dr. Gail Hecht. In August 2020, Sarah matriculated into the Loyola University Chicago Stritch School of Medicine Integrated Program in Biomedical Sciences and began her graduate education in the Biochemistry, Molecular, and Cancer Biology Program under the mentorship of Gail Hecht, MD. Sarah’s dissertation work on the genetic, in vitro phenotypic, and clinical characterization of atypical EPEC infection and pathogenesis was supported by National Institute of Allergy and Infectious Diseases (NIAID) at the National Institutes of Health R21-AI142515 and R24-AI118629. After completion of her graduate studies, Sarah will pursue a Postdoctoral Fellowship at Midwestern University in Downers Grove, IL in the lab of Dr. Kristina Martinez-Guryn.