Genetic Changes Leading to Antibiotic Resistance in Clinical Isolates of *Escherichia coli*

by

Prabhjeet Basra

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Science

in

Biology

Carleton University
Ottawa, Ontario

Abstract

The spread of antibiotic resistance has limited the use of antibiotics in clinical settings, resulting in a huge threat to public health worldwide. Multidrug resistance causes thousands of deaths annually and is currently a significant financial burden in many countries. In this study, clinical isolates of *Escherichia coli* with varying drug resistance profiles were used to understand the genetic changes contributing to quinolone and *β*-lactam resistance. Novel genetic changes in *gyrA* and *gyrB* genes, known to contribute to quinolone resistance were recorded. Furthermore, the CTX-M-14 and -15 genes were found in most *β*-lactam resistance strains. Significant correlations between mutations in novel genes and various phenotypic traits were also observed, suggesting a role for these genes in determining these phenotypes. This work further increases our knowledge regarding causes of antibiotic resistance in clinical strains and lays a foundation for future research.

Acknowledgements

I would like to extend my sincere gratitude to my supervisor Dr. Alex Wong, for his encouragement, guidance and patience in the past two years. He provided an inspiring environment, encouraging me to constantly challenge myself and think outside the box. I would like to thank my committee members, Dr. Nicolas Rodrigue and Dr. Jessica Forrest, for their guidance and thought provoking discussions. I would also like to extend my deepest gratitude towards Dr. Sylvain Pitre and Dr. Andrew Schoenrock for their help and patience in running CoEvol. Special thanks to Gabriela Bernal-Astrain and Andrew Low, for sharing growth rate data and providing some key scripts respectively. A heartfelt thank you to the members of the Wong Lab (Kamya, Ahlam, Trevor, Nicole, MK, Tom, Andrew, Gaby and Jess) for always being supportive and helpful. It's been great working with you all.

I would like to thank my friends and family, who have always been there for me and believed in me. Finally I would like to thank the two people without whom I couldn't have done this; my beautiful sister Ishwar and my amazing husband Jaspal. You two have been my constant support in keeping me focused on my goals.

Table of Contents

List of Tables

List of Figures

1 Introduction

Antibiotics are a group of compounds often isolated from microorganisms that either kill or inhibit bacterial growth by interfering with important metabolic processes. Their discovery and clinical use is one of the most important advancements in modern medicine (Andersson & Hughes, 2010; Tenover, 2006). Antibiotics were major contributors in increasing life expectancy in the mid 1900s by providing treatment for many diseases that were untreatable in the pre-antibiotic era. Over 1 million tonnes of antibiotics have been used since their development in the 1940s and have contributed extensively to the treatment of infectious diseases and to reductions in childhood mortality (Andersson & Hughes, 2010; Blair et al. 2014). With the introduction of antibiotics, deaths due to infectious diseases in the United States declined by 8.2% annually between 1938 and 1952 (Cohen, 2000). The most prominent effects were seen in patients suffering from tuberculosis and pneumonia, which were once thought to be untreatable. Similarly, deaths due to childbed fever, caused by *Streptococcus pyogenes*, declined by 50% following the introduction of the antibiotic sulphadiazine (Cohen, 2000). The end of twentieth century saw mortality from infectious diseases being replaced by mortality due to chronic diseases as a major public health concern.

However, this success was short-lived with the re-emergence of diseases like tuberculosis and cholera in the late twentieth century (Cohen, 2000). Since their initial success in clinical settings, antibiotics have become less and less effective due to the spread of antibiotic resistance worldwide, limiting their clinical application and resulting in a growing threat to public health in both industrialized and developing countries (Palmer & Kishony, 2013).

Major organizations like the Centers for Disease Control (CDC), the World Health Organization (WHO), and the European Center for Disease Prevention and Control (ECDC) have all recognized multidrug resistance (MDR) as an emergent global issue, with MDR bacteria infecting over 2 million people and causing 23,000 deaths annually in the USA (Blair et al. 2014; Roca et al. 2015). MDR is not only a health issue but also a financial burden, with increasing numbers of patients requiring hospitalization. MDR costs the European Union 1.5 billion euro annually and was cited as the greatest threat to humans in a recent World Economic Forum Global Risks report (Blair et al. 2014).

Antibiotic resistance is not a recent phenomenon. Clinically, resistance was first seen in the 1930s with the emergence of sulphonamide resistant *S. pyogenes* (Levy, 1982). Initially, a low level of resistance was observed and infections could be treated with larger doses of antibiotics, but high levels of resistance soon followed. By the 1950s, penicillin-resistant *Staphylococcus aureus* were causing serious problems in hospitals throughout the world (Finland, 1955). MDR was first seen in enteric bacteria (*Escherichia coli*, *Shigella* and *Salmonella*) in the late 1950s and early 1960s (Levy & Marshall, 2004). However, many recent studies have confirmed the hypothesis that resistant bacteria in fact predate the clinical antibiotic era. For example, the discoveries of resistant bacteria in the gut of a thousand year-old Peruvian mummy, in the Lechuguilla cave in New Mexico which has been isolated from the outside world for 4-7 million years, and in 30 000 year-old permafrost in the Yukon, all show evidence that resistance evolved pre-clinically, presumably in response to antibiotics encountered in the natural world (Bhullar et al. 2012; D 'costa et al. 2011; Santiago-Rodriguez et al. 2015).

1.1 Mechanisms of Resistance

The baseline ability of a bacterium to withstand antibiotics, as a result of inherent structural or functional characteristics, is referred to as "intrinsic resistance". Bacterial groups differ in their intrinsic resistance to different drugs – for example, penicillins are more effective against Gram-positive bacteria than Gram-negative bacteria due to the composition of the cell wall (Blair et al. 2014). The ability of bacteria to evolve resulting in protection against antibiotics, is referred to as "acquired resistance". It is the rapid spread of acquired resistance that is a worldwide concern. Bacteria acquire resistance by many mechanisms but these can be grouped into three main categories: bacteria can (1) alter their cellular membrane to decrease intracellular concentration of the drug, (2) alter the target of the drug, or (3) produce enzymes to inactivate or hydrolyze the drug (Blair et al. 2014; Levy & Marshall 2004; Palmer & Kishony 2013). Acquired resistance is gained by bacteria via two methods: *de novo* chromosomal mutations or horizontal gene transfer (HGT) (Andersson & Hughes 2010; Blair et al. 2014; Levy & Marshall 2004). Most commonly, chromosomal mutations cause resistance by altering the drug target, or by increasing efflux pump expression. Such mechanisms are common for rifampicin and fluoroquinolones (MacLean et al. 2010; Andersson and Hughes 2010). HGT, comprising of conjugation, transformation and transduction, is typically associated with mechanisms like drug modification and acquisition of novel efflux pumps, causing resistance to drugs like β-lactams (MacLean et al. 2010; Pitout et al. 2005). Even though *de novo* mutations and HGT are the most common mechanisms for acquiring resistance, the role of recombination cannot be over looked. Recombination unlinks adjacent genomic regions;

this complicates evolutionary histories but is required to carry out genotype-phenotype associations (Dettman et al. 2013). However, in this study we will focus mainly on *de novo* mutations and HGT.

1.2 Fluoroquinolones

Fluoroquinolones are broad-spectrum synthetic antibiotics that have been used widely in both hospital and community settings. The first quinolone, nalidixic acid, was introduced for clinical use in 1962. Subsequent modifications to molecular structures have led to 2nd- and 3rd-generation quinolones, such as ciprofloxacin and levofloxacin (Fèbrega et al. 2009). Quinolones function by targeting DNA gyrase and Topoisomerase IV, with gyrase being the more important target in Gram-negative bacteria such as *E. coli* (Fèbrega et al. 2009). DNA gyrase introduces negative supercoils into DNA, and unwinds positive supercoils, during diverse cellular processes, including replication and transcription (Hooper, 1999). Gyrase's activity requires the formation and re-ligation of a double-strand break (DSB) in DNA; quinolones bind to the gyrase-DSB complex, preventing re-ligation, ultimately leading to cell death (Drlica & Zhao, 1997).

Major mutations contributing to clinical quinolone resistance have been described in *E. coli* and other important pathogens. Chromosomal mutations in the *gyrA* and *gyrB* genes are the most common causes of resistance. These mutations result in conformational changes in the A or B subunits of DNA gyrase, respectively, altering the active site and hence preventing binding of quinolones to the gyrase enzyme (Bagel et al. 1999; Hopkins et al. 2005). In addition to these target gene mutations, alterations in *marR* also contribute to quinolone resistance in *E. coli*. The *marR* gene encodes a repressor of

the *marRAB* operon. Mutations that reduce binding of the MarR protein to the operator region of the operon result in increased expression of *marA.* The MarA protein in turn regulates many other genes; importantly, it up-regulates *acrAB-TolC* and *micF. acrAB-TolC* encodes an efflux pump capable of removing quinolones from the cell, while upregulation of *micF* decreases the production of OmpF porins, an important point of entry for quinolones. Thus, loss-of-function mutations in *marR* lead to a decrease in intracellular quinolone concentrations (Bagel et al. 1999; Hopkins et al. 2005). Lastly, plasmids carrying the *qnr* quinolone resistance genes also contribute to resistance against these drugs. The *qnr* genes encode pentapeptide repeat proteins (PRPs), which bind to the gyrase-quinolone complex causing conformational change in the gyrase. This conformational change results in the release of the quinolone from the gyrase-quinolone complex, allowing for the DSB to be re-ligated and providing resistance against the drug (Blair et al. 2014; Vetting et al. 2011).

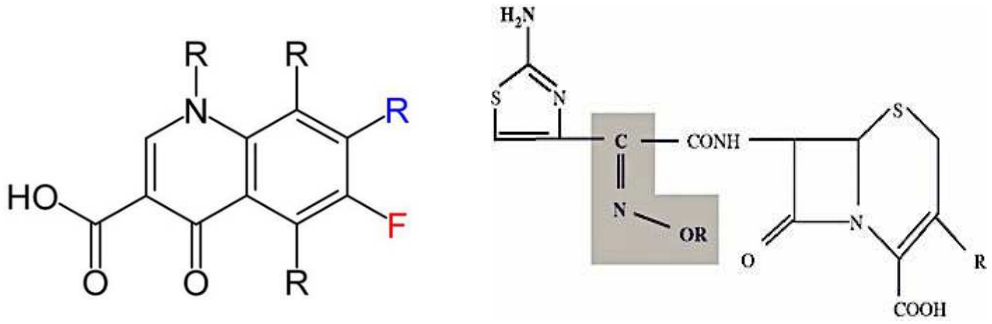

Figure 1: General structure of fluoroquinolones (left) and *β*-lactams with oxyimino side chain (right).

1.3 *β*-lactams

β-lactams were one of the first classes of broad-spectrum antibiotics to be widely used in clinical settings. This large drug class includes penicillin and ampicillin, the carbapenems, cephalosporins, and monobactams. *β*-lactams inhibit cell growth by binding to penicillin-binding-proteins (PBPs), which are cell wall synthesizing enzymes (Bycroft & Shute, 1985; Yotsuji et al. 1988). PBPs catalyze the D-ala D-ala cross linkages of the peptidoglycan wall that surrounds the bacterium. Binding of drug to these proteins results in weakening of the cell wall and inhibition of growth, eventually leading to cell death (Bycroft & Shute, 1985; Yotsuji et al. 1988). In Gram-negative bacteria *β*-lactamases, enzymes that hydrolyze *β*-lactams and render them ineffective, are the most common causes of resistance to *β*-lactams (Pitout et al. 2005). These enzymes are often encoded on plasmids that are acquired by HGT and are associated with the rapid spread of resistance. Even though plasmid-encoded *β*-lactamases are most common, resistance to *β*-lactams can also occur by chromosomal mutations. AmpC enzymes are chromosomally encoded and are inducible *β*-lactamases present in the Enterobacteriaceae. Mutation in the *ampD* or *ampR* genes, regulators of *ampC*, can lead to overexpression of AmpC and confer resistance to extended-spectrum cephalosporins. AmpC enzymes can also be located on plasmids but are usually constitutively expressed (Hanson, 2003; Jacoby, 2009). Over 400 different *β*-lactamases have been identified and the number continues to increase. Extended-spectrum *β*-lactamases (ESBLs), first described in 1983, are a rapidly evolving group of *β*-lactamases that share the ability to hydrolyze third-generation cephalosporins containing the oxyimino side chain (Pitout et al. 2005). In the 1990s, the majority of ESBLs seen in clinical settings were mutants of the TEM (capable of hydrolyzing penicillins and first generation cephalosporins) or SHV

(similar spectrum of activity to that of the TEM-1 β-lactamase, but it achieves better activity against ampicillin) types, which have evolved from narrow spectrum β-lactamases. However, in recent years, CTX-M β-lactamases (active on CefoTaXime, first isolated in Munich) have become more prevalent in clinical settings (Hawkey & Jones, 2009; Pitout et al. 2005). Over 40 different CTX-M β-lactamases have been identified worldwide and are grouped into five clusters based on their amino acid composition (Hawkey & Jones, 2009; Pitout et al. 2005).

1.4 Areas of Further Research

Even though major changes resulting in ciprofloxacin resistance are known, there is reason to believe that these known loci are not solely responsible for adaptation to quinolone antibiotics. First, several studies have identified mutations in additional loci that can result in modest changes in quinolone resistance, in *E. coli* and other bacterial species (Breidenstein et al. 2011; Tamae et al. 2008; Wong et al. 2012). For example, in a screen of the *E. coli* KEIO knockout collection, Tamae et al. (2008) identified 19 mutants with increased ciprofloxacin sensitivity, most of which had not been previously associated with quinolone susceptibility and resistance. Similarly, Breidenstein et al. (2011) found that a very large collection of genes, from different functional categories, contribute to both intrinsic and acquired resistance in *Pseudomonas aeruginosa*.

Second, adaptation to an antibiotic may not occur solely via increases in levels of resistance. A number of resistance mutations result in a cost of resistance, whereby the ability to grow in the absence of antibiotic is compromised (Andersson & Hughes 2010; Melnyk et al. 2015). Compensatory mutations – mutations elsewhere in the genome that

restore competitive ability without completely eliminating resistance – are expected to arise, and indeed do so readily during laboratory selection (e.g., Kugelberg et al. 2005). Costs of resistance, and associated compensation, can be measured easily in the lab through the use of competitive fitness assays or using pure culture growth curves. In the latter case, the fitness costs of resistance mutations may decrease growth rates, reduce population density, and/or increase lag phase (the time taken for a culture to start growing exponentially). Thus, in addition to known mediators of high-level resistance, I expect that additional resistance mutations, as well as compensatory mutations, will be selected under antibiotic pressure.

1.5 Current Work

In this thesis I set out to identify the genetic basis of quinolone and *β*-lactam resistance in a collection of pathogenic *E. coli*. I will be using thirty-nine *E. coli* strains having different drug resistance profiles (quinolone susceptible, quinolone resistant, quinolone susceptible ESBL positive and quinolone resistant ESBL positive) isolated from patients in Canada. Specifically,

1. I will assess the contributions of mutations in *gyrA*, *gyrB* and *marR* towards quinolone resistance, and that of *β*-lactamases towards *β*-lactam resistance.
2. I aim to identify potential novel contributors to quinolone resistance.
3. I will assess the genes gained and lost during the evolution of these strains.
4. I will evaluate the correlation between population structure and location of isolation.

2 Material and Methods

2.1 Bacterial Isolates

Thirty-nine clinical isolates of *E. coli*, collected as part of the CANWARD survey of antibiotic resistant pathogens in Canada (Lagacé-Wiens et al. 2013), were obtained from the Zhanel laboratory at the University of Manitoba. These isolates were collected from patients at hospitals across Canada, from a variety of non-gastrointestinal infection types, and represent a wide range of quinolone and *β*-lactam sensitivities (Appendix, Table 3). Additionally the laboratory strain *E. coli* K-12 (MG1655; NC_00913), the enterohaemorrhagic strain *E. coli 0157* (BA000007.2), and the outgroup species *E. fergusonii* (NC_011740.1) were added to the dataset.

2.2 Phenotypic Data

Quinolone and *β*-lactam resistance were measured using minimum inhibitory concentration (MIC) assays (Andrews, 2001). MICs were measured for four drugs: the quinolone ciprofloxacin, and the *β*-lactams ampicillin (a penicillin), meropenem (a carbapenem), and ceftazidime (a 3rd generation cephalosporin). The use of these three *β*-lactams enables us to distinguish between different types of *β*-lactamase, since only ESBLs will display high level resistance to ceftazidime, and only carbapenemases will confer resistance to meropenem. Overnight cultures of each strain were inoculated at a 1:100 dilution into 150 µl lysogeny broth (LB), with a 2-fold dilution of ciprofloxacin, ampicillin, ceftazidime, and meropenem ranging from 32000 ng/µl to 7.8 ng/µl, 1024 µg/ml to 0.25 µg/ml, 128 µg/ml to 0.25 µg/ml, and 16 µg/ml to 0.25 µg/ml respectively. Following 18 hours of growth at 30°C with shaking at 150 rpm, optical density (OD) was

measured at 600nm. The MIC was defined as the concentration of antibiotic that visibly inhibited growth after overnight culture.

In addition, data on growth parameters collected in two types of antibiotic-free liquid media, LB and tryptic soy broth (TSB). Growth of bacteria in an antibiotic free environment allows us to determine whether resistance is associated with impaired growth. Growth curves were collected from 150 µl cultures in 96-well plates, inoculated at a 1:100 dilution from over-night cultures. OD_{600} was measured every 37 minutes for 24 hours, and lag phase, maximum growth rate, and density at stationary phase were estimated using BioTek Gen5 software with two replicates for each strain. Values reported are averages of the two replicates.

2.3 Whole-genome Sequencing and Quality control

Genomic DNA was extracted from the 39 clinical isolates, library preparation was carried out using the Nextera XT kit and paired-end 300 bp sequencing was carried out using the Illumina MiSeq platform. Quality control on the data produced by sequencing was performed using Trimmomatic-0.32 (Bolger et al. 2014). Fifteen bases from the beginning, and one base from the end, were removed from each read. In addition, reads were trimmed using a sliding window, with each read clipped once average base-call quality score dropped below 20 in a 4-bp window, which signifies 99% accuracy when bases are called during sequencing. Reads of fewer than 36 bps were also removed. Effects of quality control were visualised by FASTQC, assuring that high quality data were used in further analyses (Patel & Jain 2012).

2.4 Genome Assembly and Annotation

Reference based alignment was carried out using *E. coli* MG1655 (K-12) as the reference genome, using Bowtie2, version 2.1.0 (Langmead et al. 2009), and single nucleotide polymorphisms (SNPs) were called using Samtools (Li et al. 2009). The quality of the alignment was assessed using Qualimap, version 2.0.1 (García-Alcalde et al. 2012) and custom Perl scripts were used to filter SNPs (coverage higher than 15, quality greater than 20 and a frequency of 80% or higher) and to construct a pseudo-alignment using SNPs called with MG1655. Quality is given on a Phred scale (where a quality of 20 corresponds to 99% accuracy) and signifies the confidence in the SNP being called.

De novo assemblies were constructed for all strains in order to conserve information regarding accessory genomes, which can be lost in a reference-based approach. The accessory genome contains strain-specific genes that are involved in processes like niche adaptation, specialization and host-switching (Dobrindt & Hacker 2001; Didelot et al. 2009). VelvetOptimiser-2.2.5 was used to determine the optimal *k-mer* length for the data and *de novo* assemblies were constructed using Velvet-1.2.10 (Zerbino & Birney 2008). Contigs smaller than 200 bps were removed to improve the quality of the assemblies. QUAST-3.1 was used to assess the assemblies (Gurevich et al. 2013). Genomes were annotated using the toolkit provided by Rapid Annotation using Subsytem Technology (RAST) for batch submission (Brettin et al. 2015).

2.5 Multi-Sequence Alignment and Phylogeny

Whole genome alignment was carried out using our *de novo* assemblies of 39 samples, along with the genomes of *E. coli* K-12 (MG1655), *E. coli 0157* and *E. fergusonii,* downloaded from GenBank. ProgressiveMauve-2.4.0 was used to align the genomes with iterative refinement using default settings (Darling et al. 2010). ProgressiveMauve is a multiple alignment tool that identifies locally collinear blocks (LCBs), each being a homologous region of sequence shared between genomes. LCBs shared among all 42 genomes are classified as the 'core genome', while others are considered to be part of the 'accessory genome'. A core genome phylogeny was constructed using Bayesian phylogenetic inference as implemented by BEAST (Drummond et al. 2012), using a general time-reversible model with gamma correction (GTRGAMMA model). The Markov chain Monte Carlo (MCMC) chain was run for 20 million generations with sampling every 1000 states.

The pseudo-alignment was also used to construct a phylogeny using Bayesian phylogenetic inference as implemented by BEAST (Drummond et al. 2012), using a general time-reversible model with gamma correction (GTRGAMMA model). The Markov chain Monte Carlo (MCMC) chain was run for 10 million generations with sampling every 1000 states. The phylogenetic trees were visualized and converted to Newick format using FigTree version 1.4.2 (Rambaut 2008).

2.6 Serotyping and Resistance Gene Identification

Serotype prediction of the strains was carried out using SerotypeFinder 1.1, a publicly available web tool at the Center for Genomic Epidemiology (CGE) (Joensen et al. 2015). SerotypeFinder was used to serotype the strains based on the O and H antigens.

The O antigen contains repeats of an oligosaccharide unit, and is part of the lipopolysaccharides present in the outer membrane of Gram-negative bacteria (Wang et al. 1998). The H antigen is the central, variable region of the flagellin protein of *E. coli* (Wang et al. 2003). Serotypes were characterized based on a threshold of 85% identity and minimum length of 60%.

ResFinder 2.1, also available at CGE, was used to identify genes associated with *β*-lactam resistance. Genes were identified based on a threshold of 98% identity and minimum length of 60% (Zankari et al. 2012). To determine the genetic basis for ciprofloxacin resistance *gyrA*, *gyrB* and *marR* genes were manually annotated. Alignments for these genes were extracted from the pseudo-alignment using a custom Perl script and visualized using MEGA 6 (Tamura et al. 2013).

2.7 Phenotype-genotype Correlations

In order to identify novel genes that might contribute to ciprofloxacin resistance three different approaches were used: CoEvol, Kover, and PPFS.

2.7.1 CoEvol

CoEvol is a Bayesian method for the detection of associations between rates of substitution at genes and rates of evolution of key phenotypic characters (Lartillot & Poujol 2011). It is a phylogenetically based approach that jointly estimates rates of sequence evolution and rates of phenotypic evolution (Lartillot & Poujol 2011). I predict that, if variation in a particular gene underlies phenotypic variation, then rates of evolution should be correlated between that gene and the phenotype of interest. Using a consensus tree constructed from the pseudo-alignment, runs were carried out using the

dNdS model. This model quantifies selection pressures by comparing the rate of synonymous substitutions (dS), to the rate of non-synonymous substitutions (dN). The ratio of dN to dS is typically regarded as reflecting the strength and nature of selection on a protein. A ratio of dN to dS higher that 1 is evidence of positive selection, lower than 1 is interpreted as purifying selection, and equal to 1 suggests neutral evolution (Lartillot & Poujol, 2011; Zhang et al. 2005).

Evolutionary rate estimation and covariance analyses were performed for 75 genes. Of these 75 genes, 63 were genes of interest – that is, genes in which mutations might affect growth rate parameters and ciprofloxacin MIC phenotypes (17 encoding efflux pumps, porins, and other transmembrane proteins; 7 encoding constituents of the primosome, 36 genes encoding for proteins involved in carbohydrate degradation and the tricarboxylic acid cycle and 3 were the known resistance genes *gyrA*, *gyrB*, and *marR*) and 12 were baseline genes not expected to be associated with any phenotypes of interest. The top five trees from BEAST were also used to test the effects of fixed tree on the data produced. CoEvol was used to infer covariance between rates of evolution of seven phenotypic characteristics (ciprofloxacin MIC, three growth parameters in both media) and substitution rates. MCMC sampling was carried out for ~1500 generations, with the first 500 removed as burn-in.

2.7.2 Kover

Kover is a machine-learning algorithm that relies on reference-free genome comparisons. Given two groups of phenotypically distinct individuals, Kover uses models

that accurately discriminate them and determines predictors of the phenotype using both conjunction (logical-AND) or disjunction (logical-OR) models (Drouin et al. 2016). Genomes assembled by Velvet were split into *k*-mers of 31 bps using the Ray Surveyor Tool, which is part of the Ray *de novo* assembler (Boisvert et al. 2010). A presence and absence matrix for *k*-mers from the 39 clinical samples and *E. coli* K-12 (MG1655) was used along with binary phenotypic data to learn and predict association between these *k*-mers and phenotypic states. Growth parameter data were converted to binary by assigning '0' to strains with a given growth parameter below the median, and '1' to strains with a given growth parameter greater than the median. For ciprofloxacin MIC data, a cutoff of 4 μg/ml was used to assign a binary phenotype (Clsi, 2014). Data were analyzed using both conjugation and disjunction models and a 20-fold cross validation, which divides the data into subsamples and trains it 20 times, using a different subsample as testing data each time.

2.7.3 PPFS

PPFS identifies SNPs that are most likely to be causally related to a phenotype based on χ^2 probability (Hall 2014). *CausalSNPs* determines the probability that the SNP changed randomly across a branch where the phenotype also changed (Hall, 2014). SNPs with a lower χ^2 probability have a higher likelihood of causing the phenotype. Given the assembled and annotated clinical genomes along with *E. coli* K-12 (MG1655), *kSNP* identified 258,426 SNPs using k=19 (Gardner & Hall 2013; Gardner & Slezak 2010). *kSNP* provides a SNP matrix, a SNP list, SNP phylogenies in various formats and other files which can be used for further analysis. These files were used by the PPFS package

to determine 'diagnostic SNPs' that predict phenotype (Hall 2014). SNPs were added until the accuracy with which true positives are predicted (PPV) reaches 0.85. MEGA 6 was used to calculate the ancestral state of each SNP at each internal node of a maximum likelihood tree (Tamura et al. 2013). Most probable sequences inferred by MEGA were used by *CausalSNPs* to identify SNPs associated with a phenotype of interest.

2.8 Genes Gained and Lost

Command line application of Basic Local Alignment Search (BLAST) version 2.2 was used to find orthologs between genes annotated by RAST for our 39 clinical samples, as well as *E. coli* K-12, *E. coli 0157* and *E. fergusonii*. Orthologs were detected as reciprocal best BLAST hits, which were detected using custom Perl scripts. Genes present in all genomes, the core genome, were removed from the analysis. A presence and absence matrix of genes constituting accessory genomes was used by Genoplast to infer losses or gains of genes along the branches of a core genome phylogeny (Didelot et al. 2009). The run was carried out for 200,000 iterations, with the first 100,00 iterations being used as burn-in. Functional classification of the genes gained and lost was determined using PantherScoring Tool-11.0 and overrepresentation of the genes in our data was determined using the Statistical overrepresentation test by PANTHER. This test first divides the data provided into PANTHER classifications, compares the percentage of GO terms in given data to the percentage in the database (*E. coli*) and calculates over or under-representation of these terms in the data provided (Mi et al. 2013).

2.9 Population Structure

To define the population structure the hierBAPS module of the Bayesian Analysis of Population Structure (BAPS) software was used (Cheng et al. 2013). hierBAPS uses hierarchical clustering of DNA sequence in an approach where data clustered at a particular stage is re-clustered at the next-stage. This nested approach provides greater resolution of population structure (Cheng et al. 2013; Willems et al. 2012). Using the core alignment produced by MAUVE, seven levels of nested clusters were found to fit our data.

3 Results

3.1 Phenotypic Variation Amongst Clinical Isolates of *E. coli*

Phenotypic data was gathered for 39 clinical isolates of *E. coli*, with varying quinolone and *β*-lactam resistance profiles, collected from hospitals across Canada. A standard lab strain, *E. coli* K-12 (MG1655), was also used. Ten phenotypic characteristics were measured for each of these 40 strains: ciprofloxacin, ampicillin, ceftazidime and meropenem MICs (a standard measure of antibiotic susceptibility), and three growth parameters (length of lag phase, maximum growth rate, and density at stationary phase) in two liquid media (LB and TSB). Variation in duration of lag phase in both media and maximum growth rate in TSB is evident (One way ANOVA: $P \leq 0.05$) (Figure 2; Appendix Table1), indicating differences in the performance of each strain under laboratory conditions. The distribution of ciprofloxacin and ceftazidime MIC values shows that some strains have very low MICs and others with high MIC values. Less variation was observed in ampicillin and meropenem MICs. Strains with MICs ≥ 4 μg/ml, ≥ 32 μg/ml, ≥ 16 μg/ml, ≥ 4 μg/ml were classified as resistant to ciprofloxacin, ampicillin, ceftazidime, and meropenem respectively (Clsi, 2014).

Correlation was assessed between all the phenotypes measured and significant correlations ($P \leq 0.05$) between maximum growth rate in LB and both length of lag phase and density at stationary phase in TSB was found. Significant correlation between ampicillin and ceftazidime MICs and ampicillin and ciprofloxacin MICs was also observed (Table 1). Few correlations were found between MICs and growth parameters, suggesting that resistance was not associated with growth in an antibiotic free environment. This can be explained by two mechanisms. First, that the resistant mutation

had no cost associated with it or second, the presence of compensatory mutations that are restoring the cost of resistance.

Table 1: Correlation (r) (top right) between 10 phenotypes measured and the significance (P value) (bottom right) shown below. Phenotypes measured were length of lag phase (Lag), maximum growth rate (Vmax), and density at stationary phase (MaxOD) in two liquid media (LB and TSB and minimum inhibitory concentration (MIC) of ciprofloxacin (Cip), ampicillin (Amp), ceftazidime (Cef) and meropenem (Mero). MIC values were log-transformed. Significant correlations are bolded.

	Vmax LB	**Lag LB**	**MaxOD LB**	**Vmax TSB**	**Lag TSB**	**MaxOD TSB**	**MIC Cip**	**MIC Amp**	**MIC Cef**	**MIC Mero**
Vmax LB	-	-0.12	-0.17	-0.05	**-0.54**	**-0.29**	0.05	**0.33**	0.12	0.11
Lag LB	0.48	-	-0.24	0.05	0.25	-0.12	0.18	-0.21	-0.09	0.07
MaxOD LB	0.28	0.13	-	0.22	-0.03	**0.58**	0.08	0.08	0.03	-0.17
Vmax TSB	0.77	0.77	0.17	-	-0.15	0.11	0.11	-0.01	0.23	0.05
Lag TSB	0.00	0.11	0.83	0.37	-	0.20	0.19	0.02	-0.05	0.02
MaxOD TSB	0.07	0.47	0.00	0.52	0.23	-	-0.07	-0.05	-0.05	-0.02
MIC Cip	0.74	0.27	0.64	0.49	0.25	0.66	-	**0.49**	**0.32**	-0.05
MIC Amp	0.04	0.18	0.63	0.93	0.88	0.77	0.00	-	0.56	0.01
MIC Cef	0.46	0.58	0.86	0.16	0.77	0.77	0.05	0.00	-	0.23
MIC Mero	0.51	0.67	0.28	0.77	0.93	0.88	0.76	0.96	0.15	-

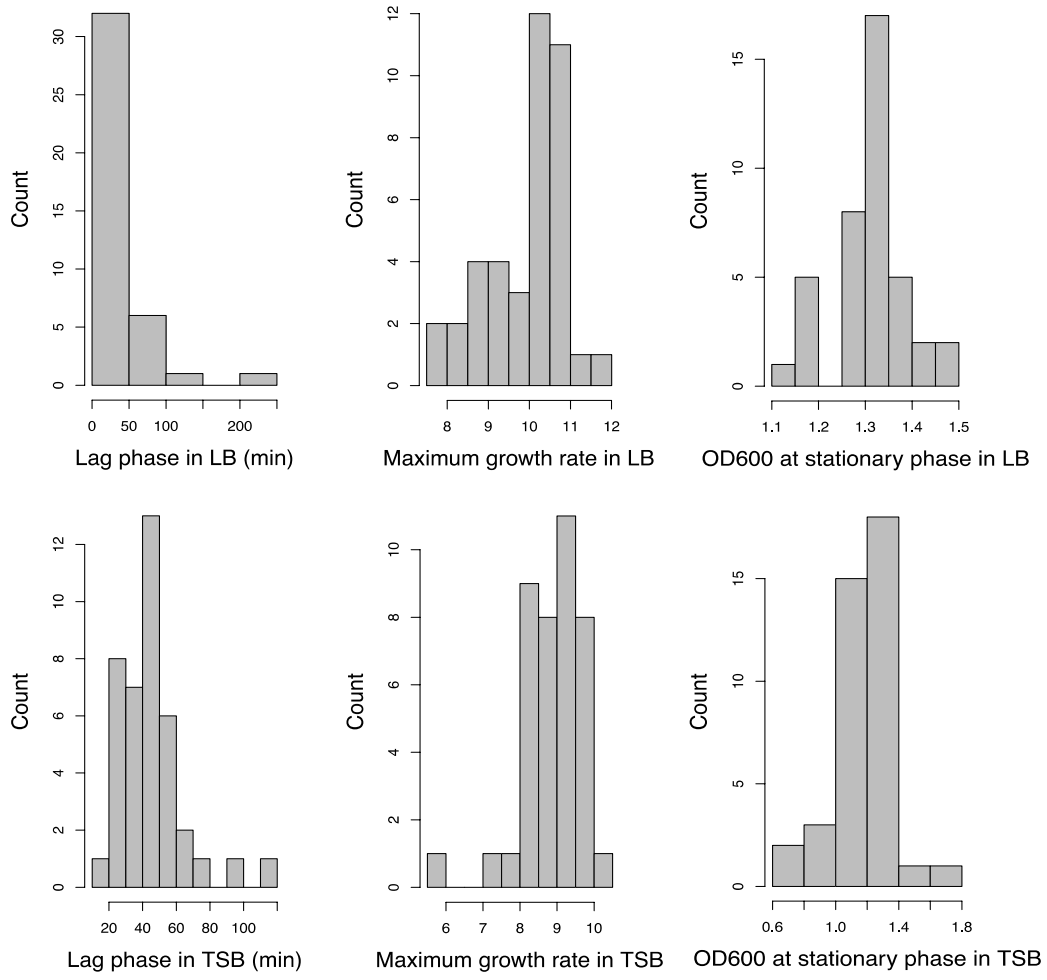

Figure 2: Distributions of growth parameters and minimum inhibitory concentration for 40 strains of *E. coli*. Duration of lag phase, maximum growth rate, and density at stationary phase were estimated from 24-hour growth curves, with OD_{600} measured every 37 minutes.

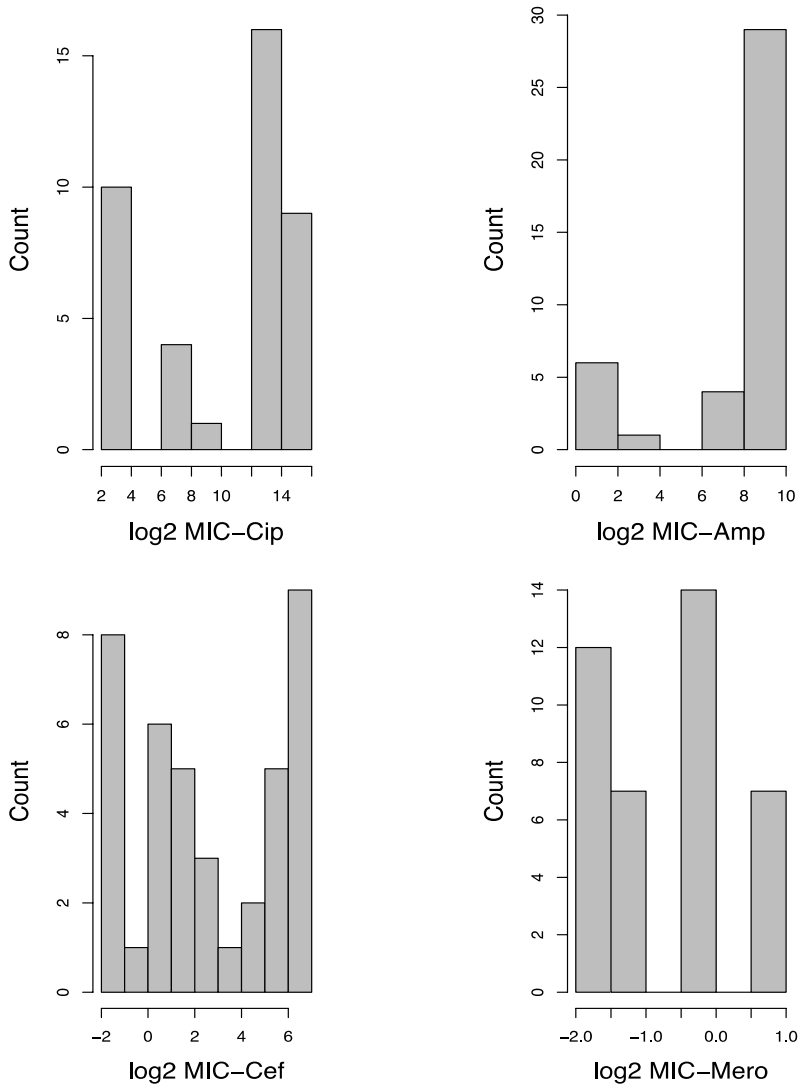

Figure 3: Minimum inhibitory concentration (MIC) was measured as the concentration of ciprofloxacin (Cip), ampicillin (Amp), ceftazidime (Cef) and meropenem (Mero) required to visibly inhibit growth of bacteria after overnight culture.

3.2 Genome Assembly and Phylogenetic Inference

De novo assemblies for the 39 clinical strains were constructed using Velvet and the quality of the assemblies was assessed using QUAST. Alignment quality is gauged based on commonly used parameters such N50, size of the largest contig, and the number of contigs. In evaluating assemblies N50 values are often used as a comparative tool. N50 is the size of the contig such that 50% of the genome assembly is contained in contigs of this length or larger (Lin et al. 2011). N50 measures the connectivity of the assembly, and higher N50 values are considered to be indicative of better alignments (Lin et al. 2011). Average N50 across all the assemblies was 2,910,985. On average the total length of the assemblies was found to be larger than MG1655 (4,639,675 bps) (Keseler et al. 2013). This is not surprising as pathogenic *E. coli* are often found to have larger genomes – for example, *E. coli* O157:H7 has a genome size of 5,705,645 bps. Hence, the length of assemblies was found to be in normal range for *E. coli* (Table 2).

A full genome alignment for all 42 genomes was constructed using ProgressiveMauve. 111,344 LCBs, which are conserved, homologous segments present in two or more genomes, were identified. A core genome alignment of 1.6Mb was extracted from the full alignment and used to construct a core genome phylogeny. The phylogeny shows two distinct clades, each containing susceptible and resistant strains, suggesting that ciprofloxacin and *β*-lactam resistance have evolved multiple times in the Canadian population (Figures 4,5). Annotations of the assemblies using RAST predicted an average of 5083 coding sequences and 52 RNAs. The number of coding sequences predicted using RAST is higher than *E. coli* K-12 (4,499 genes) but still within the range found in *E. coli.*

Reference-based alignment was carried out for 39 strains using the sequence of the standard laboratory strain *E. coli* K-12. Mean coverage for the clinical isolates ranged from 15x to 91x, and mean mapping quality ranged from 34.28 to 40.14, with higher values of mapping quality indicating that reads map uniquely to the reference genome. Reference-based alignment of each strain with MG1655 identified a total of 1,618,561 SNPs. SNPs were used to infer a phylogeny of the clinical strains and *E. coli* K-12 (Appendix, Figure1).

Table 2: *De novo* assembly parameters as quantified by Quast and annotation results determined using RAST.

Strain	Number of Contigs	Largest Contig	Total length (MB)	GC (%)	N50	Mean Coverage	Predicted Coding Sequences	Predicted RNAs
gz1	317	133996	5.05	50.49	31352	38.47	4877	33
gz2	3366	16580	4.66	49.82	1661	52.45	4949	29
gz3	65	2666445	5.18	50.56	2666445	58.08	5035	71
gz4	59	4682335	4.9	50.62	4682335	19.6	4606	61
gz5	80	4747128	5.12	50.47	4747128	23	4919	56
gz6	136	3851947	5.31	50.68	3851947	31.89	5248	55
gz8	63	2997609	5.04	50.68	2997609	33.9	4868	62
gz9	591	59904	4.89	50.6	15354	16.8	4766	27
gz10	110	4494224	5.02	50.85	4494224	49.83	4876	56
gz11	61	2295135	5.05	50.84	2224081	91.02	4903	58
gz12	90	4653115	5.16	50.78	4653115	20.64	5083	46
gz13	343	155955	5.03	50.68	31712	32.42	4920	26
gz14	235	3311323	5.13	50.72	3311323	19.08	5049	45
gz15	3140	9162	4.61	50.5	1831	53.77	4985	37
gz16	93	4593273	5.1	50.72	4593273	30.18	5013	55
gz17	124	3746536	4.98	50.77	3746536	16.07	4840	58
gz18	353	155966	5.03	50.7	32931	25.28	4897	29
gz19	142	3728458	5.06	50.74	3728458	17.56	4896	55
gz20	85	4670031	5.14	50.74	4670031	44.86	5063	65
gz21	137	4443969	5.17	50.72	4443969	46.81	5045	68
gz22	116	4581314	5.28	50.65	4581314	28.12	5079	62
gz23	3043	12323	5.1	50.31	2296	45.96	5415	47
gz24	287	222870	5.13	50.7	61547	26.43	5069	58
gz25	114	3858091	4.99	50.72	3858091	26.75	4767	58
gz26	181	4160562	5.15	50.43	4160562	18.47	5020	52
gz27	92	4819660	5.3	50.49	4819660	28.63	5109	68
gz28	133	4306205	5.17	50.58	4306205	30.01	5083	58
gz29	922	41147	5.03	50.42	9953	19.66	4893	35
gz30	124	4611195	5.12	50.68	4611195	27.52	5071	57
gz31	157	4676687	5.46	50.34	4676687	15.18	5352	41
gz32	188	4752638	5.37	50.5	4752638	18.58	5276	62
gz33	94	4596933	5.34	50.62	4596933	29.67	5251	57
gz34	380	244729	5.29	50.4	32355	28.66	5167	35

gz35	80	2297835	5.25	50.76	1464776	46.7	5204	51
gz36	164	4888157	5.45	50.7	4888157	57.54	5449	61
gz37	139	4532448	5.34	50.64	4532448	55.4	5326	48
gz38	283	2663773	4.83	50.75	2663773	74.1	4649	48
gz39	126	4532834	5.51	50.55	4532834	35.11	5535	76
gz40	67	4841694	5.08	50.78	4841694	45.53	4949	63

3.3 Genetic Basis for Quinolone and *β*-lactam Resistance

To determine the genetic basis for quinolone resistance, SNPs were analyzed at the *gyrA, gyrB* and *marR* genes for all clinical strains against *E. coli* K-12. Five mutations in *gyrA* and six mutations in *gyrB* were found, occurring in both resistant and susceptible strains (Table 2). Of the 5 mutations in *gyrA*, only 2 were present in the quinolone resistance-determining region (QRDR) (Ser83-Leu and Asp87-Asn). Both of these mutations, known as causative mutations for ciprofloxacin resistance, occur only in ciprofloxacin resistant strains except for gz10, which was classified as sensitive to ciprofloxacin according to its MIC (250ng/ml) (Figure 4). Some strains, like gz12, gz14, gz28, gz29 and gz33, were found to lack these known resistances mutation despite high ciprofloxacin MICs. For *gyrB,* the 6 mutations found were not present at the two positions associated with quinolone resistance (Asp426 and Lys447), however mutation S492N falls in the QRDR of *gyrB* (Table 3).

β-lactamase producing genes were identified using ResFinder. In the clinical isolates, 4 groups of *β*-lactamases were found: TEM (-1B, -1C, 116-like), OXA (-1 and 36-like), CTX-M (-14 and -15) and SHV (-2 and -12). TEM genes were present in both ESBL-positive and ESBL-negative strains, whereas OXA and SHV were present only in the ESBL-positive strains. CTX-M were present in all of the ESBL-positive strains except gz27, gz31 and gz38, which show presence of OXA and/or SHV *β*-lactamases (Figure 3).

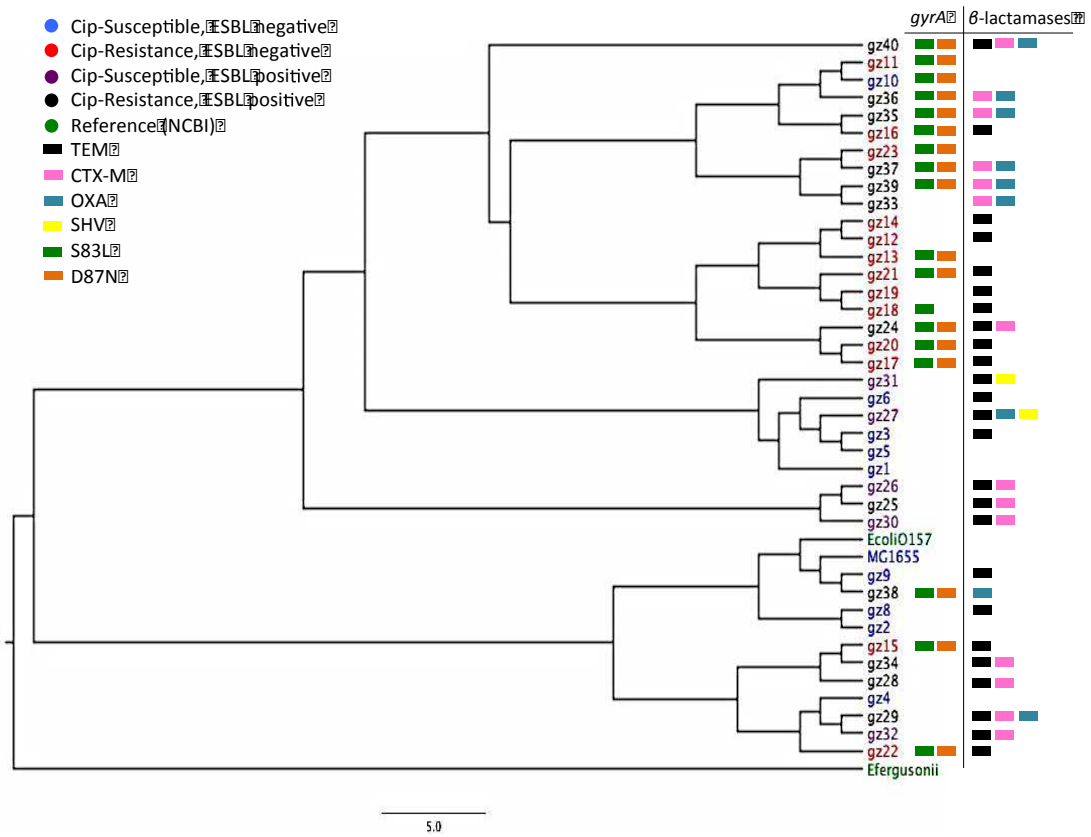

Figure 4: Phylogenetic tree constructed using Bayesian inference showing the relationship between 39 clinical strains, *E. coli* K-12, *E. fergusonii* and *E.coli* O157. The consensus phylogeny was inferred from 20 million generations using BEAST. Mutations observed in the QRDR of the *gyrA* gene, as well as presence of *β*-lactamases genes, are shown. Strain names are colour according to their resistance profile.

Table 3: Mutations present in all clinical strains in *gyrA, gyrB* and *marR* genes when aligned to *E. coli* K-12.

Strain	GyrA	GyrB	MarR
gz1	A828S	A618T	G103S, Y137H
gz2			G103S, Y137H
gz3	D678E, A828S	E185D	G103S, Y137H
gz4		A618T	G103S, Y137H
gz5		E185D	
gz6			
gz8			
gz9			
gz10	S83L, D87N, A828S		G103S, Y137H
gz11	S83L, D87N, A828S	A618T	G103S, Y137H
gz12			
gz13	S83L, D87N		G103S
gz14			
gz15	S83L, D87N, D678E, A828S	S492N, A618T, E656D	G103S
gz16	S83L, D87N, A828S		G103S, Y137H
gz17	S83L, D87N		
gz18	S83L	A618T	
gz19		A618T	
gz20	S83L, D87N, A828S	A618T	G103S, Y137H
gz21	S83L, D87N, A828S	A618T	G103S, Y137H
gz22	S83L, D87N, D678E	A618T, A653T, I663V	K62R
gz23	S83L, D87N, A828S	A618T	G103S, Y137H
gz24	S83L, D87N, A828S	A618T	G103S
gz25			G103S
gz26			
gz27	D678E, A828S	E185D	
gz28	E609G		
gz29			
gz30			
gz31			
gz32			
gz33			
gz34		E656D	
gz35	S83L, D87N, A828S	A618T	Y137H
gz36	S83L, D87N, A828S		G103S
gz37	S83L, D87N, A828S	A618T	G103S
gz38	S83L, D87N		
gz39	S83L, D87N, A828S	A618T	G103S, Y137H
gz40	S83L, D87N, A828S	A618T	G103S, Y137H

3.4 Phenotype-genotype Correlations

I used three different programs to infer correlations between genotype and seven of the measured phenotypes (ciprofloxacin MIC and lag phase, maximum growth rate, and density at stationary phase in LB and TSB). First, CoEvol was used to detect associations between rates of substitution at candidate genes and rates of evolution of key phenotypic characters. This phylogenetically based method estimates the correlation (r) between rates of evolution of a gene and phenotype being tested. The full (marginal correlation) and multiple regression (partial correlation) models were used to test the contribution of various genes to the seven phenotypes. Using marginal correlations, 8 genes were found showing high covariation (pp > 0.025 and < 0.975) with Cip-MIC. No significant covariation was observed using partial correlation. Similarly, CoEvol showed correlation between numerous genes and other phenotypes (Table 3). CoEvol only showed correlation with candidate genes, and no significant correlation was observed between our "baseline" genes with any of the phenotypes (Figure 5). Robustness to tree topology was determined by re-running analyses with the five best trees in the BEAST MCMC chain. No significant correlation was found in the data produced by different trees (Figure 2, Appendix), suggesting that CoEvol may be sensitive to underlying tree topology.

Secondly, Kover was used to estimate phenotype-genotype correlations. Kover is a machine-learning program that uses a presence and absence matrix of *k*-mers determined using Ray. This reference free method uses the data to predict associations between *k*-mers and phenotypes based on the model learned from the data provided. Kover uses rule-based models, where rules are individual units that detect the presence or

the absence of a *k-mer* in a genome. Kover predicted associations for only one gene for four of the phenotypes, and multiple genes for the other phenotypes. For ciprofloxacin MIC, Kover predicted two genes, *endA* and *hrpB*.

Lastly, PPFS was used to determine genotype-phenotype correlation. PPFS predicted genes associated with phenotypes based χ^2 probability, using SNPs called by *kSNP*. *kSNP* uses a reference free method to assemble the genomes and divides them into *k*-mers (19 bps). PPFS predicted genes associated with all phenotypes except Vmax in TSB media, for which no significant SNPs were found. Since a small k-mer (19 bps) size often returned multiple BLAST matches, PPFS produced more predicted genes for most phenotypes than other programs (Table 4).

While comparing the three programs used, little overlap, if any, was found in the genes inferred by the three programs to correlate with the phenotypes tested. There is also no functional similarity in the genes predicted by these three programs. All three methods, CoEvol, Kover and PPFS, did not show significant correlation between ciprofloxacin MIC and the known mediators of ciprofloxacin resistance, *gyrA, gyrB,* and *marR*. However, CoEvol correlation coefficients and posterior P-values tend in the "correct" direction for *gyrB* and *marR* (green datapoints in Figure 5).

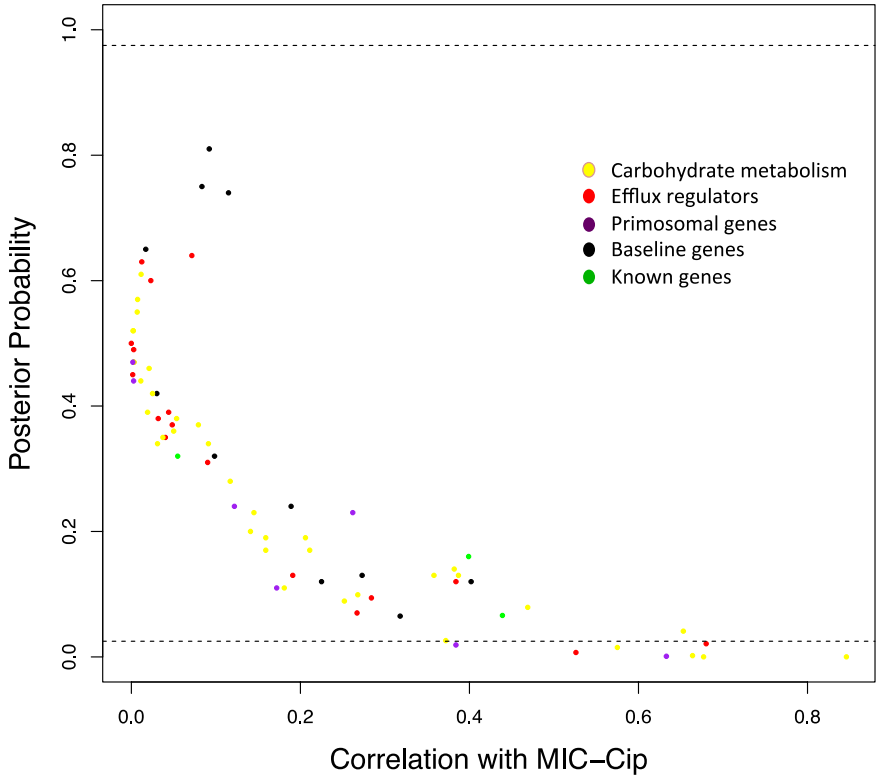

Figure 5: Covariance between rates of molecular evolution and rates of change in ciprofloxacin MIC estimated using the dSdN model in CoEvol. For each of 75 genes, the posterior probability of covariance is plotted against R^2. Dashed lines give posterior probabilities of 0.025 and 0.975.

Table 4: Predicted genes showing correlation with seven phenotypes determined using three different programs.

Phenotype	PPFS	Kover	CoEvol
MIC for Ciprofloxacin	gapB, fhuB, cobT, yhjB, glpA, yhim, mutY, fecA, modF, relA, yegH, ytfJ, pka, ybhS, ybfL, blc, rspB, yaiN, yieL, ygaH, nanK, rne, rhsD, rhsC, rhsB, rhsA, cpdB, galP, ydfE, fre, kefC, dnaJ, apbE, entS, kdpB, glnL, pitB	endA, hrpB	glpX, malZ, priA, pgi, acrD, lpd, dnaT, mdtG
Vmax LB	yhcG, ygcb, ravA, tpiA, fhlA, hybC, sslE, ygcG, acrF, rlmF	djlB	glpX, pgi, mdtG, lpd
Lag LB	ydhP, dadX, paoB	ybgD, mocA, yegT	glpX, pgi, dnaB, malZ, priA, acrB, mdtG
OD LB	bluF, ykgG, hrpA, ligT, sseA, thrS, leuS, creD, rne, nrfC, aaeB, lacZ, acpT, hisG, hsdR, tap, pepD, gadA, rhsH, alaA, yliE, ybjD, gadB, rbsR, putP, malF, tsaB, mutM, insI4, insI3, insI1, ghrB, katG, yeeP	thiL	dnaB
Vmax TSB		arsB	glpX, pgi
Lag TSB	phnJ, ygiD	insK, rsxC, mukB	glpX, malZ, pgi, dnaB, priA, acrB, fbaA, mdtA
OD TSB	insH1, cpsG	yhfT	glpx, pgi, priA

3.5 Serotyping and Population Structure

Serotypes for the 39 clinical strains were predicted using ResFinder 2.1 and were mapped to the core genome phylogeny (Figure 6). Fifteen different serotypes were found in these strains, with O25:H4 being the most common serotype. Twenty-two strains were found to be of serotype O25:H4, of which 21 were closely clustered on the phylogeny (Figure 6). Clustering of the serotypes on the tree provides confidence in the phylogeny. Two strains, gz29 and gz34, only showed the presence of H or O antigen respectively. Lack of a second antigen can be due to incomplete genome sequence or absence of H antigen can be due to lack of flagella.

Population structure was defined using hierBAPS, which grouped our data into seven clusters when two levels of hierarchy were specified, which allows for the data to be clustered twice. Of the seven clusters, two comprised of *E. fergusonii* and *E. coli* O157, for location is not known (genome sequences for these strains were obtained from GenBank). The remaining data were grouped into five clusters, which showed no correlation with the location of isolation of the strain (Figure 7). Even though more strains were isolated from Eastern Canada (29) than Western Canada (10), similar proportions of the three clusters (1, 3, 6) were observed in both Eastern and Western Canada. The fourth and fifth clusters are different between both geographic regions but are present in similar proportions. This difference, however, is inconclusive and a larger sample is required for any conclusive results.

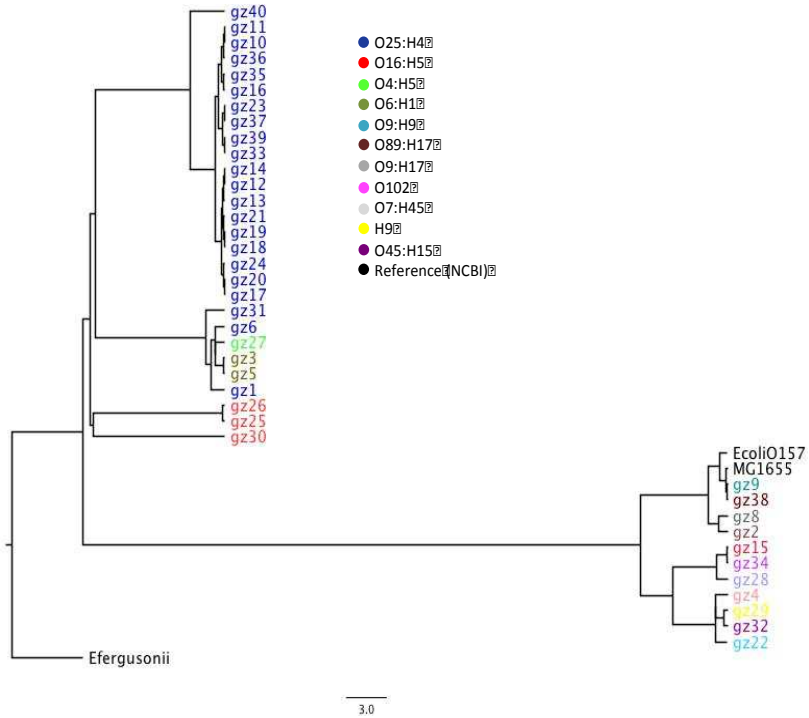

Figure 6: Phylogenetic tree constructed using Bayesian inference showing the relationship between 39 clinical strains, *E. coli* K-12, *E. fergusonii* and *E. coli* O157. The phylogeny was inferred from the core genome alignment and the consensus tree was constructed from an MCMC run of 20 million generations using BEAST. Relationships between various serotypes are shown by coloration of tip labels.

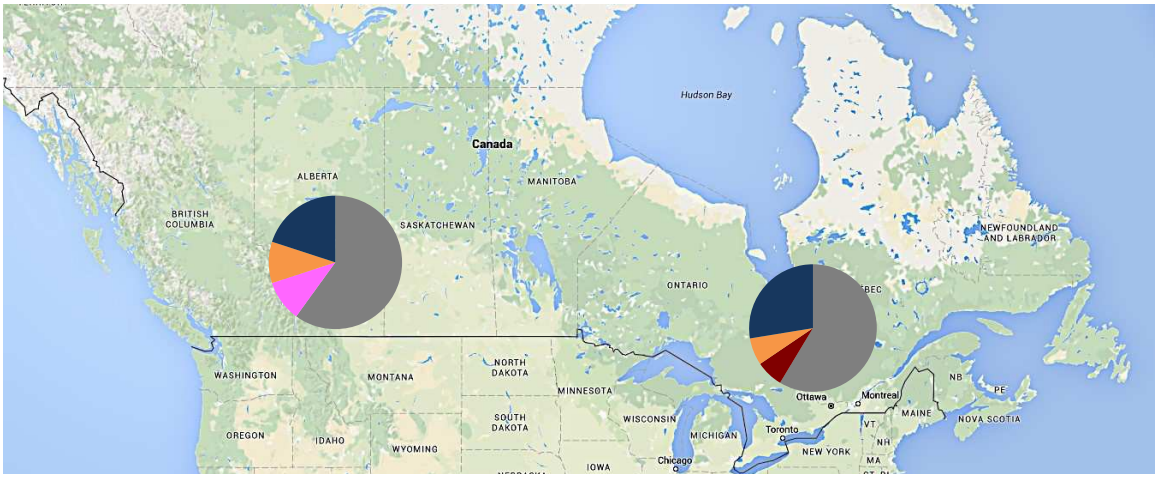

Figure 7: Map of Canada showing the proportion of five clusters seen in Eastern and Western Canada. Black line demarcates the division between Eastern and Western Canada. Cluster 1 (grey), 6(blue), 3 (orange), 2 (pink), 4 (red) are shown.

3.6 Genomic Flux

Genoplast was used to model the overall rate of genetic material being gained and lost in 39 clinical strains, *E. coli* K-12, *E. fergusonii* and *E. coli* O157. 1364 orthologous genes were found in all 42 strains, forming the core genome. The accessory genome consists of 15,442 genes present in some, but not all strains. Losses and gains of these accessory genes were modeled along 82 branches of the phylogenetic tree constructed from the core genome using BEAST. From the 200,000 iterations in Genoplast, a consensus tree showing gain (red) and loss (blue) of features has been shown (Figure 8). A large amount of genetic flux was seen at the shorter, terminal branches of the phylogeny rather than the longer internal branches (Figure 8). Net percentage gain and loss was calculated using the genes with greater than 95% probability of being lost or gained. No correlation was found in the net gain/loss percentage and the amount of accessory genes present in the strain (Appendix, Table 3). Upon determining the molecular function of genes gained and lost across the phylogeny and three significant ($P<0.05$) functional categories were found for both gains and losses (Figure 8).

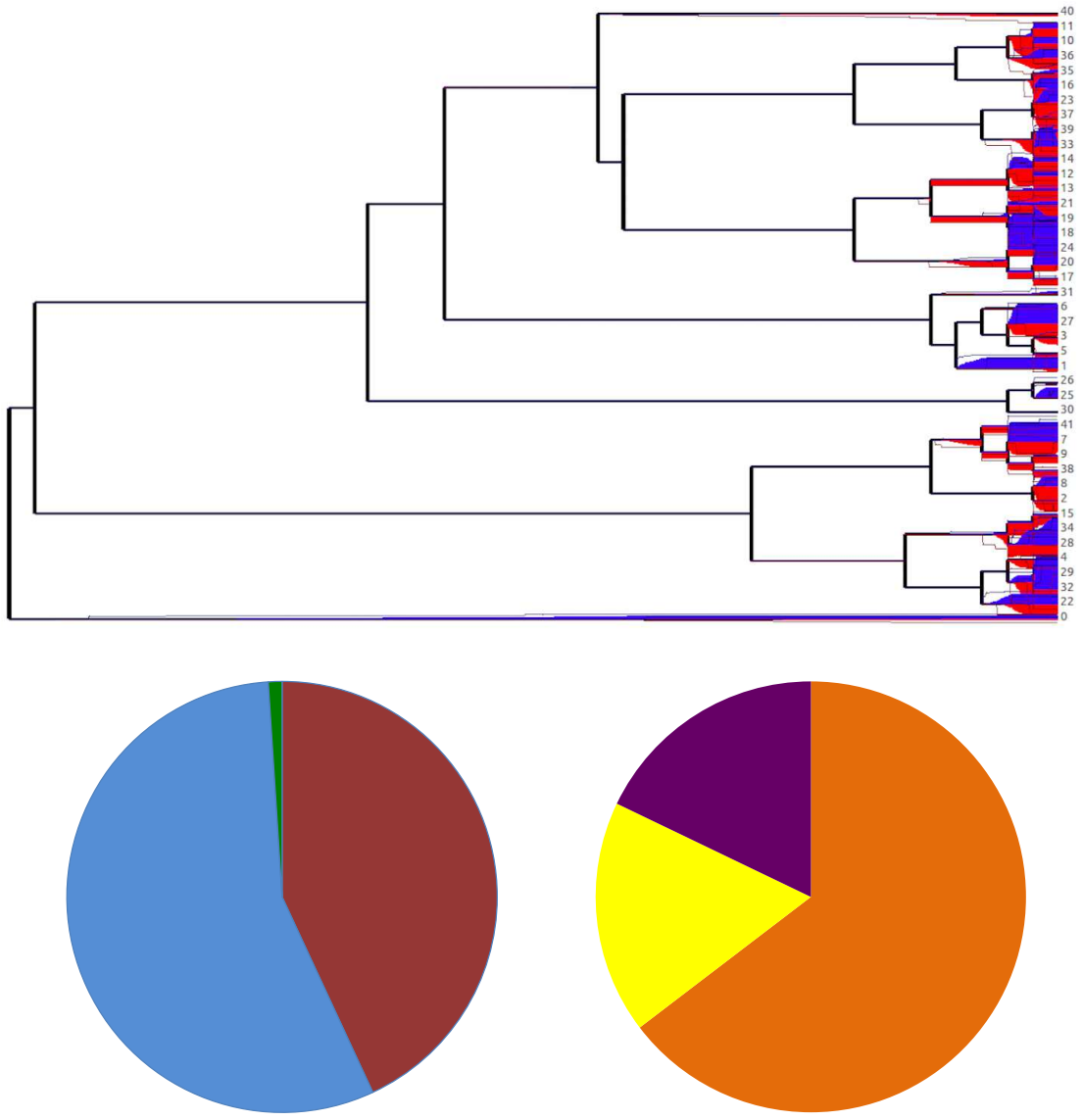

Figure 8: Consensus phylogenetic tree showing the features gained (r+) in red below and features lost in blue (r-) above the branches of the tree. The width of the red and blue lines is proportional to detected genomic gains and losses respectively, and the 95% confidence interval is shown by the lines of corresponding colours above and below the solid regions. Genes overrepresented in the list of gains (left) and losses (right) are shown at the bottom. Genes involved in DNA binding (red), nucleic acid binding (blue), structural component of ribosome (green) transporter activity (purple), oxidoreductase activity (yellow) and catalytic activity (orange) are shown.

4 Discussion

Antibiotic agents are among the most important contributors to the modernization of medicine, and it is difficult to imagine the continuation of advances of recent years without them. However, the emergence of antibiotic resistance threatens our ability to care for patients and is among the top public health threats of the 21st century. An increase in the frequency of antibiotic resistance is evident, resulting in thousands of deaths annually (Roca et al. 2015). Today, bacteria resistant against all major classes of antibiotics can be found in both clinical and community setting. Further, the threat is increased when bacteria are resistant to newer generations of antibiotics like ciprofloxacin.

One such pathogen is *E. coli*, which causes 18.66% of all hospitalised cases of infections in Canada (Lagacé-Wiens et al. 2013). Infections caused by *E. coli* are commonly treated using two groups of drugs: quinolones and *β*-lactams. In Canada, resistance to both ciprofloxacin (quinolone) and cephalosporins (*β*-lactam) is on a rise with 27% and 9.5% infections resistant to these drugs respectively (Lagacé-Wiens et al. 2013). Presence of resistance to multiple classes of antibiotics, resulting in strains with MDR phenotypes, has progressively narrowed the available treatment options for some pathogens. In order to better understand the evolution of antibiotic resistance I used clinical strains of *E. coli* to determine the contributions of mutations in *gyrA*, *gyrB* and *marR* to quinolone resistance, and the presence of *β*-lactamases towards *β*-lactam resistance. I set out to identify potential novel contributors to quinolone resistance and assess the genes gained and lost during the evolution of these strains. Lastly I evaluated correlations between population structure and location of isolation.

4.1 Genetic Basis of Quinolone Resistance

I set out to identify the genetic basis of quinolone resistance in pathogenic *E. coli*. To do so, mutations in *gyrA*, *gyrB* and *marR*, genes known to contribute to ciprofloxacin resistance, were assessed. Most ciprofloxacin resistant strains were found to carry two known causative mutations (S83L and D87N). These mutation are found in the QRDR, present between position 67 to 106 in *gyrA* of *E. coli* (Ruiz, 2003). The QRDR in DNA gyrase is present neighbouring the active site (tyrosine 122), which covalently binds to the phosphate on the DNA during the initial strand break (Drlica & Zhao, 1997; Jacoby, 2005). Even though mutations in the QRDR are often causative for quinolone resistance, high levels of resistance to fluoroquinolones can be obtained by additional mutations in *gyrA*. In our samples, three other mutation in *gyrA* (E609G, D678E and A828S) were found, of which two had been reported previously to be present in resistant strains; E609G is a novel mutation in *gyrA* (Heisig et al. 1993; Phan et al. 2015). D678E and A828S were present in both resistance and susceptible strains, suggesting these mutations are not causative of the resistance profile. E609G is also the only mutation present in the known resistance genes in strain gz28 and was not found in any other resistant or susceptible strain. Presence of this mutation in a strain with a high-level (MIC ≥8000ng/ml) of resistance, qualifies this mutation to be a viable candidate for further analysis. Also six mutations in *gyrB* (E185D, S492N, A618T, A653T, E656D and I663V) were found in our strains, however these strains lacked the known mutations (D426N and K447E) associated with ciprofloxacin resistance (Ruiz, 2003). After inspecting the gyrase genes, mutations in the *marR* gene were assessed. The *marR* gene encodes a repressor of

the *marRAB* operon. Mutations that reduce binding of the MarR protein result in increased expression of *marA*. *marA* up-regulates *acrAB-TolC* and *micF,* which are known to increase efflux and decrease influx of quinolones respectively (Hopkins et al. 2005). Three mutations were found in *marR* (K62R, G103S and Y137H), that have all been previously recognized to be present in ciprofloxacin resistance strains (Zayed et al. 2015). However, G103S and Y137H were found to be present in both resistant and susceptible strains, suggesting that these mutations are not causative of the resistance profile. Upon a closer look, it was recognized that most strains with high-level resistance carried multiple mutations in multiple genes (Table 2, Appendix Table1). However, gz12, gz14, gz29 and gz33 all have no mutations in these known genes and hence other novel genetic changes should play in a role in quinolone resistance in these strains. Genetic changes in *parC* or *parE*, two subunits of topoisomerase IV, can potentially contribute to this phenotype. Topoisomerase IV, like gyrase A, is involved is relaxing positive supercoils and has been previously linked to fluoroquinolone resistance (Drlica & Zhao, 1997). Presence of *qnr* genes, which bind to the gyrase-DNA complex, can also be potential contributors to resistance in these strains. Lastly, aminoglycoside modifying enzyme class aac(6')Ib-cr, which along with acetylating aminoglycosides have also been shown to acetylate ciprofloxacin, can also contribute to resistance (Vetting et al. 2008).

4.2 Novel Candidate Genes for Quinolone Resistance

Three computational approaches (CoEvol, PPFS, Kover) were used to infer novel changes leading to quinolone resistance. Little overlap in the genes selected by the three programs to correlate with various phenotypes was observed. This lack of overlap can be

attributed to the three programs relying on three different approaches to determine associations. Also many of the genes selected by Kover and PPFS were not tested using CoEvol, and reducing any overlap in the three programs. Lack of association between *gyrA* and the resistance phenotype can be explained by a number of resistant strains not carrying mutations in this gene. Furthermore, both Kover and PPFS rely on a *k-mer* approach, which requires the same SNP to occur in all strains to be detected as a change associated with the phenotype. This decreases the program's ability to detect association of a phenotype with genes like *marR*, where mutations at multiple sites in the gene can result in the phenotype of interest. CoEvol also showed no correlation between the data produced by different trees, suggesting that a fixed tree will have an impact on the results. This is particularly worrisome when working with bacterial strains since it is difficult to construct the "true" tree as different parts of the genome will have different evolutionary histories due to recombination and horizontal gene transfer. Even though none of these three approaches identified correlations between ciprofloxacin resistance and the three known mediators of resistance, some interesting candidates were identified.

The genes predicted to be associated with ciprofloxacin MIC can be divided into three groups based on molecular function: transporter activity (17%), binding (15%), and catalytic activity (49%). The transporter group mostly comprises of genes involved in transmembrane transportation with some genes (13%) involved in carbohydrate transportation. CoEvol showed significant associations between the *mdtG* and *acrD* genes and MICs for ciprofloxacin. The *mdtG* gene encodes an efflux transporter involved in multidrug transportation, and overexpression of this gene has been shown to result in resistance to fosfomycin and deoxycholate (Fàbrega et al. 2010; Nishino & Yamaguchi,

2001). The *acrD* gene encodes a component of AcrAD-TolC, a multidrug efflux transport system (Saier et al. 1994). Mutations in *acrD* have been shown to increase sensitivity to aminoglycosides, and its overexpression reduces the cytoplasmic concentration of fluoroquinolones (Eaves et al. 2004; Swick et al. 2011).

Other genes like *modF, ybhS, fecA, fhuB, pitB, kefC, kdpB* and *galP* all encode transporters, which can potentially play a role in drug efflux. Of the genes involved in binding, 30% encode products binding to nucleic acids, whereas others are protein or calcium ion binding. Two such genes are *priA* and *dnaT*; both encode proteins binding to the DNA and are part of the primosome. The primosome is a protein complex that restarts stalled replication forks (Manhart & McHenry 2015), and so it is possible that mutations in primosomal components either contribute some degree of quinolone resistance and/or relieve costs associated with *gyrA* or *gyrB* mutations. Mutations in these genes may compensate for costs of resistance in *gyrA* or *gyrB* mutants, given the importance of DNA gyrase in unwinding DNA ahead of the replication fork (Chompoux 2001), and previous observations of genetic interactions between *gyrB* mutants and constituents of the primosome (Grompone et al. 2003). Finally, genes predicted for growth rate parameters are largely (78%) involved in catalytic activity, with rest involved in antioxidant, binding and transporter activity.

4.3 Genetic Basis of *β*-lactam Resistance and Serotyping

To identify the genetic basis for *β*-lactam resistance, ResFinder was used to find genes encoding *β*-lactamases. The CTX-M genes were present in all of the ESBL-positive strains except gz27, gz31 and gz38, which showed presence of OXA and/or SHV

β-lactamases. CTX-M-type *β*-lactamases have become the dominant ESBLs worldwide during the past two decades (Cantón et al. 2012; Olesen et al. 2013; Peirano & Pitout 2010). CTX-M types have taken over, replacing the previously common TEM and SHV-types, which often conferred resistance to 1st and 2nd generation *β*-lactams (Cantón et al. 2012).

Over 100 variants of the CTX-M genes have been recorded, of which only CTX-M-14 and CTX-M-15 were found in our sample; both are the only variants currently recoded in North America (Cantón et al. 2012; Davies & Davies 2010). On further analysis, it found that many of the strains carrying CTX-M genes belonged to the O25:H4 serotype. Over 50% of our samples belonged to O25:H4 serotype. Both CTX-M and O25:H4 have been associated with *E. coli* strain ST131 (Cantón et al. 2012; Davies & Davies 2010; Peirano & Pitout 2010). ST131 is the predominant *E. coli* lineage among the extra-intestinal pathogenic *E. coli* (ExPEC), and ST131 strains are commonly reported to be ESBL positive with CTX-M-15 genes and almost always resistant to fluoroquinolones. Currently, this strain is causing serious public health concerns due to its multidrug resistance and presence in both community and hospital settings (Nicolas-Chanoine et al. 2014).

4.4 Genes Gained and Lost

I also set out to determine the genes gained and lost during the evolution of these strains. Many physiological and virulence properties of bacteria are conferred by different subsets of genes that enable different lifestyles in these pathogens (Dobrindt & Hacker 2001). Bacteria show a huge variation in the size of their genomes not only at the genus

level but also within a species. This variation in genome size is captured by the distinction between the bacterial 'core' and 'accessory' genomes. The core genome consists of highly conserved genes that are involved in basic cellular processes, whereas the accessory genome contains strain-specific genes that are involved in processes like niche adaptation, specialization and host-switching (Dobrindt & Hacker 2001; Didelot et al. 2009). Thus, a substantial portion of genetic diversity in bacteria is attributed to the gain and loss of genetic material, collectively termed as genetic flux, in the accessory genome (Didelot et al. 2009; Wren 2000).

To measure genetic flux in our strain collection, Genoplast was used, which infers genes that are gained and lost along the branches of a phylogenetic tree. No correlation between the size of accessory genome and net flux was observed suggesting that the number of genes in a genome does not dictate the amount of genetic flux. Also, a greater amount of genetic flux was seen at the shorter, terminal branches of the phylogeny rather than the longer branches. This pattern of genetic flux is suggestive of purifying selection, where changes are occurring over a short period of time but eventually selected against and removed from the population.

Functional analysis of the genes gained and lost was carried out. It was found that genes gained across the phylogeny showed enrichment of those involved in DNA binding and structural constituents of ribosomes ($P \leq 0.05$). The DNA binding genes can further be divided into DNA topoisomerase like activity, damage DNA binding and DNA repair. Gaining of genes involved in DNA repair functions across the phylogeny may provide the genetic changes required for ciprofloxacin resistance, as most of the stains in this phylogeny are ciprofloxacin resistant. Enrichment of genes involved in transporter,

oxidoreductase and catalytic activity was found among the genes lost across the phylogeny.

4.5 Population Structure

Lastly, I set out to determine the correlation between population structure and location of isolation of the samples. Our strains were divided into 7 clusters of which two included strains for which information regarding location of collection was not known. The remaining strains were divided into 5 clusters. No correlation was found between these 5 clusters and location of isolation suggesting that these strains can be found anywhere in Canada. However, it is recognized that this analysis was carried out with a small number of samples and conclusive results would require much larger sample. This analysis does provide base for further research and hypothesis formation.

5 Conclusion

In this thesis I set out to investigate the contributions of known mediators, and to predict novel genes, contributing to quinolone and *β*-lactam resistance. Most resistant strains were found to carry previously recorded mutations in *gyrA, gyrB* and *marR* genes. I also set out to determine the changes leading to *β*-lactam resistance and found the presence of CTX-M-14 and -15 genes, variants found in North America, in most ESBL positive strains.

Novel mutations in both *gyrA* and *gyrB* were observed, which are candidates for further analysis. However, not all resistant strains carried the mutations in genes known to contribute to quinolone resistance, reiterating the need to further investigate the genotypic changes leading to resistance. In an aim to discover novel genes that correlate with the resistance phenotype, analyses were carried out using three different programs and found numerous novel genes that correlate with ciprofloxacin MICs. Further, the role of genetic flux was investigated in these clinical strains and found a gain in genes involved in DNA binding and DNA repair across the phylogeny.

Future work will be aimed towards analyzing the novel genes selected by our analysis. In particular, genes encoding efflux pumps (*mdtG* and *acrD*) and components of primosome (*priA* and *dnaT*) are of interest. I would like to assess the effects of knocking out these genes on ciprofloxacin MICs. Novel mutations in *gyrA* and *gyrB* genes found during our analysis, will be further investigated. It is recognised that none of the programs showed significant correlation between genes known to cause ciprofloxacin resistance and ciprofloxacin MICs. This lack of correlation is attributed to a small sample size and carrying out this analysis with a larger sample size will be beneficial.

6 Abbreviations

Abbreviation	Explanation
Amp	Ampicillin
CDC	Center of Disease Control
Cef	Ceftazidime
CGE	Center for Genomic Epidemiology
Cip	Ciprofloxacin
dN	Rate of non-synonymous mutations
dS	Rate of synonymous mutations
DSB	Double stranded break
ECDC	European Center for Disease Prevention and Control
ExPEC	Extra-intestinal pathogenic E. coli
HGT	Horizontal gene transfer
Lag	Lag phase
LB	Lysogeny broth
LCBs	Locally collinear blocks
MaxOD	Maximum optical density
MCMC	Markov chain Monte Carlo
MDR	Multidrug resistant
Mero	Meropenem
MIC	Minimum inhibitory concentration
OD	Optical density
PBP	Penicillin binding protein
QRDR	Quinolone resistance determining region
RAST	by Rapid Annotation using Subsytem Technology
SNP	Single nucleotide polymorphism
TSB	Tryptic soy broth
Vmax	Maximum growth rate
WHO	World Health Organization

7 Appendices

Table 1: One-way ANOVA analysis carried out to determine variation among strains in growth parameters

Phenotype	F value	P value
Vmax LB	1.379	0.158
Lag LB	23.43	2.00E-16
OD LB	0.746	0.819
Vmax TSB	2.372	0.00429
Lag TSB	7.611	2.59E-09
OD TSB	1.276	0.223

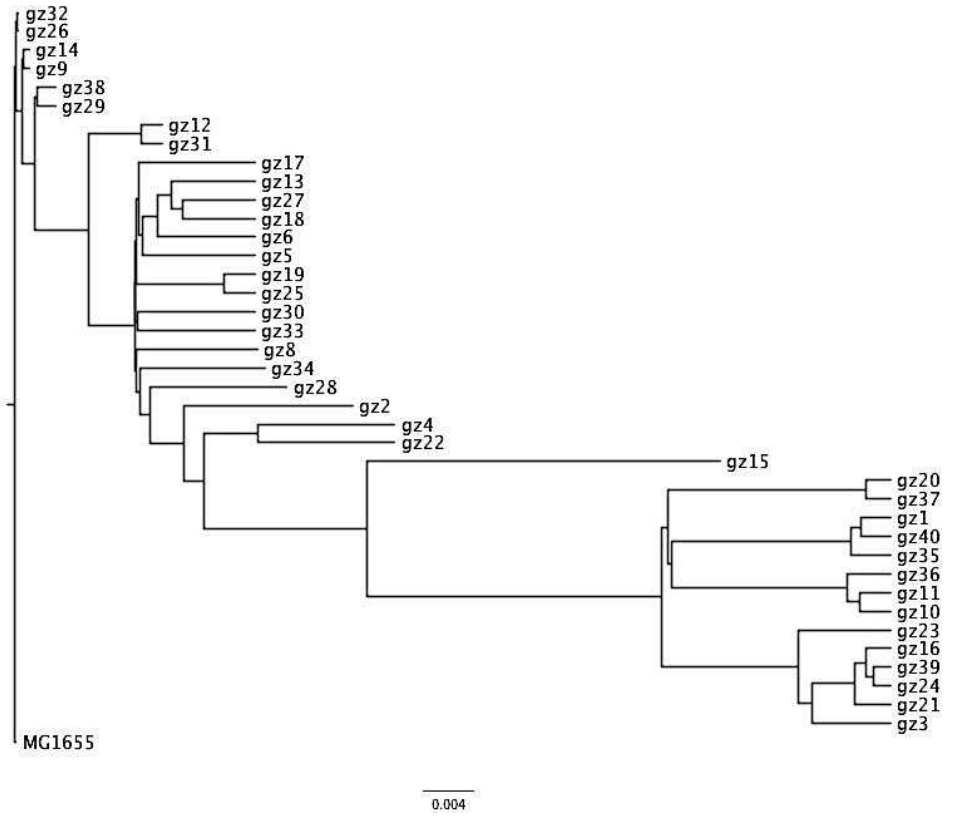

Figure 1: Phylogenetic tree constructed using Bayesian inference showing the relationship between clinical strains and MG1655. The phylogeny was inferred from whole genome alignments with a total of 637, 533 SNPs. The consensus tree was constructed from MCMC run10 million generation using BEAST. Phylogeny used as consensus tree for CoEvol.

Table 2: Data for growth rate parameters (length of lag phase (Lag), maximum growth rate (Vmax), and density at stationary phase (Max OD)) in both LB and TSB media and minimum inhibitory concentration (MIC) for ciprofloxacin (Cip), ampicillin (Amp), ceftazidime (Cef) and meropenem (Mero) collected for 40 strains

Sample	Vmax LB	Lag LB	Max OD LB	Vmax TSB	Lag TSB	Max OD TSB	MIC Cip (ng/ml)	MIC Amp (ug/ml)	MIC Cef (ug/ml)	MIC Mero (ug/ml)
gz1	9.04	53.18	1.20	8.46	62.33	0.91	7.8	4	0.25	0.25
gz2	9.46	36.33	1.42	9.16	39.07	1.23	15.6	8	0.25	0.25
gz3	10.69	45.27	1.26	8.41	46.13	1.07	15.6	512	4	2
gz4	10.54	23.33	1.29	8.14	20.48	1.25	7.8	1	0.25	1
gz5	9.18	41.58	1.29	9.04	46.51	1.17	7.8	4	1	1
gz6	10.58	41.58	1.18	8.37	41.54	0.99	7.8	1024	4	1
gz8	10.14	19.15	1.27	7.92	19.08	1.18	7.8	1024	4	0.5
gz9	10.35	31.44	1.16	9.37	37.07	1.09	250	1024	2	0.25
gz10	8.74	14.15	1.30	9.11	37.37	1.07	250	4	2	2
gz11	9.92	238.41	1.14	9.51	51.15	0.96	16000	4	4	2
gz12	10.43	39.44	1.31	8.77	21.24	1.21	16000	1024	8	1
gz13	10.05	34.36	1.30	5.99	57.51	1.30	32000	256	0.5	0.5
gz14	10.32	36.48	1.34	10.03	45.50	1.26	16000	1024	0.5	2
gz15	9.24	63.18	1.19	8.43	49.37	1.04	32000	256	0.5	0.25
gz16	10.04	60.52	1.19	8.31	48.12	1.03	16000	256	8	0.25
gz17	10.61	45.05	1.33	8.92	60.21	1.19	32000	1024	2	0.5
gz18	10.31	37.40	1.32	8.24	55.34	1.20	16000	1024	2	1
gz19	10.15	45.53	1.36	9.57	44.29	1.27	16000	1024	4	0.25
gz20	10.65	47.22	1.37	8.93	32.01	1.28	16000	1024	2	0.25
gz21	7.58	109.26	1.49	8.11	94.58	1.44	32000	1024	0.5	0.25
gz22	10.95	42.04	1.32	8.54	54.50	1.11	8000	512	2	1
gz23	10.51	41.56	1.33	8.66	23.06	0.64	8000	1024	8	0.5
gz24	9.59	32.07	1.30	8.81	34.11	1.19	16000	512	64	1
gz25	11.18	43.48	1.40	9.32	27.42	1.33	8000	1024	64	2
gz26	8.82	63.29	1.27	7.34	76.52	1.30	7.8	1024	128	1
gz27	11.51	52.39	1.34	9.82	23.49	1.30	7.8	1024	16	1
gz28	10.62	32.55	1.34	9.49	43.17	1.28	8000	1024	64	0.25
gz29	10.64	42.23	1.36	9.07	22.13	0.76	8000	1024	128	0.25
gz30	10.42	33.50	1.35	8.58	23.25	1.24	500	1024	32	2
gz31	10.84	36.09	1.29	8.28	45.32	1.16	250	256	128	1
gz32	8.96	30.35	1.42	9.43	43.30	1.31	125	1024	32	0.5
gz33	10.45	43.15	1.30	9.19	54.16	1.29	32000	1024	128	0.5
gz34	8.69	43.42	1.30	9.16	28.34	1.14	16000	1024	128	1
gz35	8.14	32.46	1.26	9.59	112.48	1.16	32000	1024	128	2

gz36	10.00	63.10	1.34	8.70	53.49	1.25	32000	1024	128	1
gz37	10.14	32.15	1.35	9.53	44.40	1.21	16000	1024	64	1
gz38	10.01	39.30	1.32	9.54	36.46	1.14	32000	1024	64	0.5
gz39	10.57	44.00	1.34	9.18	34.51	1.21	16000	1024	128	0.25
gz40	8.30	35.56	1.30	9.58	48.43	1.25	32000	1024	128	1
MG1655	7.58	46.47	1.49	9.95	47.33	1.71	7.8	1	0.5	0.25

Table 3: Details about location of isolation, infection source, serotyping, presence of β-lactamases and population clustering of the thirty-nine clinical strains.

Sample	Location	Infection Source	Serotype	*β*-lactamses	Population Cluster
gz1	Moncton		O25:H4		1
gz2	Moncton	Genitourinary	O21:H25		6
gz3	Moncton	Genitourinary	O6:H1	blaTEM-1B	4
gz4	Moncton	Genitourinary	O23:H45		1
gz5	Moncton	Genitourinary	O6:H1	blaTEM-1B	4
gz6	Moncton	Genitourinary	O25:H4	blaTEM-1C	1
gz8	Winnipeg	Respiratory	O9:H17	blaTEM-1B	6
gz9	London	Genitourinary	O9:H9		6
gz10	London	Genitourinary	O25:H4		1
gz11	London	Respiratory	O25:H4		1
gz12	Moncton	Genitourinary	O25:H4	blaTEM-1B	1
gz13	Moncton	Genitourinary	O25:H4		1
gz14	Moncton	Blood	O25:H4	blaTEM-1B	1
gz15	Winnipeg	Respiratory	O1:H6	blaTEM-116-like	6
gz16	Halifax	Internal Wound	O25:H4	blaTEM-1B	1
gz17	Winnipeg	Genitourinary	O25:H4	blaTEM-1B	1
gz18	Montreal	Genitourinary	O25:H4	blaTEM-1B	1
gz19	Montreal	Respiratory	O25:H4	blaTEM-1B	1
gz20	London	Genitourinary	O25:H4		1
gz21	Moncton	Genitourinary	O25:H4	blaTEM-1B	1
gz22	Halifax	Blood	O45:H6	blaTEM-1B	6
gz23	Halifax	Blood	O25:H4	blaTEM-1B	1
gz24	Vancouver	Respiratory	O25:H4	blaCTX-M-14,blaTEM-1B	1
gz25	Toronto	Blood	O16:H5	blaCTX-M-14,blaTEM-1B	3
gz26	Toronto	Blood	O16:H5	blaCTX-M-14,blaTEM-1B	3
gz27	Edmontom	Blood	O4:H5	blaOXA-36-like,blaSHV-2,blaTEM-1B	2
gz28	Toronto	Blood	O7:H45	blaCTX-M-14,blaTEM-1B	6

gz29	Montreal	Blood	H9	blaCTX-M-15,blaOXA-1,blaTEM-1B	6
gz30	Vancouver	Respiratory	O16:H5	blaCTX-M-14,blaTEM-1B	3
gz31	Saskatoon	Blood	O25:H4	blaSHV-12,blaTEM-1B	1
gz32	Toronto	Blood	O45:H15	blaCTX-M-14,blaTEM-1B	6
gz33	Winnepeg	Respiratory	O25:H4	blaCTX-M-15,blaOXA-1	1
gz34	Montreal	Genitourinary	O102	blaCTX-M-15,blaTEM-1B	6
gz35	Vancouver	Blood	O25:H4	blaCTX-M-15,blaOXA-1	1
gz36	Vancouver	Blood	O25:H4	blaCTX-M-15,blaOXA-1	1
gz37	Halifax	Blood	O25:H4	blaCTX-M-15,blaOXA-1	1
gz38	Toronto	Genitourinary	O89:H17	blaOXA-1	6
gz39	Montreal	Blood	O25:H4	blaCTX-M-15,blaOXA-1	1
gz40	Ottawa	Blood	O25:H4	blaCTX-M-15,blaOXA-1,blaTEM-1B	1

Table 4: Details about number of genes present in the accessory genome and net percentage gain and loss for the terminal branches of the phylogeny.

Strain	Number of genes in Accessory Genome	Net Percentage Gain	Net Percentage Loss
gz1	3502	2.25	6.53
gz2	3195	1.37	0
gz3	3627	0.82	0.14
gz4	3188	1.47	0.04
gz5	3518	1.17	0.67
gz6	3774	1.25	0.46
gz8	3519	1.99	1.42
gz9	3276	1.83	0.15
gz10	3501	2.11	0.89
gz11	3518	1.22	0.56
gz12	3743	1.92	0
gz13	3539	2.62	0.2
gz14	3660	2.59	0
gz15	3198	1.08	0.24
gz16	3521	2.44	1.39
gz17	3472	0.04	0
gz18	3528	4	0.07
gz19	3522	1.5	0
gz20	3576	2.39	3.09
gz21	3628	2.18	1.87
gz22	7596	2.05	4.67

gz23	3497	2.61	2.18
gz24	3873	3.99	12.13
gz25	3381	1.79	2.36
gz26	3579	0.76	1.01
gz27	3434	3.03	6.44
gz28	3688	3.7	2.49
gz29	3403	0.72	4.05
gz30	3629	2.53	1.27
gz31	3724	2.32	1.26
gz32	4051	0.81	2.65
gz33	3706	11.73	3.15
gz34	3831	1.19	1.4
gz35	3731	2.42	0.14
gz36	3878	2.5	1.8
gz37	4999	1.84	0
gz38	3182	0.6	0.07
gz39	4019	3.51	0
gz40	3540	1.57	0.16
MG1655	2991	3.76	10.54
E. fergusonii	2510	2.53	10.03
***E. coli* 0157**	4133	1.44	0.18

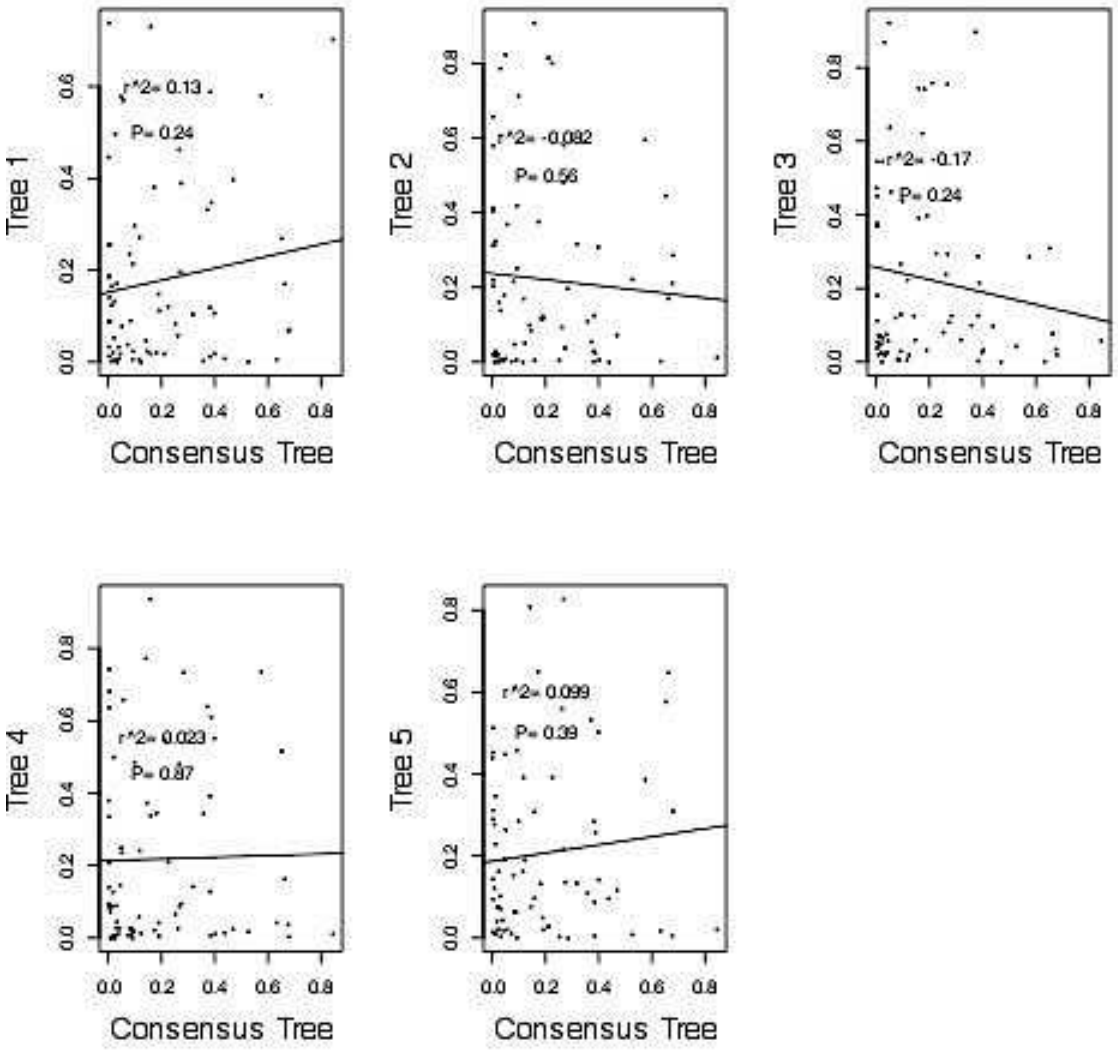

Figure 2: Correlation between data produced by CoEvol using the consensus and top five trees. Results produced by consensus tree and the top 5 trees did not show high correlation.

8 References

Andersson, D. I., & Hughes, D. (2010). Antibiotic resistance and its cost: is it possible to reverse resistance? *Nature Reviews. Microbiology*, *8*(4), 260–271.

Andrews, J. M. (2001). Determination of minimum inhibitory concentrations. *The Journal of Antimicrobial Chemotherapy*, *48 Suppl 1*, 5–16.

Bagel, S., Hüllen, V., Wiedemann, B., & Heisig, P. (1999). Impact of gyrA and parC mutations on quinolone resistance, doubling time, and supercoiling degree of *Escherichia coli*. *Antimicrobial Agents and Chemotherapy*, *43*(4), 868–875.

Bhullar, K., Waglechner, N., Pawlowski, A., Koteva, K., Banks, E. D., Johnston, M. D., Barton, H.A., Wright, G. (2012). Antibiotic Resistance Is Prevalent in an Isolated Cave Microbiome. *PLoS ONE*, *7*(4), e34953.

Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O., & Piddock, L. J. V. (2014). Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology*, *13*(1), 42–51.

Boisvert, S., Laviolette, F., & Corbeil, J. (2010). Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *Journal of Computational Biology*, *17*(11), 1519–1533.

Bolger, A. M., Lohse, M., Usadel, B., Planck, M., Plant, M., & Mühlenberg, A. (2014). Trimmomatic : A flexible trimmer for Illumina Sequence Data, *btu,* 170.

Breidenstein, E. B. M., de la Fuente-Núñez, C., & Hancock, R. E. W. (2011). *Pseudomonas aeruginosa*: All roads lead to resistance. *Trends in Microbiology*, *19*(8), 419–426.

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., Olson, R., Overbeek, R., Parrello, B., Pusch, G.D., Shukla, M., & Xia, F. (2015). RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Scientific Reports*, *5*, 8365.

Bycroft, B. W., & Shute, R. E. (1985). The molecular basis for the mode of action of Beta-lactam antibiotics and mechanisms of resistance. *Pharmaceutical Research*, *2*(1), 3–14.

Cantón, R., González-Alba, J. M., & Galán, J. C. (2012). CTX-M Enzymes: Origin and Diffusion. *Frontiers in Microbiology*, *3*.

Cheng, L., Connor, T. R., Sirén, J., Aanensen, D. M., & Corander, J. (2013). Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Molecular Biology and Evolution*, *30*(5), 1224–8.

Clsi. (2014). M100-S24 Performance Standards for Antimicrobial Susceptibility Testing; Twenty-Fourth Informational Supplement An informational supplement for global application developed through the Clinical and Laboratory Standards Institute consensus process.

Cohen, M. L. (2000). Changing patterns of infectious disease. *Nature*, *406*(6797), 762–7.

D 'costa, V. M., King, C. E., Kalan, L., Morar, M., Sung, W. W. L., Schwarz, C., Froese, D., Zazula, G., Calmels, F. & Wright, G. D. (2011). Antibiotic resistance is ancient. *Nature*, *477*(7365),457-461.

Darling, A. E., Mau, B., & Perna, N. T. (2010). progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE*, *5*(6), e11147.

Davies, J., & Davies, D. (2010). Origins and evolution of antibiotic resistance.

Microbiology and Molecular Biology Reviews, *74*(3), 417–33.

Didelot, X., Darling, A., & Falush, D. (2009). Inferring genomic flux in bacteria. *Genome Research*, *19*(2), 306–17.

Dobrindt, U., & Hacker, J. (2001). Whole genome plasticity in pathogenic bacteria. *Current Opinion in Microbiology*, *4*(5), 550–7.

Drlica, K., & Zhao, X. (1997). DNA gyrase, topoisomerase IV, and the 4-quinolones. *Microbiology and Molecular Biology Reviews*, *61*(3), 377–392.

Drouin, A., Giguère, S., Déraspe, M., Marchand, M., Tyers, M., Loo, V. G., V. G., Bourgault, A.M., Laviolette, & Corbeil, J. (2016). *Predictive computational phenotyping and biomarker discovery using reference-free genome comparisons*. *bioRxiv,* 045153.

Drummond, A. J., Suchard, M. a., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*(8), 1969–1973.

Eaves, D. J., Ricci, V., & Piddock, L. J. V. (2004). Serovar Typhimurium: Role in Multiple Antibiotic Resistance. *Antimicrobial Agents and Chemotherapy*, *48*(4), 1145–1150.

Fàbrega, A., Martin, R. G., Rosner, J. L., Tavio, M. M., & Vila, J. (2010). Constitutive SoxS Expression in a Fluoroquinolone-Resistant Strain with a Truncated SoxR Protein and Identification of a New Member of the marA-soxS-rob Regulon, mdtG. *Antimicrobial Agents and Chemotherapy*, *54*(3), 1218–1225.

Fèbrega, A., Madurga, S., Giralt, E., & Vila, J. (2009). Mechanism of action of and resistance to quinolones. *Microbial Biotechnology*, *2*(1), 40–61.

Finland, M. (1955). Emergence of antibiotic-resistant bacteria. *The New England Journal of Medicine*, *253*(21), 909–22.

García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L. M., Götz, S., Tarazona, S., S., Dopazo, J., Meyer, T.F., & Conesa, A. (2012). Qualimap: Evaluating next-generation sequencing alignment data. *Bioinformatics*, *28*(20), 2678–2679.

Gardner, S. N., & Hall, B. G. (2013). When Whole-Genome Alignments Just Won't Work: kSNP v2 Software for Alignment-Free SNP Discovery and Phylogenetics of Hundreds of Microbial Genomes. *PLoS ONE*, *8*(12), e81760.

Gardner, S. N., & Slezak, T. (2010). Scalable SNP Analyses of 100+ Bacterial or Viral Genomes. *Journal of Forensic Research*, *01*, 107.

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–5.

Hall, B. G. (2014). SNP-Associations and Phenotype Predictions from Hundreds of Microbial Genomes without Genome Alignments. *PLoS ONE*, *9*(2), e90490.

Hanson, N. D. (2003). AmpC *β*-lactamases: what do we need to know for the future? *Journal of Antimicrobial Chemotherapy*, *52*, 2–4.

Hawkey, P. M., & Jones, A. M. (2009). The changing epidemiology of resistance. *Journal of Antimicrobial Chemotherapy*, *64*(Suppl 1), i3-i10.

Heisig, P., Schedletzky, H., & Falkenstein-Paul, H. (1993). Mutations in the gyrA Gene of a Highly Fluoroquinolone- Resistant Clinical Isolate of *Escherichia coli*. *Antimicrobial Agents and Chemotherapy*, *37*(4), 696–701.

Hopkins, K. L., Davies, R. H., & Threlfall, E. J. (2005). Mechanisms of quinolone resistance in *Escherichia coli* and *Salmonella*: Recent developments. *International*

Journal of Antimicrobial Agents, *25*, 358–373.

Jacoby, G. A. (2005). Mechanisms of Resistance to Quinolones. *Clinical Infectious Diseases*, *41*, S120–S126.

Jacoby, G. A. (2009). AmpC beta-lactamases. *Clinical Microbiology Reviews*, *22*(1), 161–82.

Joensen, K. G., Tetzschner, A. M. M., Iguchi, A., Aarestrup, F. M., & Scheutz, F. (2015). Rapid and Easy In Silico Serotyping of Escherichia coli Isolates by Use of Whole-Genome Sequencing Data. *Journal of Clinical Microbiology*, *53*(8), 2410–26.

Keseler, I. M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martínez, C., C., Fulcher, C., Huerta, A.M., Kothari, A., Krummenacker, M. & Latendresse, M. (2013). EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Research*, *41*, D605–12.

Kugelberg, E., Löfmark, S., Wretlind, B., & Andersson, D. I. (2005). Reduction of the fitness burden of quinolone resistance in *Pseudomonas aeruginosa*. *Journal of Antimicrobial Chemotherapy*, *55*(1), 22–30.

Lagacé-Wiens, P. R. S., Adam, H. J., Low, D. E., Blondeau, J. M., Baxter, M. R., Denisuik, A. J., Nichol, K.A., Walkty, A., Karlowsky, J.A., Mulvey, M.R., Hoban, D.J., & Zhanel, G. G. (2013). Trends in antibiotic resistance over time among pathogens from Canadian hospitals: results of the CANWARD study 2007-11. *The Journal of Antimicrobial Chemotherapy*, *68 Suppl 1*, i23–9.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25.

Lartillot, N., & Poujol, R. (2011). A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. *Molecular Biology and Evolution*, *28*(1), 729–744.

Levy, S. (1982). Microbial Resistance to Antibiotcs. *The Lancet*, *320*(8289), 83–88.

Levy, S. B., & Marshall, B. (2004). Antibacterial resistance worldwide: causes, challenges and responses. *Nature Medicine*, *10*(12 Suppl), S122–9.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Lin, Y., Li, J., Shen, H., Zhang, L., Papasian, C. J., & Deng, H.-W. (2011). Comparative studies of de novo assembly tools for next-generation sequencing technologies. *Bioinformatics (Oxford, England)*, *27*(15), 2031–7.

MacLean, R. C., Hall, A. R., Perron, G. G., & Buckling, A. (2010). The Evolution of Antibiotic Resistance: Insight into the Roles of Molecular Mechanisms of Resistance and Treatment Context. *Discovery Medicine*, *10*(51), 112–118.

Manhart, C. M., & McHenry, C. S. (2015). Identification of Subunit Binding Positions on a Model Fork and Displacements That Occur during Sequential Assembly of the *Escherichia coli* Primosome. *Journal of Biological Chemistry*, *290*(17), 10828–10839.

Melnyk, A. H., Wong, A., & Kassen, R. (2014). The fitness costs of antibiotic resistance mutations. *Evolutionary Applications*, *8*(3), 273–283.

Mi, H., Muruganujan, A., Casagrande, J. T., & Thomas, P. D. (2013). Large-scale gene function analysis with the PANTHER classification system. *Nature Protocols*, *8*(8),

1551–1566.

Nicolas-Chanoine, M.-H., Bertrand, X., & Madec, J.-Y. (2014). *Escherichia coli* ST131, an Intriguing Clonal Group. *Clinical Microbiology Reviews*, *27*(3), 543–574.

Nishino, K., & Yamaguchi, A. (2001). Analysis of a Complete Library of Putative Drug Transporter Genes in *Escherichia coli*. *Journal of Bacteriology*, *183*(20), 5803–5812.

Olesen, B., Hansen, D. S., Nilsson, F., Frimodt-Møller, J., Leihof, R. F., Struve, C., Scheutz, F., Johnston, B., Krogfelt, K.A., & Johnson, J. R. (2013). Prevalence and Characteristics of the Epidemic Multiresistant *Escherichia coli* ST131 Clonal Group among Extended-Spectrum Beta- Lactamase-Producing *E. coli* Isolates in Copenhagen, Denmark. *Journal of Clinical Microbiolog*, *51*(6), 1779–1785.

Palmer, A. C., & Kishony, R. (2013). Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance. *Nature Reviews. Genetics*, *14*(4), 243–8.

Patel, R. K., & Jain, M. (2012). NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PloS One*, *7*(2), e30619.

Peirano, G., & Pitout, J. D. D. (2010). Molecular epidemiology of *Escherichia coli* producing CTX-M β-lactamases: the worldwide emergence of clone ST131 O25:H4. *International Journal of Antimicrobial Agents*, *35*(4), 316–321.

Phan, M. D., Forde, B. M., Peters, K. M., Sarkar, S., Hancock, S., Stanton-Cook, M., Zakour, N.L.B., Upton, M., Beatson, S.A., & Schembri, M. A. (2015). Molecular Characterization of a Multidrug Resistance IncF Plasmid from the Globally Disseminated *Escherichia coli* ST131 Clone. *PLoS ONE*, *10*(4), e0122369.

Pitout, J. D. D., Nordmann, P., Laupland, K. B., & Poirel, L. (2005). Emergence of

Enterobacteriaceae producing extended-spectrum β-lactamases (ESBLs) in the community. *Journal of Antimicrobial Chemotherapy*, *56*, 52–59.

Roca, I., Akova, M., Baquero, F., Carlet, J., Cavaleri, M., Coenen, S., Cohen, J., Findlay, D., Gyssens, I., Heure, O.E., Kahlmeter, G., & Vila, J. (2015). The global threat of antimicrobial resistance: Science for intervention. *New Microbes and New Infections*, *6*, 22–9.

Ruiz, J. (2003). Mechanisms of resistance to quinolones: target alterations, decreased accumulation and DNA gyrase protection. *Journal of Antimicrobial Chemotherapy*, *51*, 1109–1117.

Saier, M. H., Tam, R., Reizer, A., & Reizer, J. (1994). Two novel families of bacterial membrane proteins concerned with nodulation, cell division and transport. *Molecular Microbiology*, *11*(5), 841–7.

Santiago-Rodriguez, T. M., Fornaciari, G., Luciani, S., Dowd, S. E., Toranzos, G. A., Marota, I., & Cano, R. J. (2015). Gut Microbiome of an 11 th Century A.D. Pre-Columbian Andean Mummy, *10*(9), e0138135.

Swick, M. C., Morgan-Linnell, S. K., Carlson, K. M., & Zechiedrich, L. (2011). Expression of Multidrug Efflux Pump Genes acrAB-tolC, mdfA, and norE in *Escherichia coli* Clinical Isolates as a Function of Fluoroquinolone and Multidrug Resistance. *Antimicrobial Agents and Chemotherapy, 55*(2), 921–924.

Tamae, C., Liu, A., Kim, K., Sitz, D., Hong, J., Becket, E., Bui, A., Solaimani, P., Tran, K.P., Yang, H., & Miller, J. H. (2008). Determination of antibiotic hypersensitivity among 4,000 single-gene-knockout mutants of *Escherichia coli*. *Journal of Bacteriology*, *190*(17), 5981–5988.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, *30*(12), 2725–9.

Tenover, F. C. (2006). Mechanisms of antimicrobial resistance in bacteria. *American Journal of Infection Control*, *34*(5 Suppl 1), S3–10.

Vetting, M. W., Chi, |, Park, H., Hegde, S. S., Jacoby, G. A., Hooper, D. C., & Blanchard, J. S. (2008). Mechanistic and Structural Analysis of Aminoglycoside N-Acetyltransferase AAC(6′)-Ib and Its Bifunctional, Fluoroquinolone-Active AAC(6′)-Ib-cr Variant. *Biochemistry*, *47*(37), 9823–9835.

Vetting, M. W., Hegde, S. S., Wang, M., Jacoby, G. A., Hooper, D. C., & Blanchard, J. S. (2011). Structure of QnrB1, a plasmid-mediated fluoroquinolone resistance factor. *The Journal of Biological Chemistry*, *286*(28), 25265–73.

Wang, L., Curd, H., Qu, W., & Reeves, P. R. (1998). Sequencing of *Escherichia coli* O111 O-antigen gene cluster and identification of O111-specific genes. *Journal of Clinical Microbiology*, *36*(11), 3182–7.

Wang, L., Rothemund, D., Curd, H., & Reeves, P. R. (2003). Species-wide variation in the *Escherichia coli* flagellin (H-antigen) gene. *Journal of Bacteriology*, *185*(9), 2936–43.

Willems, R. J. L., Top, J., van Schaik, W., Leavis, H., Bonten, M., Sirén, J., … Corander, J. (2012). Restricted gene flow among hospital subpopulations of Enterococcus faecium. *mBio*, *3*(4), e00151–12.

Wong, A., Rodrigue, N., & Kassen, R. (2012). Genomics of Adaptation during Experimental Evolution of the Opportunistic Pathogen *Pseudomonas aeruginosa*.

PLoS Genetics, *8*(9), e1002928.

Wren, B. W. (2000). Microbial genome analysis: insights into virulence, host adaptation and evolution. *Nature Reviews. Genetics*, *1*(1), 30–9.

Yotsuji, A., Mitsuyama, J., Hori, R., Yasuda, T., Saikawa, I., Inoue, M., & Mitsuhashi, S. (1988). Mechanism of action of cephalosporins and resistance caused by decreased affinity for penicillin-binding proteins in Bacteroides fragilis. *Antimicrobial Agents and Chemotherapy*, *32*(12), 1848–53.

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M., & Larsen, M. V. (2012). Identification of acquired antimicrobial resistance genes. *The Journal of Antimicrobial Chemotherapy*, *67*(11), 2640–4.

Zayed, A. A.-F., Essam, T. M., Hashem, A.-G. M., & El-Tayeb, O. M. (2015). "Supermutators" found amongst highly levofloxacin-resistant E. coli isolates: a rapid protocol for the detection of mutation sites. *Emerging Microbes & Infections*, *4*(1), e4.

Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, *18*(5), 821–9.

Zhang, J., Nielsen, R., & Yang, Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution*, *22*(12), 2472–2479.