A genome-based approach to study the ecology and epidemiology of Extended-Spectrum Beta-Lactamases (ESBLs) producing *E. coli*

Doctoral thesis for a doctoral degree

Submitted to the Faculty of Medicine

in partial fulfilment of the requirements

for the Ph.D. degree

of the Faculties of Veterinary Medicine and Medicine

of the Justus-Liebig-University Giessen, Germany

By

Hiren Ghosh

of

Shalpara, India.

From Institute of Medical Microbiology

Director: Prof. Dr. Trinad Chakraborty

Faculty of Medicine

Justus-Liebig-University Giessen

Supervisor: Prof. Dr. Trinad Chakraborty

Co-Supervisor: Prof. Dr. Christoph Lämmler

Contents

1. List of publications

During his doctoral work, the author was involved with the following publications, which are part of this thesis.

1. Detection of translocatable units in a *bla*$_{\text{CTX-M-15}}$ extended-spectrum β-lactamase-producing ST131 *Escherichia coli* isolate using a hybrid sequencing approach.

Hiren Ghosh, Swapnil Doijad, Boyke Bunk, Linda Falgenhauer, Yancheng Yao, Cathrin Spröer, Katrin Gentil, Judith Schmiedel, Can Imirzalioglu, Jörg Overmann and Trinad Chakraborty. International Journal of Antimicrobial Agents, 2016 ;47(3):245–247

2. Emergence of *bla*$_{\text{CTX-M-27}}$-encoding *Escherichia coli* ST131 lineage C1/M27 clone in clinical isolates in Germany

Hiren Ghosh, Swapnil Doijad, Linda Falgenhauer, Moritz Fritzenwanker, Can Imirzalioglu and Trinad Chakraborty* Emerging Infectious Diseases 2017; 23 (10):1754-1756.

3. Complete Genome Sequence of *bla*$_{\text{CTX-M-27}}$ -Encoding *Escherichia coli* Strain H105 of ST131 Lineage C1/H30R

Hiren Ghosh, Boyke Bunk, Swapnil Doijad, Judith Schmiedel, Linda Falgenhauer, Cathrin Spröer, Can Imirzalioglu, Jörg Overmann, Trinad Chakraborty* Genome Announc. 2017; 5(31): e00736-17

4. Circulation of clonal populations of fluoroquinolone-resistant CTX-M-15-producing *Escherichia coli* ST410 in humans and animals in Germany

Linda Falgenhauer, Can Imirzalioglu, **Hiren Ghosh**, Konrad Gwozdzinski, Judith Schmiedel, [..], Trinad Chakraborty Int J Antimicrob Agents. 2016; 47:457–65.

5. Chromosomal Locations of mcr-1 and *bla*$_{\text{CTX-M-15}}$ in Fluoroquinolone-Resistant *Escherichia coli* ST410

Linda Falgenhauer, Said-Elias Waezsada, Konrad Gwozdzinski, **Hiren Ghosh**, Swapnil Doijad, [..], Trinad Chakraborty* Emerg Infect Dis. 2016; 22(9):1689-1691

"In such cases, the thoughtless person playing with penicillin is morally responsible for the death of the man who finally succumbs to infection with the penicillin-resistant organism. I hope this evil can be averted."

….Sir Alexander Fleming

2. Introduction

2.1. The emergence and mechanism of antimicrobial resistance

Antibiotics are one of the milestone discoveries in the medical science of the twentieth century. Since the 1940s, the successful use of antibiotics has revolutionized the practice of modern medicine. Aside from curing direct bacterial infection antibiotics are playing an important role in surgery, organ transplantation, cancer chemotherapy, and countless lives have saved around the world [1–3]. Unfortunately, in the modern times, antibiotics face a degree of obsolescence because of rapid worldwide appearance and spread of untreatable drug-resistant bacteria [4,5]. Antimicrobial-resistance became one of the global problems, which transcends through national, international boundaries, socio-economic divisions classes and affects human, animal and environmental health equally [4].

In principle, antibiotic resistance developed when bacteria undergo a genetic change to minimize the effects of antibiotics. Traditional mechanisms behind bacterial antimicrobial resistance broadly divided into two major genetic strategies: spontaneous mutation in genes and acquisition of external genetic material via horizontal gene transfer (HGT).

Mutational resistance: Antimicrobial resistance-conferring mutation is a consequence of the multi-step evolutionary process. When a population of microbes are challenged with antibiotics, a subpopulation of them evolves to evade the drug's action due to the mutation-selection pressure rendered upon them, as a consequence previously antibiotic susceptible bacterial population ends up being resistant to the drug [5]. On the other hand, at the same time, antibiotics eliminate the antibiotic-susceptible population from the community and drug-resistant population gradually emerged. In general, the underlying mechanism of antimicrobial resistance occurred via one of the following mechanisms: **I) Alteration of intracellular drug target**: a common strategy of bacteria to develop the antibiotic-resistant mutant by modification

of the drug target site. Few classical examples include a) bacterial resistance developed due to a single step point mutation in 16S rRNA methylase encoding (i.e. *rmt*A or *rmtD*) genes conferring aminoglycoside mutant [6]. b) Another well-characterized example: mutation in chromosomal DNA in gyrase encoding genes (i.e. *gyr*A/*gyr*B), which resulted in the fluoroquinolone-resistant mutant [7]. c) Mutation in topoisomerase encoding chromosomal genes (i.e. *parC*/*parE*) also associated with quinolones resistant mutant [8]. d) Finally, another good example, the alternation of penicillin-binding protein (PBP) reduced the affinity of beta-lactams group of antibiotic and give rise to penicillin-resistant bacteria [9] (Figure1). **II) Alteration of membrane permeability:** another clinically significant mechanism in gram-negative bacteria in preventing antibiotic influx to the intracellular compartment through the modification of permeability of the membrane. One classical example of the porin mediates resistant generally observed in clinically *Pseudomonas. aeruginosa*, where due to lack of OprD porins shows resistance to imipenem [10] In few cases *Klebsiella. pneumonia* with porin mutant has caused the global outbreak also. (Figure 2.1). **III) Increased Efflux pumps:** Overexpression of the bacterial efflux pump, actively transport antibiotic out of the cell and produce a drug-resistant phenotype. Another example, observes in tetracycline resistant gram-negative bacteria, where tetracycline pumps out from the cell by Tet efflux pumps. (Figure 2.1**)**. **IV) Overproduction of antibiotic-inactivating enzymes:** The production of specific enzyme molecules capable to change the antibiotic molecule is also a well-known mechanism in both gram positive and negative bacteria. Few classical examples include a) the inactivation of aminoglycosides by aminoglycoside-modifying enzymes (AMEs) by Acinetobacter spp. [11] b) Trimethoprim resistant resulted by the overproduction of dihydrofolate reductase. (Figure 2.1). **V) Alternation of global metabolic pathways:** Through the years of evolution bacterial phenotypic resistance mechanism may develop by altering global metabolic pathway that bypasses the antibiotic mechanism. For example, certain sulfonamide-resistant bacteria do not require extracellular para-aminobenzoic acid (PABA) instead, like mammalian cells, they turn to utilizing preformed folic acid [12]. The emergence of antimicrobial resistance can be explained by using Darwinian's survival of the fittest model, i.e., beneficial mutations occurred in pre-existing genes of the bacterial chromosome that are positively selected by environmental forces. However, the exact nature of the drug resistance mechanisms still poorly understood, because of few additional unidentified mechanisms also involved in antibiotic

resistance. Research strategies to effectively combat antimicrobial resistance include the identification of potential resistance genes, as well as the expression pattern of those genes by using multi-omics approaches (e.g. RNA-seq) are needed.

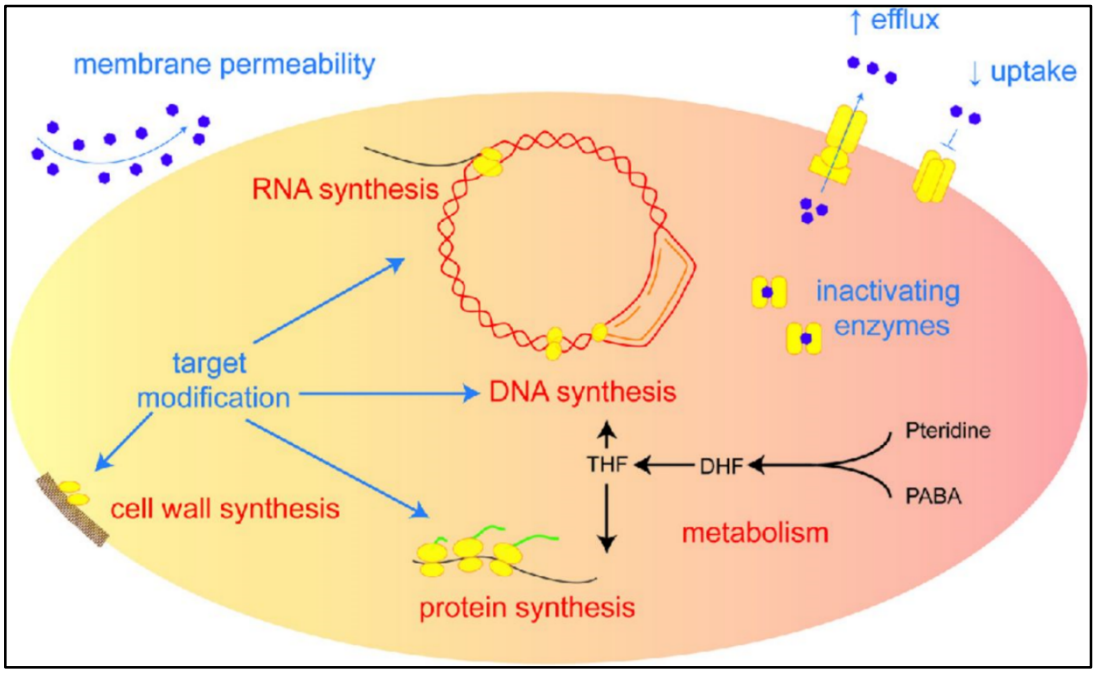


Figure 2.1: Diagrammatic representation of the different mechanism by which bacteria can develop resistance to antibiotics (Source: Park J.A, 2016). Example included cellular targets for antibiotics include DNA replication (e.g. fluoroquinalones), alternation of protein synthesis (e.g. aminoglycosides), cell wall integrity changes (e.g. pencillins) and folic acid metabolism (e.g. sulfonamides).

Acquisition of external genetic material via horizontal gene transfer: Emergence and the dissemination of antibiotic resistance may develop vertically by the mutation in the chromosomal gene followed by selection pressure. Bacterial resistance may also occur by acquiring foreign resistance genes from other bacteria via horizontal gene transfer [13,14]. The HGT mechanism can promote commensal and environmental nonpathogenic isolate to antibiotic-resistant ones through one of the following modes of genetic exchange [15].

Conjugation (sexual mode of genetic transfer)**:** Bacterial conjugation is one of the principal conduits for direct gene transfer in between two compatible bacterial cells. During this process, the transfer of DNA occurred through a multi-step contact-dependent process membrane-associated macromolecular machinery called Type IV secretion system (T4SS). Conjugative machinery may enable the mobilization of plasmids mediated AMR gene transfer in many types of ecosystems. A well-demonstrated conjugation mediates antibiotic resistance gene dissemination of

bla$_{CTX-M}$ ESBL genes, which have disseminated to various narrow and broad host range plasmids within Enterobacteriaceae [16] (Figure 2.2).

Transformation (incorporation of naked DNA): Certain bacteria appeared to be capable of direct uptake and integration fragments of extracellular DNA from the environment by homologous recombination, a process called transformation [17]. Transformation is not only a useful method in molecular cloning to produce multiple copies of a recombinant DNA molecule, but also a powerful mechanism of gene transfer in a natural bacterial population [18]. Transformation mediated antibiotic resistance gene dissemination has well documented in many gram-negative bacteria [19] (Figure 2.2).

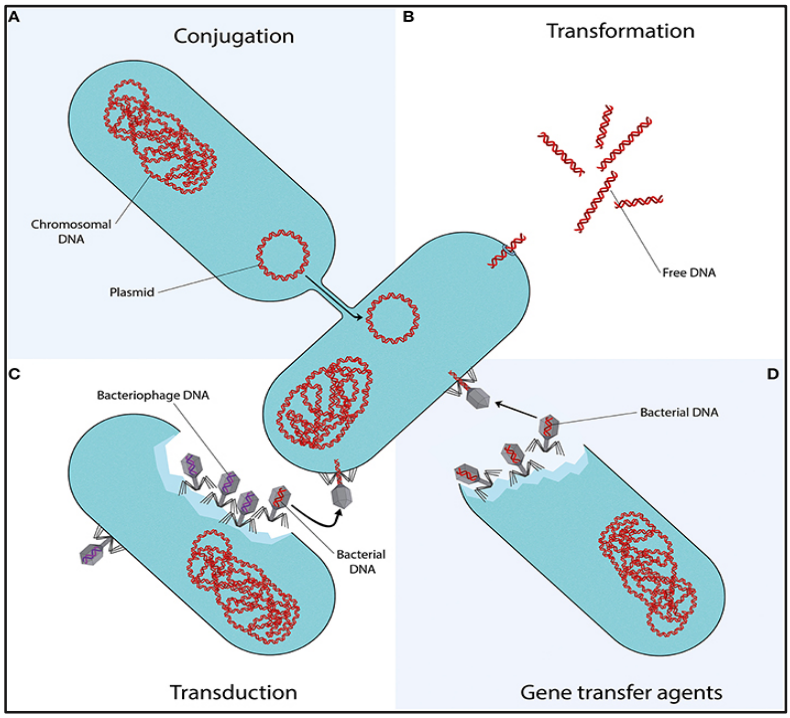


Figure 2.2: Mechanisms of horizontal gene transfer (Source: Christian, 2016).

Transduction (bacteriophage mediated): Bacteriophages are viruses that infect bacteria and play an important role in bacterial evolution in the different ecological niche. Transduction is facilitated by transferring virulence and antimicrobial resistance genes to new bacterial hosts. Through specialized or generalized transduction, bacteriophages can transfer genes that are advantageous to their microbial hosts, in turn promoting their own survival and dissemination. The mobilization or transfer of antibiotic resistance genes (ARGs) by bacteriophages has been documented for various bacterial species e.g. *E. coli, Salmonella* [20] (Figure 2.2).

Since the last few decades, the worldwide drug resistance scenario has been accelerated dramatically because of overuse of antibiotics, and on the other hand, the new classes of antibiotics have been not discovered. According to recent surveillance reports, antibiotic resistance currently incurs an estimated loss of around 50,000 lives in Europe and the United States, costing to estimated €2.5 billion annually [21,22]. In fact, the World Health Organization (WHO) has warned antibiotic resistance to be one of the most complex public health threats of the 21st century [23]. For many years, the population structure of common gram-negative bacteria such as *E. coli*, *K. pneumonia* and *Acinetobacter baumannii* were relatively predictable and stable over time**.** During the 1980s the rise of multi-drug resistant gram-negative bacteria (MDRGN) from Enterobacteriaceae family that produce extended-spectrum β-lactamase (ESBLs) or carbapenemase enzymes are spread throughout the world. Among these groups of bacteria, *E. coli* are of particular clinical concerned worldwide. *E. coli* can survive a wide range of hosts and rapidly acquire resistance to key antibiotics and represent one of the most prominent threats for the successful treatment of infections caused by these bacteria. Moreover, pathogenic *E. coli* possessing resistance to other antimicrobial agents, including aminoglycosides, sulphonamides and fluoroquinolones are associated with significant morbidity and mortality in human and animal healthcare [24–26]. Nevertheless, misuse and overuse of antibiotics in human medicine, agriculture, and veterinary medicine are primarily accelerating the process of antibiotic resistance. Other factors such as wastewater treatment plants, food, and the environment also plays a crucial role in its spread.

2.2. Ecological compartments and evolution of antibiotic resistance

The traditional healthcare organization system is divided into two major compartments, e.g. community and hospital [27,28]. Both the compartments are inextricably linked to the dissemination of antibiotic resistance. Therefore, to combating antibiotic resistance a “One Health” conceptual framework that could promote the integration of public health, food, environment, and animal health surveillance approaches are needed

2.2.1. The hospital compartment

Hospital environments have traditionally been assumed mainly to serve as the incubator for the breeding of antibiotic-resistant bacteria. Here, bacteria proliferate in an environment, filled with immunocompromised patients, and antibiotic pressure is continuously eliminating the drug-susceptible bacteria. Numerous clinical studies have

demonstrated that the human intestinal tract is most complex and densely populated microbial ecosystem, which act as an important reservoir for many gram-negative bacteria. Human gut also offers ample opportunities for bacteria to gene flow via the horizontal transfer, including antibiotic resistance genes [29]. In general, exposure to antibiotics does not induce resistance mechanism but in a few cases, long-term antibiotic therapy promotes proliferation of the emergence of new resistant mutants by accelerating the selection pressure [30]. Once resistant pathogens have evolved and subsequent spread from patient to patient and resulted as an untreatable pan-resistant pathogen [31]. Few overarching factors that drive antimicrobial-drug resistance in hospital-acquired infections are described below:

I). Poor infection control and hygiene in health care settings:

Studies repeatedly reported that several antimicrobial resistant outbreaks are directly associated with lack of access to safe infection control in hospital settings. The most frequently described organisms associated with nosocomial infections are “ESKAPE” pathogens *(Enterococcus faecium*, *Staphylococcus aureus*, *K. pneumoniae*, *Acinetobacter baumannii*, *P. aeruginosa*, and *Enterobacter* species). Furthermore, most of the ESKAPE pathogens are often multidrug resistant, which becomes a major challenge in clinical practice due to the high susceptibility to immunocompromised or elderly patients [32]. Hand-hygiene practices and inadequate sanitary conditions are playing a major role in the dissemination of ESKAPE pathogens. Appropriate hand hygiene can minimize microorganisms acquired on the hands of the healthcare workers during daily duties when there is a direct contact with blood, body fluids, secretions, excretions and known and unknown contaminated equipment or surfaces of medical devices [33].

II). Colonization pressure and selection:

The role of antibiotic consumption in the hospital is promoting the influence of “selection density” in a geographic area. In the hospital, the density-dependent selection is very high because of a depleted choice of antibiotics and a reduced diversification strategy. Furthermore, the accumulation of the few patients in small spaces such as intensive-care unit (ICU) or specific wards and to a rapid resistant clone circulation within such a population also contributing the role in selection density. Considered these above facts together, excess or overprescribed antibiotics and lack of guidance to use antibiotics also play a major role in spreading drug-resistant bacteria [34] (Figure 2.3).

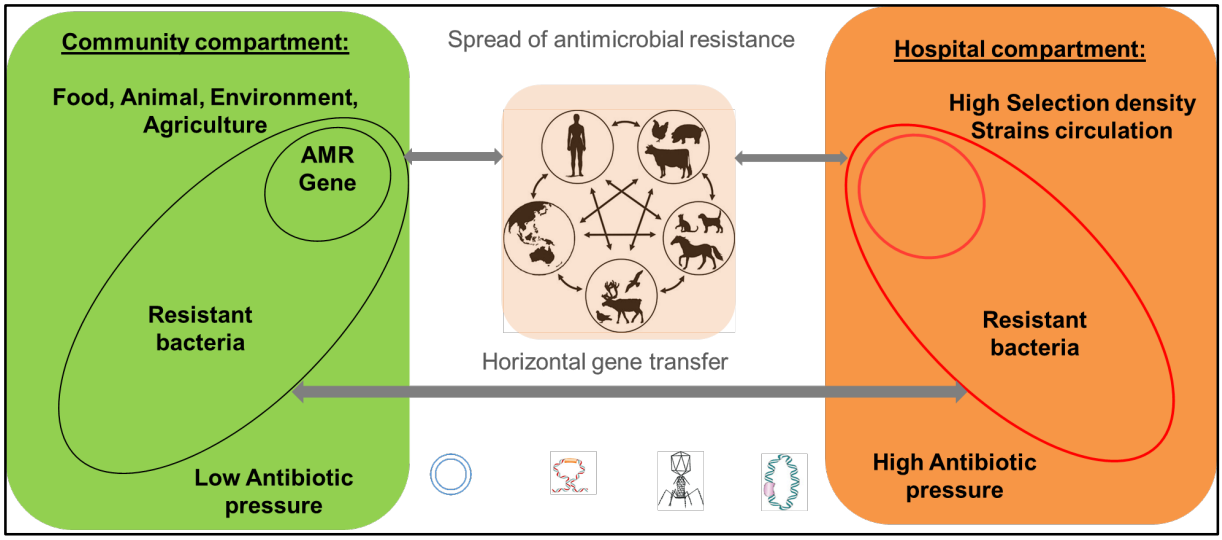


Figure 2.3: Diagrammatic representation of the transfer of antimicrobial resistance bacteria among different compartments (Source: Cantón, 2013 and Stokes, HW 2011). Arrows represent the routes of transmission of AMR between the community and hospital compartments.

2.2.2. Community compartments

Classically, the "environmental resistome" gene pool comprises both from human activities and from ecological, environmental factors. It is hypothesized, that when identical antibiotic-resistant bacteria (ARB) that present in environmentally and clinically relevant pathogens, probably represents a risk of the transmission cycle. These environmental present antibiotic-resistant bacteria are assumed as vast reservoirs, which comprise the phylogenetically diverse group of bacteria. Thus, some intermediary agents or vectors are required to complete the transmission cycle, e.g. via food chain or via mobile genetic elements [35]. The dissemination of ARB across the environment compartment is caused by two major mechanisms: HGT; mutation and genetic recombination [36]. However, it is unlikely that all antibiotic resistance genes (ARGs) that has arisen and spread in clinical settings are always being evolved and transferred from environmental bacteria. Environmental hotspots, where ARB are abundant or the transfer of ARGs may be promoted, are critical points for resistance control. These locations comprise habitats that are influenced by human activities, such as wastewater treatment plants (e.g. Hospital, urban and industrial), waste and intensive food-production facilities such as agriculture purposes or animal husbandry [37]. It is well established that antibiotic selective pressure is the main causative agents for the emergence of dangerous superbugs in both hospital and community compartments. Recent studies showed in the evolutionary scale, the existence of antimicrobial resistance genes in environment hundreds of millions of years ago. In a

recent study has reported the presence of antibiotic resistance gene in Yanomami tribes of the Amazon jungle of Venezuela who had never been exposed to antibiotics [38]. Earlier studies have observed the presence of the ABR gene in the gut bacteria inside 1000-year-old mummies from the Inca Empire is resistant to most of today's antibiotics [39]. Overall, the evidence suggests that microbes have long evolved the capability to counteract toxins, including antibiotics, and that preventing drug resistance may be harder than scientists thought (Figure 2.3).

2.3. Mobilization of antibiotic resistance genes via mobile genetic elements

Mobile genetic elements (MGEs) are segments of DNA that can play a pivotal role in the dissemination of resistance, virulence, and adaptation factors, in between bacteria, and even within species. Traditionally, MGEs classified as into different categories such as plasmids, transposons, bacteriophages, integrative-conjugative elements. The major MGEs involved in the spread and expression of clinically important antimicrobial resistance in bacteria are described below.

Plasmids: Plasmids are extrachromosomal circular or linear genetic element that occurs in many bacterial strains and can replicate independently of the bacterial chromosome. Plasmids are extraordinarily versatile in their size and gene content and are readily transmissible between host cells. Plasmids typically harbor a wide spectrum of beneficial host traits, such as antibiotic resistance, heavy-metal-resistance genes and adaptation factor for the survival of the different ecological niche. Conjugative plasmids serve a central role as the vehicle for the dissemination of antibiotic resistance gene and directly linked with many outbreaks [40]. Plasmid classification scheme provides useful information insight into the epidemiology and antibiotic resistance gene transfer. Based on the replicon typing, plasmid has been classified into several incompatibility (Inc) groups, refers to the inability of two plasmids to be sharing similar replicon and partition system. Conventional single-locus based typing schemes have been widely used for a long time; they provide low-resolution compare to multiple loci. Currently in the genomics era based on genome sequence high-resolution classification scheme also available such as plasmid multi-locus sequence typing (pMLST) schemes or based on core genome single nucleotide polymorphism (SNPs) phylogenetic approaches. Due to high genomic plasticity and diverseness plasmid subtyping always challenging [41].

Transposable elements (TEs): TEs are also known as "jumping genes" or transposons that are scattered throughout the genome of an organism. Transposons are DNA sequence that can exchange a linear piece of DNA with varying degrees of site selectivity that ranging size from ~2.5 to ~25 kb. In some transposons, the inverted terminal repeats presents with the lack of additional gene sequences and is capable of independent movement, called insertion sequences (IS element). Some transposable elements excise from the original site and insert into the new site (cut and paste), whereas others use replicative mechanisms to create a copy at a new site. In case of bacteria, mostly transposon can jump from chromosomal DNA to a plasmid or vice versa. Transposons carry a gene that encodes transposase(s), the enzyme(s) responsible for recombination of the transposon into another DNA molecule. Studies on transposable genetic elements in bacteria have not only given insight into the spread of antibiotic resistance but also in the process of DNA movement [42,43] (Figure 2.4).

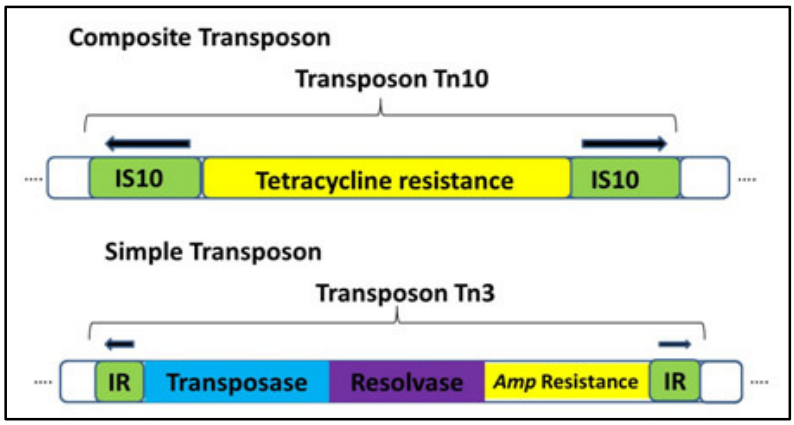


Figure 2.4: Detailed structure of Transposable elements with an example Tn10 and Tn3 transposon (Source: Gaikwad B. K, 2015).

Bacteriophages: Bacteriophages or phages are viruses that consist of a DNA or RNA genome surrounded by a protein coat (capsid) and invade the bacterial cell. They are the most abundant and diverse biological entities in the biosphere, with a huge population structure (~10^{30}–10^{32}). Phages are another major vehicle to shuttle DNA between bacteria, numerous phages are also capable of transducing plasmids, including those encoding antibiotic resistance. Phages act as an important vechiles for genetic exchange via generalized or specialized transduction.There are several genetic elements composed of phages or phage-related elements, which play a major role in the mobilization of genes in bacteria [44,45].

Integrative and Conjugative Elements (ICEs): are diverse groups of mobile genetic elements found in both gram-positive and gram-negative bacteria. These genetic elements primarily reside in a host chromosome, but retain the ability to excise and to transfer by conjugation. ICEs use a range of mechanisms to promote their core functions of integration; excision, transfer, and regulation, there are common features that unify the group. A set of genes in ICEs enables conjugative transfer and control of element removal and integration into the host chromosome. These features indicate that ICEs are directly involved in the processes of horizontal transfer of genetic determinants, which increase the adaptive potential of bacterial species, and can function as universal mobilizing factors for other genetic elements. Mobile between cells using conjugation machinery and able to integrate into DNA sites via site-specific recombination [46] (Figure 2.5).

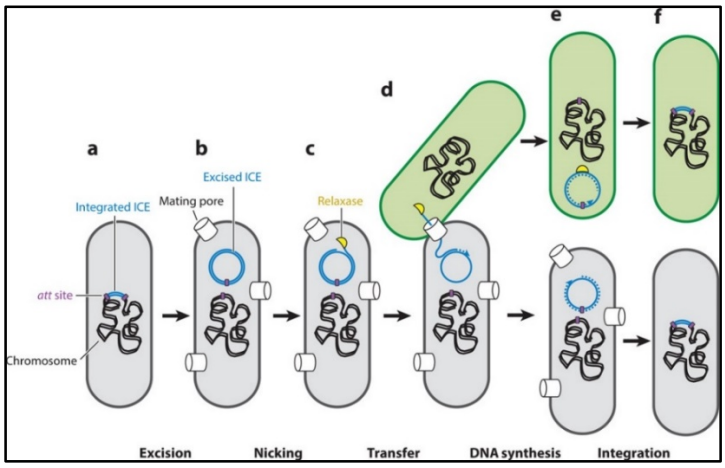


Figure 2.5: Schematic representation of life cycle of integrative and conjugative element (Source Johnson CM, 2015).

Translocatable units (TUs): The insertion sequence IS*26* plays a key role in disseminating antibiotic resistance genes in Gram-negative bacteria, forming regions containing more than one antibiotic resistance genes that are flanked by a single copy of IS*26*. "A composite transposon is a mobile genetic element consisting of two insertion sequences (ISs) flanking a segment of cargo DNA often containing antibiotic resistance (AR) genes are termed as Tus" [47][48]. In this mechanism, the donor IS26 site recognize another IS26 in the recipient site and via RecA-independent, genetic rearrangement happens. The rearrangement of TUs is schematically shown below (Figure 2.6).

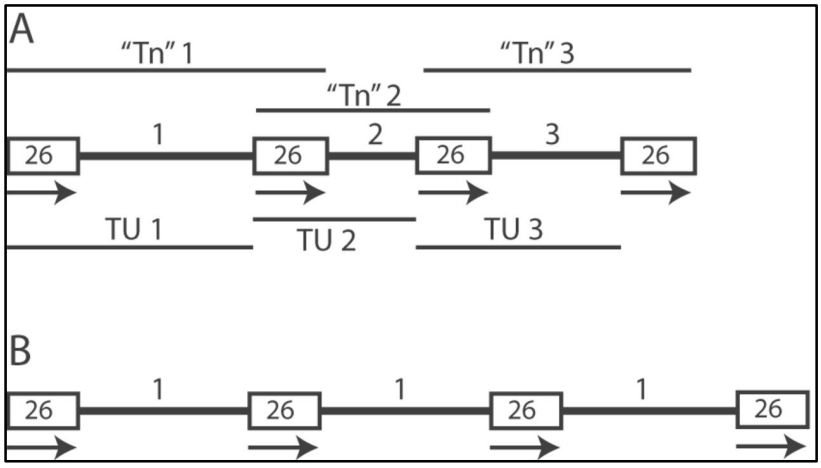


Figure 2.6: Diagrammatic representation of formation of translocatable units. Here boxes and line represents IS26s carrier DNA segments (Source Harmer JC, 2014).

2.4. Decline the effectiveness of the antibiotic: conservation vs innovation

Despite the real and growing public health threat posed by ongoing explosion antibiotic resistance, in recent decades, the discovery and development of new antibiotics have slowed down dramatically. Simultaneously the emergence of antibiotic-resistant pathogens has accelerated, allowing for life-threatening infections that will not respond to available antibiotic treatment. We are in the midst of the crisis of the antibiotic, which similar to managing others natural resources, such as oil, water, and fisheries. Excessive use and misuse of antibiotics in human and animal health are leading to global warnings that in the near future without effective antimicrobials for prevention and treatment of infections. However, to maintain long-term of the effectiveness of antibiotic, individual patients, doctors, pharmaceutical companies, hospitals, and even countries have little incentive to use antibiotics. The decline of the effectiveness of the antibiotic depends on a combination of the following reasons [49,50].

Human responsibility: Although antibiotic resistance evolves naturally via natural selection through random mutation, human activity is playing a major role to accelerate the promotion of resistance. The misuse of antibiotics and biocides in human or animal healthcare practice is a significant driver of the emergence and spread of antibiotic-resistant bacteria. The extensive use of the broad spectrum of antibiotics in agricultural applications has also promoted the development of dangerous extreme drug-resistant bacteria in the soil with the risk for a further circulation to human or animal via direct contact or the food chain. Urbanization and global travel are also intensifying the spread of bacterial infections. According to IMS Health MIDAS report, antibiotic

consumption of the human in low and middle–income countries have increased 36% between 2000 and 2011, mostly last resort classes of antibiotics e.g. carbapenems, and polymyxins. As a consequence, today more than 70% of pathogenic bacteria are drug resistant [50–52].

Innovation failure: One major reason for the declining trends in antibiotic research is that over the past decades the resistance rate is spreading faster than the introduction of new antibiotics in the clinical setting. The primary reason for failure in discovery is due to poor penetration of compounds into bacterial cells. For example, in gram-negative bacteria, the presence of the double membrane along with a variety of efflux pumps that expel drugs out of the cell, making it difficult to design new antibiotics target [50–52].

Economic and regulatory reasons: In addition to the previously discussed factors, several economic factors also play a major role in the burden of the new antibiotic and which also affects directly or indirectly physicians, patients, healthcare administrators, pharmaceutical producers. One major reason is the increased costs, and the amount of time needed to put a drug on the market. Due to the increased production costs, pharmaceutical companies cannot cover the risk of research and development because for fear of resistance growing very fast and antibiotics having a short life span. For Food and Drug Administration Safety regulation, some of the drugs are impossible to approve due to the demand for lower antibiotic development costs. For example, in between 1980 and 2009, there are ~60 new antibiotics approved, but unfortunately pharmaceutical company incentive to sell due to production cost [50–52].

2.5. On the hunt for new antibiotics

As mentioned in the earlier section, in recent decades due to the emergence of the deadly drug resistant superbug we are nearing a crisis point of the effective antibiotics. In general, most antibacterial drugs are target at intracellular processes and needed to penetrate the membrane barrier. Due to double membrane along with the drug efflux pump always making difficulties to enter those drug molecules. Alternative approaches: i) by changing the drug target site (e.g. *fim*H inhibitors, quorum sensing inhibitoryn molecules), ii) the accurate measure the drug kinetics and penetration, iii) find out some alternate way to overcome the membrane barrier (e.g. Porin-mediate permeability) and iv) finally, the use of natural product could also be alternate new generation drug therapy. Recent studies pinpointed the most promising source of

antibiotics from the natural resource. For example, in a recent study Chung et al. have highlighted the development of a synthetic peptide called DRGN-1, which very similar to antimicrobial active histone H1-derived peptide (VK25) from the Komodo dragon DRGN-1 exhibits more antimicrobial and anti-biofilm activity compare to VK25 [53]. Recently, scientists have pinpointed a promising source of the antibiotic from leaf-cutter ants. They reported how ant uses specific bacteria, which secretes antimicrobial chemicals against certain fungi and other microbes [54]. In another study Maffioli et al. have found a new kind of compound named pseudouridimycin (PUM) from buried in dirt, which effective against deadly drug-resistant bacteria infection [55]. Interestingly, in another study Zipperer et al. have identified lugdunin (thiazolidine-containing cyclic peptide) a new type of antibiotic from human nose, which can be capable of fighting against many superbugs i.e. *S. aureus* [56].

2.6. Prevalence and spread of ESBLs –producing Enterobacteriaceae

The worldwide emergence of gram-negative antibiotic resistant pathogens are one of the most serious problems in the realm of bacterial infections. In the last decade, the prevalence of Enterobacteriaceae that produce ESBLs or carbapenemase enzymes continues to increase at alarming rates. In 1983 first plasmid-mediated ESBL was reported in a *K pneumoniae* isolate in Germany, and subsequently identified in Argentina, France and Italy [57,58]. ESBLs are enzymes produced by a variety of gram-negative bacteria that can hydrolyse the amide bond in the β-lactam which confer an increased resistance to commonly used antibiotics such as penicillins, cephalosporins, cephamycins, and carbapenems [59,60]. Based on functional and substrate profile in general, beta-lactamases can be divided into three major groups; TEM, SHV, and CTX-M and some minor groups such as OXA, PER, GES, VEB. Each major group can be further divided into subtypes (e.g. TEM: TEM-1, TEM-2 and TEM-3, CTX- M: CTX-M-1, CTX-M-2, CTX-M-8, CTX-M-9, CTX-M-15, CTX-M-27 and so on) [61]. ESBLs are predominantly observed in *K pneumonia* and *E. coli*, but may also be found in other species of Enterobacteriaceae. These bacteria are common causes of urinary tract infections (UTIs) and cause sepsis, respiratory tract- and intra-abdominal infections [62]. During the 1990s, the prevalence of ESBLs was higher TEM or SHV types and a very low frequency of others ESBLs such as CTX-M, VEB, and GES enzymes. Since the twentieth century, the scenario has changed, the incidence of

CTX-M upsurge dramatically with the highest number of observed variants. CTX-M β-lactamase is considered as a major contributor in the paradigm of spreading resistance mechanism in gram-negative Enterobacteriaceae spp. [63]. Previous studies suggested over the last 10 years only CTX-M ESBL enzymes are emerging and other ESBL enzymes have nearly disappeared in Enterobacteriaceae. Genes conferring CTX-M-type β-lactamase resistance are commonly plasmids driven and can be transferred between unrelated species by horizontal gene transfer. In Enterobacteriaceae, IncFI/FII, IncI, IncL/M, IncA/C, and IncK plasmids are commonly associated with genes for CTX-M enzymes [64]. Concerning the geographic distribution particular types of CTX-M enzymes are also associated with geographical regions. For example, CTX-M-14 type ESBL are frequent in Asian countries (dominant in China, Taiwan, and South Korea) whereas CTX-M-15/1 are dominant in Europe and USA. Overall CTX-M-15 and CTX-M-1 are widely distributed and disseminated worldwide [65]. Retrospective cohort studies showed third-generation cephalosporins (3GCREB) resistant Enterobacteriaceae has emerged, mostly due to the occurrence of ESBLs, which is an ongoing problem in Europe as well as worldwide [66]. In Europe, the prevalence of ESBLs significantly increase in mainly *E. coli* and *Klebsiella spp*, since the first report in 1983. This displacement might have occurred not only as a consequence of the extraordinary dissemination of the corresponding $bla_{\text{CTX-M}}$, over time. The prevalence of ESBL- carriage has changed significantly in different parts of the world (Figure 7). ESBL producers are not only resistant to beta-lactam antibiotics, but also, by co-resistances, are resistant to different other antibiotic and antimicrobial agents, such as aminoglycosides, trimethoprim, sulfamethoxazole and several heavy metals and detergent [67].

Clearly, there is an immediate need for improving infection control and rapid detection of ESBL, identify the source and find a solution of treatment. If not stopped, ESBL encoding Enterobacteriaceae will emerge as one of the major causes of death in the coming decades (Figure 2.7).

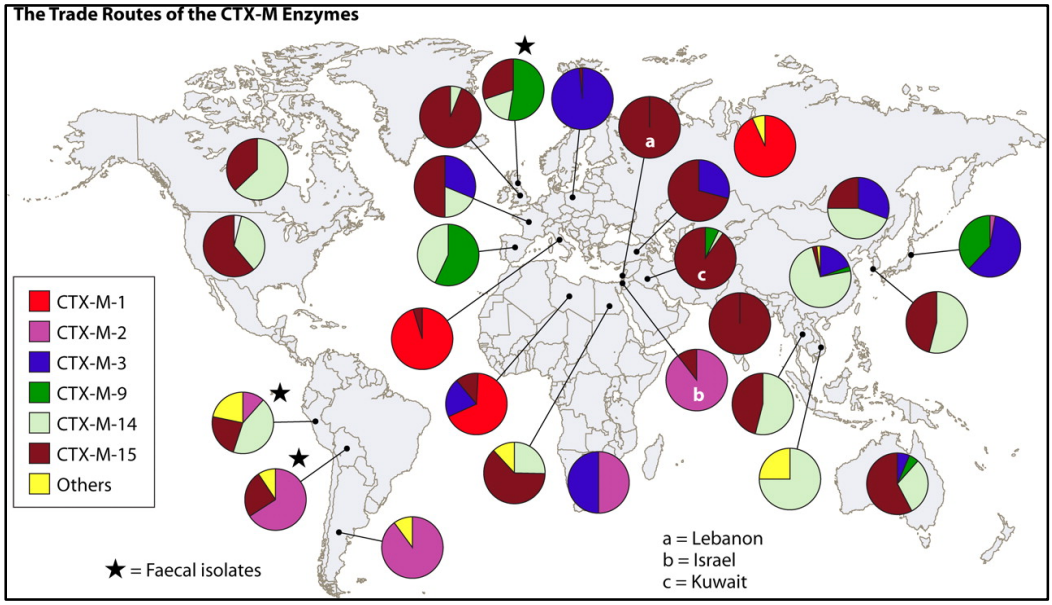


Figure 2.7: Worldwide prevalence of different classes of CTX-M β-lactamases (Source Davies, 2010).

2.7. Microbial identification and strain typing for the clinical setting

The development of quick and reliable microbial typing methods has transformed the surveillance and infection control. Bacterial genotyping methods can provide a comprehensive assessment of the unique label of intra and inter species microbial community as well as other features e.g. virulence factor,resistance geneand epidemiological information [68]. Typing systems for discriminating between different bacterial isolates are either phenotypic or genotypic level.

Phenotypic typing: It is determined by the morphology of colonies or by expressing gene products to distinguish among different microorganisms. Phenotypic methods based on several approaches, such as, serotype (determinants antigenic expressed on the cell surface), phage-typing (identify the viruses that infect and destroy bacterial cells), bio-type (classified together based on specific variation in biologic behavior), antibiogram (based on susceptibility of bacterial isolates to a panel of antimicrobial agents), Multilocus enzyme electrophoresis (MLEE) (characterizes the cellular proteins by electrophoretically separating them in a gel matrix), bacteriocin typing (identify the toxins produced by bacteria) and recently MALDI–TOF (mass spectroscopy of whole cell composition) also become a part of bacterial typing [69]. Phenotype-based microbial typing methods are successfully used in food, water, clinical, and pharmaceutical microbiological testing laboratories. However, the major drawback of such typing methods is that they are time-consuming, less discriminatory and the complexity of interpretation and phenotypic characteristics can vary in different

conditions. In recent years and with the advent of several new methodologies over the last two decades, DNA based typing methods have progressively replaced phenotypic assays.

Genotypic typing: Genome-based typing methods are indispensable to study the epidemiology of microbial community, which not only provides bacterial species identification but also subtypes. DNA-based microbial identification methods are theoretically more reliable because nucleic acid sequences are highly conserved in most microbial species. Applicable genotypic based typing methods include DNA–DNA hybridization, PFGE, PCR, 16S, and 23S rRNA sequencing, multilocus sequence typing (MLST), Multilocus variable number of tandem repeats analysis (MLVA) and whole genome sequencing-based typing. Some of the over used genotype based typing methods outlined below.

Pulse field gel electrophoresis (PFGE): PFGE is molecular-based laboratory subtyping strategy that is used for epidemiological studies in worldwide. Here the genomic DNA is fragmented by using restriction enzymes. The fragments are then run on a gel with an alternate electric field change and separated based on their sizes. Compare to the conventional agarose gel. PFGE differs from static to an alternate electric filed and capable of separating larger (10 to 800 kb) fragment. Once the DNA fragments and fingerprint produce on the gel, the fingerprint data could be, analyzed using different software packages e.g. BioNumerics (http://www.applied-maths.com/bionumerics). After analyzing the local public health laboratory pattern, it is also possible to compare the local fingerprint pattern with national wide or global database (e.g. PulseNet central database) to find strains similarity. Probably the PFGE typing method still consider as the ‘gold standard’ of choice in the typing of human bacterial pathogens and the investigation of disease an outbreak. PFGE is relatively costly, labor-intensive, lacks reproducibility, and time-consuming to obtain a result. The degree of discrimination also depends on the choice of restriction enzymes [70] (Figure 2.8).

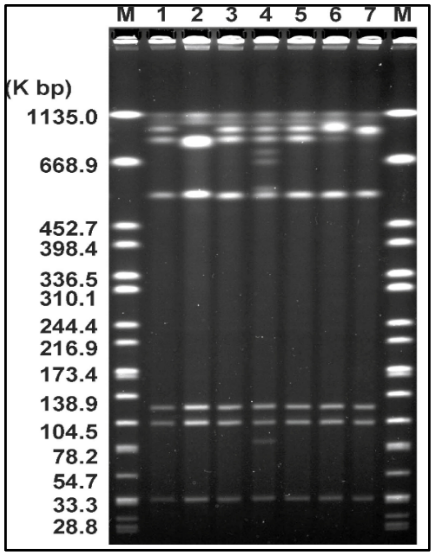


Figure 2.8: Representative PFGE of seven clinical *E. coli* isolates (Source Hirai, 2013).

Multilocus sequence typing (MLST): The Multilocus-sequencing typing (MLST) provides a high-resolution typing approach to uncover allelic variants based on i seven housekeeping genes (450 to 500 bp). The MLST approach allows detecting variations in the different loci within a species and permits the identification of the identical clone. MLST typing exhibit lack enough resolution to investigate the epidemiological concordance such as to distinguish outbreak and non-outbreak isolates. There are several available MLST typing schemes hosted on different publicly available websites. For example, for *E. coli* there are three different MLST schemes available. Namely, Mark Achtman's seven genes scheme, the Pasteur Institute eight genes scheme, and Shiga toxin producing *E. coli*, 15-genes scheme [71]. There are several software packages, or online resources available to analyze MLST data from sequencing (E.g. https://github.com/tseemann/mlst, https://cge.cbs.dtu.dk/services/MLST/). Although, MLST is currently one of the most popular genotyping methods, but it is a time and cost-intensive method. However this method is widely used to generate evidence in population genetics and reconstruct microevolution of epidemic bacteria by selecting appropriate housekeeping genes [72] (Figure 2.9).

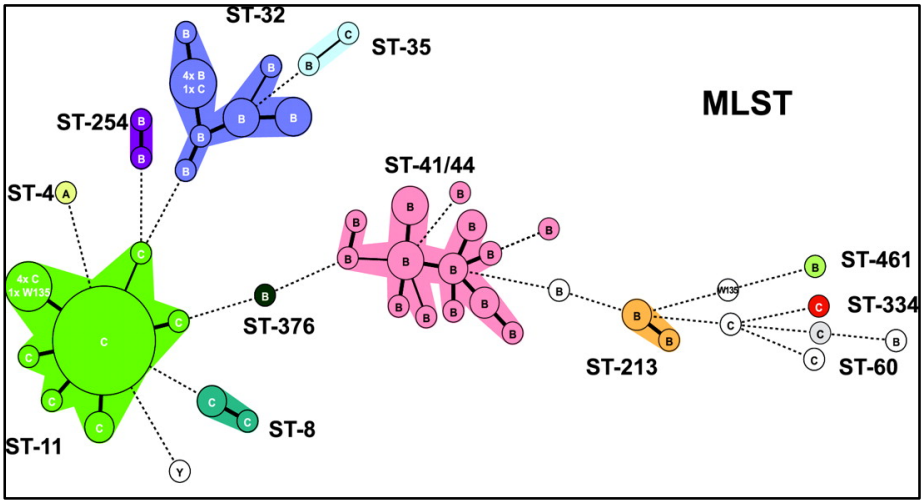


Figure 2.9: Minimum spanning tree analysis of 85 *N. eningitidis*. Here each circle corresponds to indicate the number of isolates of a particular type. The sequence type has given beside the circle. Each Thick, solid lines denote single-locus variants; whereas the dotted lines indicate types that differ in more than two loci. (Source Schouls ML, 2006).

Multilocus variable number of tandem repeats analysis (MLVA): The MLVA is a technique related to MLST based on PCR amplification, and is a fast andportable method that analyzes multiple VNTR loci of rapidly mutating repetitive DNA sequences. Although MLVA is faster, easier and more inexpensive to perform than MLST, there are issues with reproducibility and validation. Therefore, MLVA is not suitable for long-term epidemiological surveillance [73].

Ribotyping: This method involves fingerprinting of genomic DNA and relies on the relative stability of the 16S, 23S, and 5S rRNA genes coding for ribosomal RNA operons. Conceptually, ribotyping is similar to probing restriction fragments of chromosomal DNA with cloned probes. The genes are fragmented using restriction enzymes, and resulting DNA fragments separated by electrophoresis show polymorphic regions with varied size and sequence. Several ribotyping-based studies have successfully shown taxonomic classification of complex microbial communities in water, soil, and other environments [74].

Repetitive sequence-based PCR (rep-PCR): Microbial genomes contain numerous non-coding, repetitive DNA sequences (33 and 40 bp) that occupy intergenic regions and their arrangement varies between strains (500 to 1,000 copies per genome). The rep-PCR technique relies on amplifying and separation these repetitive sequences to produce unique DNA profiles or fingerprints for individual bacterial strains. The rep-

PCR is a useful tool for strain tracking in routine pathogen surveillance and disease outbreak investigations. There are three families of repetitive sequences exist, the 35-40 bp repetitive extragenic palindromic (REP) sequence, the 124–127 bp enterobacterial repetitive intergenic consensus (ERIC) sequence and the 154 bp BOX element which frequently use in REP-PCR assays [75].

DNA Microarrays: This method offers a rapid, specific and cost-efficient approach and has been proven particularly effective for detecting the presence of a gene of interest. DNA microarray is a collection of thousands of spotted DNA probes immobilized in an ordered pattern onto a solid surface (a microscope glass slides or silicon chips or nylon membrane). The probes can be PCR products of open reading frames (ORFs), amplified from the available sequenced genome. The target molecule is initially fluorescently labeled later hybridized by base pair matching to its cognate recognition probe, which is presented on the plate. Since the probes on the microarray are organism-specific, the detection signals generated upon hybridization provide the basis for pathogen identification. DNA microarray not only used for genome-wide comparison of their genetic contents, but also have been used to measure changes in expression levels and to detect mutations [76].

Whole genome sequencing (WGS): The whole genome sequencing (WGS) method provides a superior alternative when compared to the genotyping methods discussed earlier and introduced a significant breakthrough in microbial epidemiological studies. . Currently, with the advancement of high-throughput benchtop NGS instruments like 454, Illumina, PacBio and Oxford Nanopore, WGS whole genome could be achieved in less than 24 hours. At present, data interpretations from all WGS platforms are still computationally challenging. During the last decades, the rapid technological improvement in sequencing technology and subsequent bioinformatics analysis has transformed our understanding of biological knowledge [77,78]. Looking back over the past decade the development of WGS technology could be marked by three major phases as follows

Table 1: Comparison of most common bacterial typing techniques (adapted from Adzitey et al. 2013, Li et al. 2013, Sabat et al. 2013).

Typing Methods	Methods description	No. of marker	Advantages and Limitations
MLST	PCR amplification of housekeeping genes to create an allelic profile	7	Highly reproducible, can be used to type almost all strains of bacteria, moderate discriminatory, uses only seven loci, not suitable for routine infection control or outbreak investigation
RAPD	PCR amplification with arbitrary primers, separation of the amplicons by gel electrophoresis	NA	Highly laboratory dependent, low reproducibility, PCR bias, needs high stringent PCR conditions
PFGE	Phenotypic characterization of the electrophoretic mobility of housekeeping enzyme	NA	Widely used molecular typing method, highly reproducible, time consuming, requires skilled technician, does not discriminate between unrelated isolates
RFLP	Digestion of genomic DNA with restriction enzymes to produce multiple short restriction fragments	NA	Cheap, rapid, readily available and easy to perform, Has average reproducibility, discriminatory power moderate
AFLP	Digestion of genomic DNA with two restriction enzymes, ligation of restriction fragments	NA	High reproducibility requires less amount of DNA, no knowledge of DNA sequence is required prior to analysis, produce highly informative fingerprints, inability to differentiate different fragments with similar size (homology) and scoring bias
MLVA	PCR amplification of VNTR loci followed by sizing of the PCR products to create an allelic profile	10-80	Widely used to assess the molecular fingerprint of micro-organisms, useful to identify composition of bacterial populations, requires a trained and skilled technician, is not a practical routine subtyping method
16S-23S ITS PCR	PCR amplification of rps genes to create an allelic profile	53	High specificity, non-specific PCR amplification due to primer bias
Plasmid analysis	Determine the diversity and plasmid size by using PFGE or by sequencing	NA	Easy to perform and to interpret the results, plasmids can readily be lost or acquired and can make genetically related isolates to have different plasmid profiles
Microarray	High-throughput and versatile technology, detecting polymorphisms and mutations	200-300	Simultaneous detection of presence of a large number of genomic loci, limited number of loci detection, complexity and cost is not suitable for regular clinical purposes
Whole genome	Detection of similarities/differences in the pan-genomic or distributed genes	>1000	single-nucleotide resolution, high-throughput, high reproducibility, high sensitivity, cheaper cost, faster performance, interactive bioinformatics tool, statistical analysis, high flexibility, highly informative

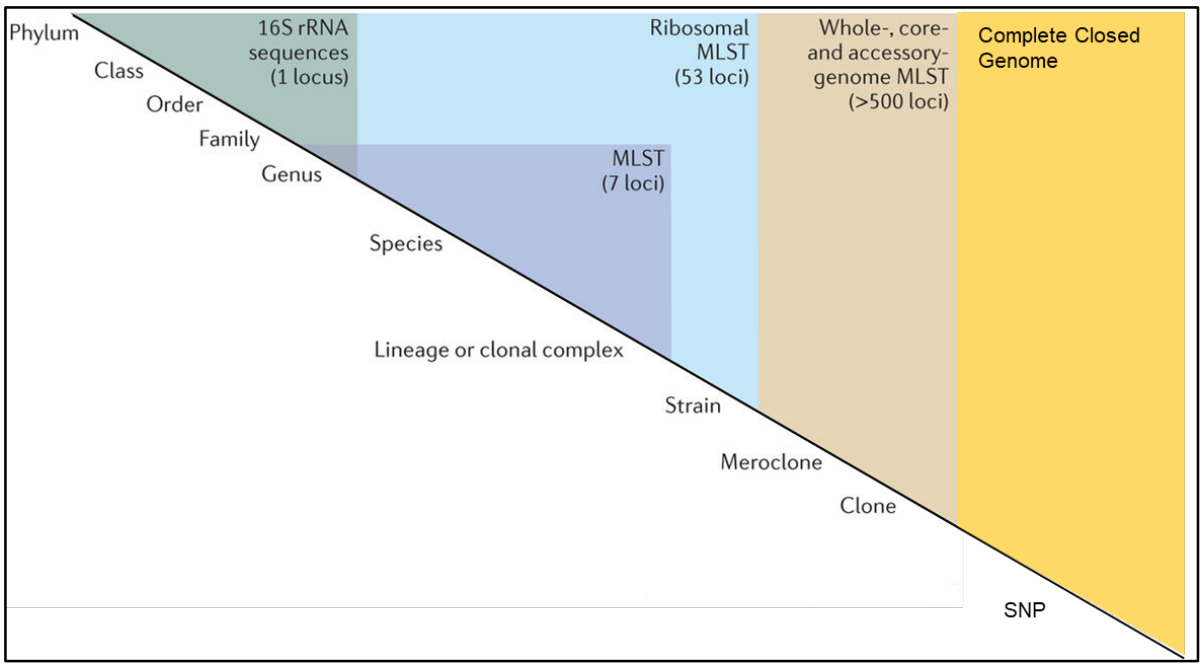


Figure 2.10: Schematic representation of different typing methods and their resolution levels within a typical population structure (Source Maiden, 2013 & modified) Different levels of loci information can be associated with hierarchical nomenclature schemes.

2.8. Genome sequencing

2.8.1. First revolution in sequencing

In the the 1990s, the bacterial genome-sequencing revolution was initiated by the introduction of automated DNA sequencing technologies. However, the development started in the late 1970s, by Sanger and later Maxam and Gilbert in 1977. Both these platforms were based on sequencing by synthesis (SBS). Sanger's sequencing methods were popular comparing to Maxam and Gilbert method DNA sequencing that is also called as the 'First- Generation Sequencing' (FGS). With Sanger's technique, Sanger and colleagues successfully deciphered the PhiX174 bacteriophage genome. The basic mechanism used deoxynucleotide (dNTPs) and fluorescently labeled modified dideoxynucleotides (ddNTPs) to act as chain terminators as dideoxynucleotides lack of 3´OH and stop chain elongation further separated in a parallel fashion on acrylamide gels in capillary electrophoresis to obtain the final output [79,80].

2.8.2. Second revolution in sequencing

There was another continual modification of sequencing concurrent with the development of Sanger sequencing, although the basic principles have remained

same with an addition of massively parallel sequencing of millions of DNA fragments simultaneously. In the second revolution, major game changers were introduced by 454 Life Sciences, Life Technologies and Illumina sequencing technology. In 2005 the first pyrosequencing "sequencing-by-synthesis" platform became commercially available by 454 Life Sciences, which was later acquired by Roche [81,82]. The technology behind pyrosequencing chemistry was based on emulsion PCR [83] (emPCR) of DNA coated beads and light is detected and measured after it is produced by luciferase sulfurylase reaction. In the first step, double-stranded DNA (dsDNA) is fragmented by nebulization to shorter lengths ranging between 100 bp–3 kb. The blunt-ended fragment followed by phosphorylated and ligated to adaptor into dsDNA. Next, dsDNA fragments are then separated into single-stranded DNA and the fragmented DNAs mixed with the micro-sized agarose beads (~28μm diameter). Each bead surface carried oligonucleotide sequence, which is complementary to an adaptor sequence of fragmented DNA and one DNA fragment bind with one bead. After mixing, the library fragment is captured by the specific oligonucleotides on the beads and the complex mixed with oil-water emulsion. Next, the fragmented DNA on the bead is amplified by an emulsion PCR, which produces around one million copies of each fragmented DNA. Each bead is then placed into a picotiter plate, which allows placement of only one bead and is mixed with DNA polymerase and sequencing buffer. The next step is pyrosequencing where each plate acts as a flow cell and each time depended on the incorporation of a new nucleotide released light by the activity of pyrophosphates, ATP sulfurylase, and luciferase, which determined incorporation events of each nucleotide. At each time, the light signal emitted by the amplified reaction is captured by using by CCD (charge-coupled device) camera, which further proceeds to generate a flowgram to get a final sequence. The 454 platforms have been used in a great variety of applications such as clinical microbiology and functional genomics, and its long reads have made it especially appealing for studies of microbiomes since only longer reads can generally be identified with greater accuracy and precision of the complex region of a genome. Few major drawbacks of 454 detection relates to the homopolymers present in the DNA to be sequenced i.e. consecutive stretch of a similar bases) which often can result in an error, and cause an insertion-deletion (indel). A major factor relates tothe cost of the instrument (~$500,000) and cost per megabases (~$60) [84]. Another game changer in this era was Illumina (Figure 2.11) sequencing technology, which began in mid-1998 with an

Indian-born British chemist Prof. Shankar Balasubramanian and Prof. Herchel Smith from Medicinal Chemistry in the Department of Chemistry the University of Cambridge. The major idea behind was “Sequencing by synthesis" (SBS), which developed in Solexa in 1998. In 2007, Solexa was acquired by Illumina, and large-scale microbe, human and other animals sequencing has been achieved by this approach. The Illumina sequencing workflows include three basic steps:

1. Library Preparation: In the first step of the Illumina protocol, an adenylated DNA adapter is ligated with 5′ and 3′ randomly fragmented DNA. Alternatively, “tagmentation” combines the fragmentation and ligation reactions into a single step that greatly increases the efficiency of the library preparation process.
2. Cluster Generation: For cluster generation, a parallel PCR bridge amplification takes place to generate the dense cluster where each of the DNA fragment is binds and forms a bridge.
3. Sequencing by synthesis: After cluster generation of sequencing, four fluorescent-labeled nucleotides are sequenced of each cluster on the flow cell surface in a parallel fashion. In each cycle, a single labeled deoxynucleotide triphosphate (dNTP) is added to the nucleic acid chain. The labeled nucleotide serves as a terminator for polymerization, after each dNTP incorporation, and the fluorescent dye is imaged to identify the base and then enzymatically cleaved to allow incorporation of the next nucleotide [85].

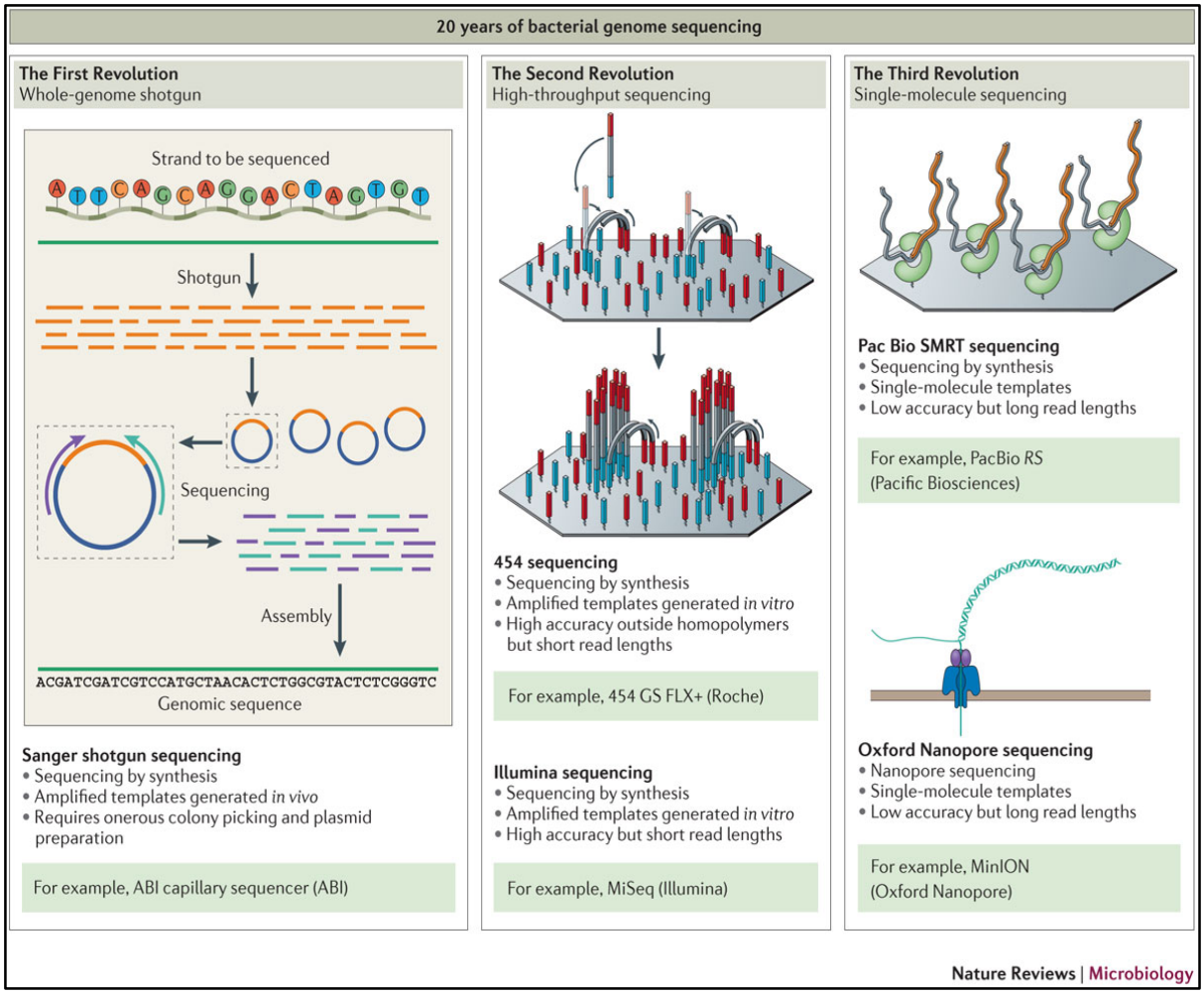


Figure 2.11: The three revolutions in sequencing technology that have transformed the landscape of clinical microbiology (Source Loman, 2015).

2.8.3 Third revolution in sequencing (3rd GS)

Another revolution in the sequencing technology was achieved by introduction of single molecule sequencing by Pacific Biosciences (Pacbio) and Oxford Nanopore. The first single-molecule real-time (SMRT) sequencing was developed from Pacific Biosciences. Unlike first and second-generation sequencing technologies, PacBio do not require amplification of DNA.

In principle, the SMRT technology follows two main steps. Firstly, SMRT bell generation: where fragmented dsDNA is ligated to the hairpin adaptor to both the ends called a SMRTbell (Figure 2.12). The fragment size varies from 250 to ~10kb. In next step the library is then load onto a SMRT cell that contains an array of several thousand (~75,000) of individual Pico liter wells termed as zero-mode waveguides (ZMWs). ZMWs are nanoscale metal apertures within which no propagating modes of light can

exist and it acts as a detector during sequencing, when DNA molecule passed inside the ZMWs via polymerase each time fluorescent bases added and light emitted. Each time the fluorescent recorded by an array of the movie of light pulses [86,87] and fluorescence color determined which nucleotide has incorporated (Figure 2.13). An important advantage of PacBio sequencing, compare to earlier technology is the longer read length with less runtime. These long reads are extremely valuable to resolve the complex repetitive regions of the genome. PacBio also can detect DNA methylation, which is very helpful to study biological regulation, such as gene expression, gene silencing, and host–pathogen interactions. In October 2015, PacBio has launched another new sequencing platform the Sequel system, with ~1 million ZMWs per SMRT cell, and up to seven times more reads with half the instrument cost.

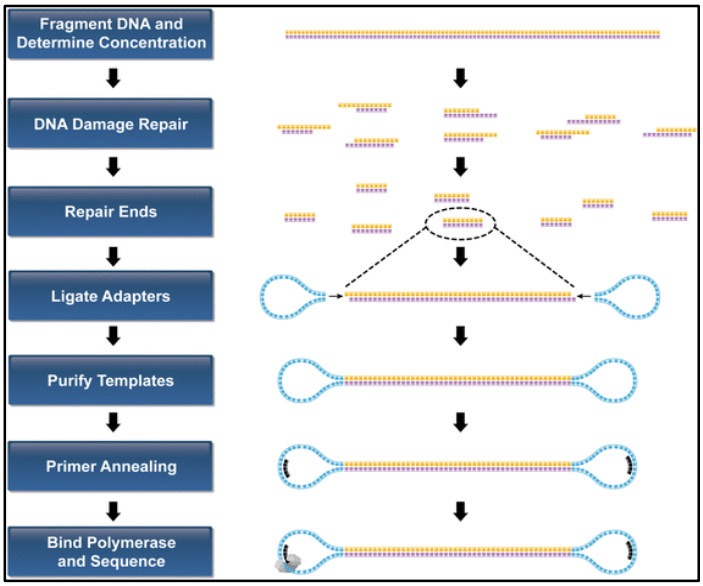


Figure 2.12: Diagrammatic representation of PacBio SMRTbell preparation workflow Firstly the input sample dsDNA is fragmented into desire size. The ends are then repaired and ligated to hairpin adaptors (blue), which forms a closed circle. Later purified templates are submitted into a sequencer (Source: Kong N, 2017).

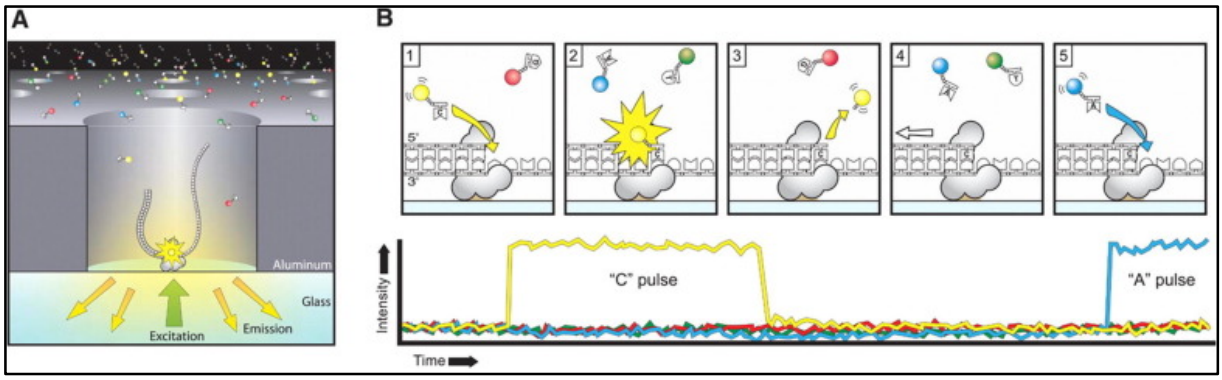


Figure 2.13. Diagrammatic representation of SMRT sequencing technology (Source: Rhoads A, 2015).

Another cutting-edge real-time sequencing technique was introduced by Oxford Nanopore technology (ONT). ONT also referred as fourth generation sequencing, as sequencing performed here without intervening PCR amplification. It has the potential to quickly and reliably sequence within a minimum effort and short time (e.g., entire human genome for less than $1000). The principle behind real-time nanopore sequencing is to apply an electrical potential across a membrane, and the changes of electric current are monitored to determine whether any analyte has passed through the protein-inserted (α-hemolysin) pore. When analytes such ssDNA molecule is passed through the pore, each base is disrupting the current to a different extent which allowing nucleotides to be identified by the detector [88,89]. This technology not only allows us to study the DNA molecule but RNA and protein molecules also could be identified. The error rate of ONT is still high, base calling algorithms needed in order to reduce error rates.

2.9. From reads to the complete finished genome

The new generation of sequencing technologies is revolutionizing in the field of molecular biology and genetics paradigm. Currently, with the advancement of sequencing technology, genome sequencing becomes a routine and affordable health care procedure. Some new sequencing technologies have spurred by rapidly decreasing costs of genome sequencing. Despite this advancement in sequencing the major bottleneck bioinformatics analysis, particularly in genome assembly, error correction and gaps present in final assembly are always lagging behind.

2.9.1. Preassembly steps

Before running the genome assembly, quality assessment of raw reads are needed because good quality data leads to an accurate assembly. In raw read cleaning steps, the base quality, GC content, presence of adaptor sequences, abundance repeat and duplicate sequences should be assessed. Also in Illumina sequencing technology PhiX phage DNA is often added to the sequencing reaction, which should be removed before assembly. There are several command line or web based tools that are publicly available for conducting a quality assessment of raw FASTQ files. Tools such as FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) or multqc [90] provide quality statistics of raw reads. Quality Trimming of low-quality bases (<Q30)

can be performed by using tools such as Trimmomatic [91] or cutadapt (http://cutadapt.readthedocs.io)

2.9.2. De Novo assembly approaches

The strategies adopted by genome assembly can be widely divided into two categories: de-novo assembly (if no prior knowledge is available) and reference based assembly (complete reference genome to guide) [92,93]. The basic strategy of De novo assembly is the process of stitch together the short NGS overlapping reads into contiguous sequences (contigs) without the use of any reference genome as a guide. The most efficient assemblers for short-read sequence assembler are typically followed one of several strategies such as greedy, overlap layout consensus (OLC) [94], de Bruijn and string graphs approach [95]. The choice of the paradigm depends on the characteristics of the sequencing data being assembled. For example, OLC or string graph is mostly used for longer, more inaccurate sequencing data. The approaches follows three major phases: initially all overlaps reads are first identified, secondly the overlaps, information is then constructed as a graph and finally the consensus sequence is inferred from the graph (Hamiltonian path). First generation genome assembler such as Newbler, TIGR, Arachne, Celera Assembler (CABOG), Minimus, Edena, CAP, PCAP followed the OLC approach [96]. The major dis-advantages of OLC approach are to handle repeated parts of a genome, correct order of the contigs and thereby being a time-consuming approach. Introduction of the most widely used de Bruijn graph led to significant improvement in the genome assembly. Here assembly based on k-mer graphs (Figure 2.14) extracted from the input reads. The first step here is starting with a k-mer size from read information and split the original raw reads. Next, a directed graph use to connecting pairs of k-mers with overlaps between the first k-1 nucleotides and the last k-1 nucleotides. The direction of the arrow goes from the k-mer, whose last k-1 (Eulerian path) nucleotides are overlapping, to the k-mer, whose first k-1 nucleotides are overlapping [97]. Tools such as Velvet, ALLPATH, ABySS, SOAPdenovo, SPades. Compare to earlier strategies major advantage here no calculation of pairwise alignments needed, which speed up the assembly. The disadvantage here k-mer graphs are sensitive to choose the parameter k Even though there are more than fifty assemblers available, no particular package stands out as the best solution for de novo genome assembly. Performance of an assembler would largely depend on sophisticated properties of genome content, and NGS data, such as error rate, read depth, final error correction [98,99] (Figure 2.15).

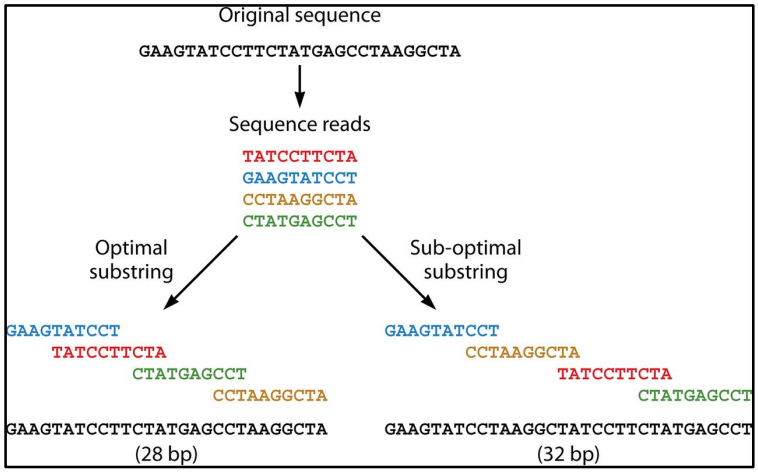


Figure 2.14 : Simplified genome assembly steps.

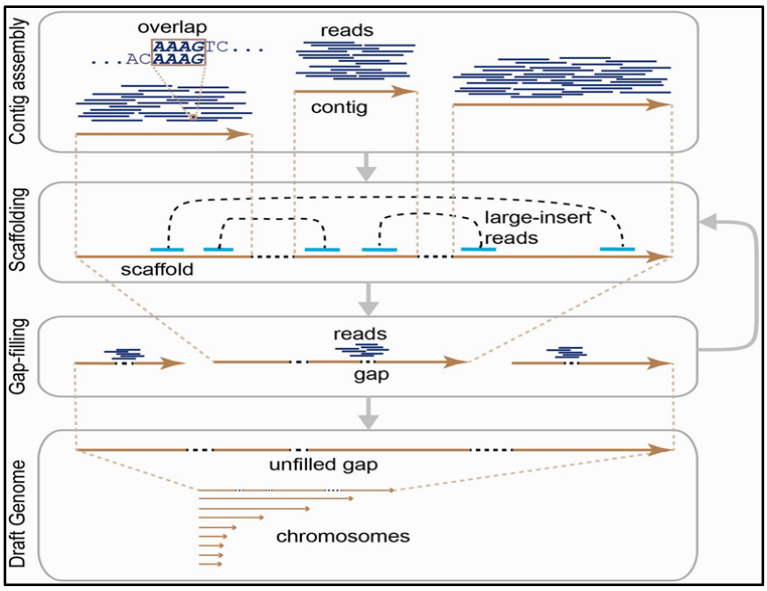


Figure 2.15: General workflow of the de novo assembly of a whole genome. Initially, by overlapping reads, contigs are assembled from short reads before scaffolding, and the remaining gaps are filled. Through this procedure, a draft genome consisting of chromosomes are built. Some unfilled gaps may remain in the draft genome (Source Sohn, 2016).

2.9.3 The coming era of long-reads sequencing and hybrid assembly

The development of third generation sequencing (TGS) technologies like single molecule Real-Time (SMRT™) sequencing and Oxford Nanopore Technologies (ONT) methods has been a major driving force behind the rapid advancements in genomics in last five years. Notably, the advent of second generation sequencing (SGS) is capable of generating hundreds of thousands or millions of short DNA sequence reads at a relatively low cost. Recent years, TGS is capable of generating long sequence reads allows without amplification-free approaches in a shorter time and at even lower

costs per instrument run. Long reads sequencing could also be used as the intermediate bridge for short reads assembly, which is very useful for investigating structural and copy number variation within de novo assemblies of the sampled populations. Combined hybrid analysis of long-read and short-read sequencing technologies not only offer to improve the assembly quality, N50 values, but also fill the gaps generated by conventional NGS platforms [100] (Figure 2.16). Long reads sequencing platforms has also followed a different algorithm for de novo assembly and analysis protocol, such as PacBio, have specific de novo assembly algorithm, e.g. Hierarchical Genome Assembly Process (HGAP) [101], FALCON toolkit (https://github.com/PacificBiosciences/FALCON), Canu [102].Nanopore sequencer also use different assembler e.g. GMAP, Graphmap [103] and MHAP [104].

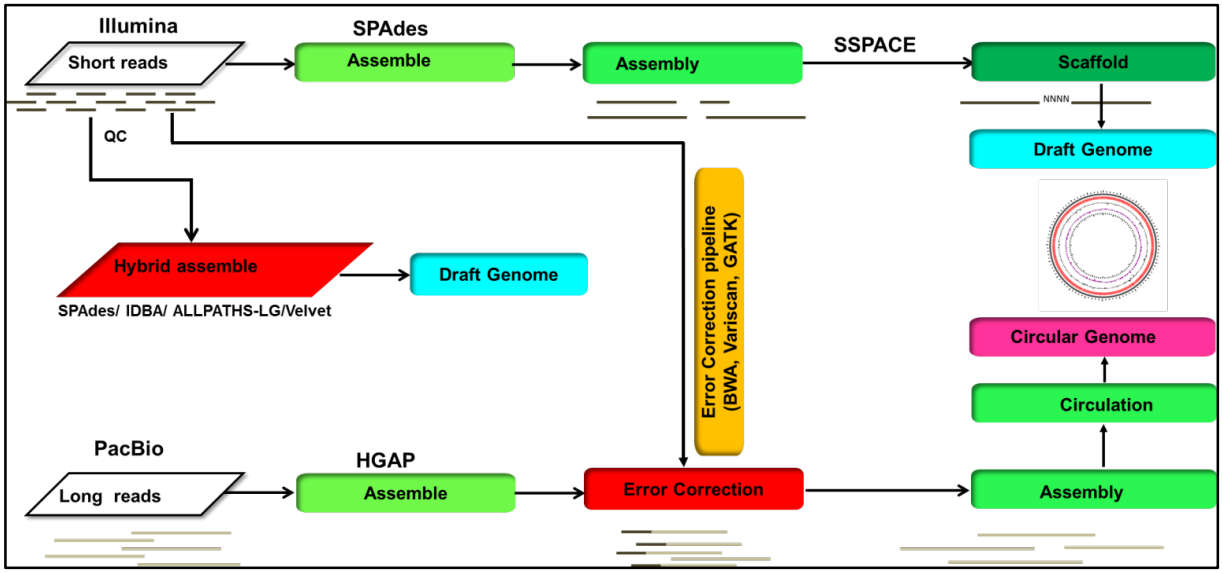


Figure 2.16: In house workflow for producing hybrid assemblies by using short reads and long reads sequencing approaches.

2.9.4. Post-assembly data analysis

Once contigs are assembled from the sequencing reads, the next step is to evaluate the quality of genome assembly. In order to obtained a gap-free, high quality assembled genome we need pass through draft assembly into a variety of post assembly protocols.

Assembly Quality assessment: It is important to evaluate the performance of the quality of the genome assembly. To evaluate, the user needs a statistical approach for quantifying the error rates of the assemblies. There are several standard metrics such as N50 (length of the contig that represent 50% of total assembly length) and counting

numbers are often used for assembly evaluation. Tools such as QUAST, KAT, Icarus, misFinder are widely used [105,106] to evaluate assembly quality for downstream analyses.

Assembly quality Improvement: Genome assembly from short-read sequencing technologies can have many internal errors such as genomic rearrangement and also fragmented into contigs. The incomplete and fragmented assembly leads to errors in gene identification in downstream genome annotation, which subsequent introduce biases in gene content within species, as well as gene gain and loss between species. Several comprehensive post-assembly genome quality improvement pipelines are available for polishing the assembly, e.g. eg. PAGIT (post-assembly genome-improvement toolkit) and Pilon. These tools improve draft genome assemblies by aligning sequences against contig ends, variant detection and performing local assemblies to produce gap-spanning contigs [107,108]. Once the genome assembly quality improvments are done, the draft genome are ready to run annotation steps**.**

Genome annotation: The massive genome sequencing data using NGS technologies are demanding automated systems capable of accurate annotation of a complete genome in a short time. In order to understand the functional content of a sequenced genome (e.g. ORF, tRNA, rRNA) and to do many downstream genome analyses high-quality annotations are needed.. The underlying mechanism behind most of the annotation software can be divided into two distinct stages. In the first stage, the step the expressed sequence tags (ESTs), proteins, Untranslated regions and so on, are identified and aligned to genome. Once a sequence has been defined, in the 'annotation' phase, these data is synthesized into gene name. Because this process is intrinsically complicated and involves so many different tools, the programs that assemble to compute data and use to create genome annotations are generally are referreded as annotation pipelines. Current pipelines are focused mainly on the annotation of protein-coding genes and rRNA element inside the cell. Tools such as Prokka, RAST, DIYA, RATT, BASys, and BEACON are widely used for bacterial genome annotation [109–113].

With the advancement of NGS bacterial genome sequencing, become easy and accessible in short time with limited cost. There has been a needed to develop new bioinformatics tools and areas, such as comparative genomics. The following section of the study will describe fundamental features of genome analysis.

Phylogeny: The phylogenetic or evolutionary tree represents the evolutionary relationships between genomes. In a molecular epidemiological study, phylogeny is routinely used to generate the outbreak of the transmission network. There are several algorithms available for the construction of the phylogenetic tree. e.g..: Distance-matrix methods: In this method, the distance between every pair of sequences is calculated by multiple sequence alignment and the resulting distance used for the construction of final tree. One most commonly used a distance matrix method is Neighbor-joining method, where the clustering algorithm measure the distance matrix to construct the phylogenetic tree. Another popular approach Maximum parsimony method where phylogenetic tree constructed based on the evolutionary events observed in the sequence data. Maximum likelihood: here the tree constructed based on probability distributions. This method uses several substitution models to calculate the probability of particular mutations. Finally, Bayesian inference: This approach uses different substitution model to construct the tree. For example, one of the most popular model is general time reversible model (GTR), which is frequently used. Commonly used tools for construction phylogeny are PHYLIP, PhyML, RAxML , and GARLI, etc. [114–116]. To visualize phylogenetic trees, common desktop based GUI tools such as FigTree (http://tree.bio.ed.ac.uk/software/figtree/), Dendroscope, MEGA, or Web-based applications such as iTOL (http://itol.embl.de/) (190), EvolView [117] and phandango [118] could be used.

Pan and core genome analysis: In 2005, Tettelin et al. introduced the concept of pan-genome. The pan-genome or supragenome has been defined as the entire genomic repertoire of a given species. According to the gene, content pangenome is divided into three groups: core (shared by all genomes), dispensable or accessory genome (which contains genes present between two and *n*–1 strains), and strain- (or isolate) specific unique genes (present only in a single strain). The relation of pan and core genome always depends on the size of the dataset. Few common tools for pan genome analysisare: Panseq (Pan-Genome Sequence Analysis Program), PGAT (Prokaryotic Genome Analysis Tool), PGAP (Pan-Genome Analysis Pipeline), BPGA, and Roary [119–122].

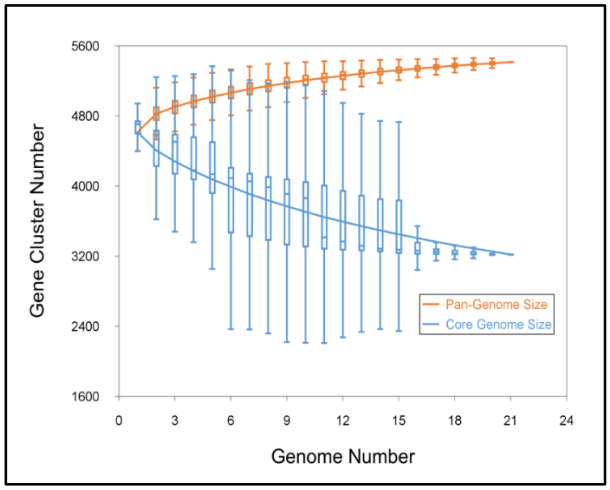


Figure 2.17: Schematic representation of Pan and core genome analysis.

Genome-wide association studies (GWAS): GWAS are the relatively new way of analyzing large-scale microbial genome sequencing data and becoming a popular day-to-day. In the last decade, GWAS have revolutionized a remarkable advent in human genetics to identify the genetic variation that influences the phenotype. The principle underlying GWAS is to identify certain SNPs at positions that are associated within a group of strains and influence the phenotype. by using several statistical model. In contrast to human genetics, bacterial genetics are relatively difficult due to the high plasticity of the genome caused by mobile genetic elements. Few existing approaches rely on genetic variation in the pan-genome with the phenotypes such as antimicrobial resistance or host adaption gene. Commonly use tools for GWAS include PLINK [123] and Scoary [124].

2.9.4 Genome characterization by using Web-Based tools:

Currently, high-throughput sequencing can generate high-quality sequence data within a short time period and at limited costs. In public health laboratories once whole-genome sequencing reads are generated, in next step theidentification of species, genome annotation, pathogen characterization, epidemiological profile, virulence and antimicrobial-resistance genes prediction can became a routine clinical practice. Several integrated and comprehensive web based tools are available for computing such information, which could be included in infection control settings without relatively deep bioinformatics knowledge, e.g.:

Specis level identification: Taxonomic classification from sequening reads is one of the most important tasks in clinical microbiology. Species level identification is very helpful in disease outbreak analysis. For species level identification, some useful, user-

friendly popular web based tools as discussed here. a) KmerFinder. KmerFinder is a method based on K-mer statistics across the entire genome for identifying bacterial species from raw sequencing reads or from contigs. The tool is accessible as a web-based platform (https://cge.cbs.dtu.dk/services/KmerFinder/) b) SpeciesFinder: SpeciesFinder predicts the prokaryotic species based on the 16S rRNA gene from NGS reads. A web service of SpecisFinder is available at (https://cge.cbs.dtu.dk/services/SpeciesFinder/). c) JSpeciesWS: JSpeciesWS is a computational pairwise comparison approach in between two species belonging same species or not based on their Average Nucleotide Identity (ANI) value. (http://jspecies.ribohost.com/jspeciesws/).

Virulence gene finding: Identification and characterization of virulence genes are a common practice in the clinical microbiology laboratory. There are several web-based tools available for detection virulence gene. a) Virulence factor database (VFDB): The VFDB is a web-based virulence factors database which provides in-depth coverage of virulence for bacterial pathogens [125] (http://www.mgc.ac.cn/VFs/), b) VirulenceFinder: The VirulenceFinder is a web-based approach to identify virulence genes. The tool uses BLASTn approaches to identify the virulence genes and can be accessed online (https://cge.cbs.dtu.dk/services/VirulenceFinder/)

Antimicrobial resistance gene finding: To identify the resistance genes, one of the most important tasks in all microbiology laboratories. Two most popular web-based antibiotic resistance gene prediction tools are a) ResFinder: identifies horizontally acquired resistance genes and finds the chromosomal mutations for acquiring resistance. ResFinder uses BLASTn for identification of acquired antimicrobial resistance genes. It is accessible via online (https://cge.cbs.dtu.dk/services/ResFinder/). b) RGI/CARD: is a manually curated database of antibiotic resistance gene. For detection of the antibiotic resistance gene, the CARD uses protein homology models and also predicts their products and associated phenotypes. (https://card.mcmaster.ca/).

Genome annotation: As it has been mentioned in the preceding section, genome annotation is a key process for identifying the protein-coding and non-coding regions of a genome gene location and functions. One of the most popular web based annotation: a) Rapid Annotation Using Subsystem Technology (RAST) is an automated web-based tool that can be used to annotate assembled contigs. The algorithm uses GLIMMER3 to identify gene–candidate, closest neighbor identification

by using BLASTP, the compare nearest neighbor by BLASTX finally overlap pattern remove by BLAST and SEED k-mer-based annotation algorithm [110]

Plasmid Characterization: As mentioned earlier section, plasmids act as a potential vehicle for harbor and spreading of resistance gene. The detection of plasmids Inc types and frequencies are important. Few popular web-servers for plasmid detection and pMLST prediction a) PlasmidFinder: PlasmidFinder [126] and pMLST, which provide *in silico* detection and characterization

Phage detection: Bacteriophage are one of the most dynamic part of bacterial genome sequencing data) PHAST (PHAge Search Tool) [127] is a web server to identify, annotate and graphically display prophage sequences from sequenced data within bacterial genomes or plasmids. PHASTER is the current upgraded version of the popular PHAST web server (http://phaster.ca/), which capable to identify, annotate and graphically display prophage sequence.

The web-based application evolved mostly in the last decade. The major advantages of web-based analysis that it is accessible everywhere without local installation that saves resources. Typically, the minimum requirement for a client would be a web browser on the PC. The major drawback of web-based tools are security risks a user could lose control of critical data in the unsecured server. Sometimes server failure on the host side or large, which could be slow down the analysis and undocumented changes made to the server misleading in the analysis.

2.10. Objectives of the study

Advancements and extensive applications of next-generation sequencing (NGS) technologies are providing a new and powerful way to obtain insights into genomics. Sequencing technology has been evolved at an impressive speed, with drastically reduced time, cost and higher throughput compared to low-resolution typing methods. The overall aim of this thesis was to investigate the molecular ecology and epidemiology of ESBL producing *E. coli* in Germany by adapting whole genome sequencing technology.

The specific objectives of the present dissertation were:

Objective I

To investigate the population structure and genomic relatedness of ESBL-producing *E. coli* isolates recovered from the human, animal and food across Germany, in between 2009 to 2016.

Objective II

To understand the abundance and diversity of the common (e.g. ST410) and host specific clone (e.g. ST131) of *ESBL*- producing *E. coli* across and detailed study of genome level relatedness among them.

Objective III

Adapting third generation single molecule sequencing, to study the genome plasticity of a globally disseminated clone (*E. coli* ST131) on a single base resolution level. Also to, evaluate the role of mobile genetic elements in adaption of *E. coli* ST131 into the human and animal.

Objective IV

Finally to study the potential role of translocatable units (Tus) in the dissemination of antibiotic resistance genes.

3. Population structure of multidrug resistant ESBL encoding *E. coli* in Germany

3.1 Publication

Manuscript under preparation (This study is the part of unpublished thesis work of HG).

3.2 Contribution

The author (H.G) was part of designing of the study and drafting the manuscript under the supervision of T.C. He performed pre-processed of raw reads and subsequent bioinformatics analysis and interpretation of the data.

3.3 Abstract

The global incidence of ESBL-producing *E. coli* has been an ever-growing burden in the hospital and community settings. Infections caused by ESBL-producing *E. coli* are associated with higher rates of mortality, morbidity as well as health care costs. Despite the growing importance of this pathogen, there is limited knowledge about the population structure and epidemiology of ESBL encoding *E. coli*. The potential reservoir and transmission of this bacterium remain inconclusive due to lack of large-scale population-based cohort studies. Here we performed whole genome sequence (WGS) of a subset of (n=4386) of ESBL-*E.coli.* Our aim was to investigate the population structure and genomic relatedness of a systematic collection of multidrug resistant *ESBL- E. coli* isolates (n=281) recovered from human, animal, food and the environment across Germany, between 2009 to 2016.

Our in-depth genomic and phylogenetic analysis elucidated a highly diverse population of ESBL *E. coli* are circulating in different hosts. In spite of large genetic diversity, a limited number of clones that emerged in the diverse background and highly associated with the spread of the antibiotic resistance gene. The predominant allele $bla_{\text{ctx-M-15/1/14}}$ was mostly associated with the conjugative IncF plasmids connoting transmission potential. Our genome-wide association (GWAS) study showed that the isolates from different ecological compartments share not only core genes but also there is always bidirectional gene flow in the accessory genomes, across the different hosts. Single nucleotide polymorphism (SNP) clustering approach shows that the

entire population could be grouped into nine major and five minor clusters. Overall, we identified the existence of most common and niche specific, highly adapted antibiotic resistant clone in a diverse ecological niche. For example, *E. coli* ST131 was identified in humans and companion animals whereas ST410 were distributed in different niche and circulating across humans, animal and the environment. Our study highlight that there are significant differences in strain-specific transmission across diverse hosts.

3.4 Background and Introduction

E. coli predominantly colonizes in the gastrointestinal tract of humans and many other warm-blooded animals. Pathogenic *E. coli* has a potential role to carry out a wide range of life threatening infections, including urinary tract infection (UTIs), bacteremia, meningitis in particular in the immunocompromised host [128][129]. According to the recent report, it has been estimated that on an average 150 million UTIs cases occur per year worldwide which costs the global economy in excess of 6 billion US dollars annually [130]. By the mechanism of horizontal gene transfer these bacterium acquired several mobile genetic elements (MGEs) such as genomic/pathogenicity islands, prophages, specific virulence factors that contribute to fitness advantage [131]. Recently a majority of these strains began to gain attention due to acquiring antibiotic resistance and form as a pathogenic organism to an extensive drug resistance (EDR) pathogen, which imposing serious therapeutic challenges. The worldwide emergence of ESBL encoding *E. coli* particularly CTX-M 15/1 producers among humans, companion animal, livestock to food became a major public health issue, offering a limited treatment option leading to higher morbidity and mortality [132]. ESBL encoding genes are mostly believed to be located on conjugative plasmids and less frequently on the chromosome which can be disseminate through horizontal gene transfer. ESBLs can efficiently hydrolyze the beta lactam ring of β-lactam antibiotics, such as third-generation cephalosporins and monobactams, as a result, these bacteria confer resistance to beta-lactam group of antibiotics such as penicillins, cephalosporins. Until now, a limited evidence has been speculated about the mechanism of spreading of ESBL genes into the different host via food chain or via direct contact of humans and animals [133]. In recent years, with the advent of high-throughput DNA sequencing technologies, it is possible to generate whole-genome sequence data from large numbers of clinically relevant bacterial isolates in a short time and it is possible to compare genomic differences in SNPs derived genotypes and virulence gene profiles

between different isolates and to draw a transmission and evolutionary pathways. In a previous study, our group phenotypically characterized the epidemiology of ESBL-*E. coli* from extra intestinal infections in humans and companion animals that were collected over a 5 yrs. time period in Germany [134]. In continuation of our previous work, this study was aimed to investigate the population structure, genetic composition of ESBL producer and genetic relatedness of ESBL-producing *E. coli* isolated from a shared population across Germany.

3.5 Material and methods

A total of 4386 non-repetitive ESBL-producing *E. coli* isolates were recovered throughout Germany in between 2009 to 2016 from various hosts (human, animal, food and environment). Initially, species level identification was performed using MALDI-TOF and ESBL production was confirmed using the double disk synergy test, and the presence of a particular ESBL gene was confirmed using allele-specific PCR. In the current study a representative subset of 281 isolates, human (n=83), livestock (n=71), companion animal (n=66), and food (n=61) were selected for sequencing and analyzed to assess the genetic relatedness among all those isolates.

Genomic DNA was isolated using a Purelink™ DNeasy Kit (Invitrogen, Darmstadt, Germany) according to the manufacturer's instructions. WGS was carried out on a MiSeq platform (Illumina Netherlands BV, Eindhoven, The Netherlands) using an Illumina Nextera XT library with 2 × 300 bp paired-end reads. Raw reads were assembled using SPAdes Assembler v.3.5 [135] and finally contigs with a size >500 bp were used for annotation and subsequent analysis.

3.6 Results and discussion

3.6.1. Sequencing and Genomic information

The draft genome sequences obtained from Illumina sequencing yielded, median value numbers of contig > 500bp is 116 (± 53) and genome size median value 5.04 mb (± .28 mb) median GC content 50.66 N50 value 155655 (± 84795). MLST identified the presence of 87 different ST types among all the 281 studied isolates. Overall in our studied population $bal_{\text{CTX-M-1}}$, $bal_{\text{CTX-M-15,}}$ $bal_{\text{CTX-M-14}}$ were predominant in the compartments (Figure 3.1 & 3.2). Our MLST analysis reviled fourteen major ST types (ST10, ST410, ST131, ST744, ST88, ST167, ST224, ST101, ST69, ST117, ST135,

ST361, ST648 and ST354). Among all fourteen ST types, nine ST types (n<5) were predominant in all four compartments. Human isolates (~51%, 43/84) were members of the major ST types such as, ST131, S10, ST224, whereas isolates in companion animal (~39 %, 25/64) were predominantly associated ST10, ST410, ST131, ST88, ST167, in the case of livestock 46% (33/71) isolates were predominant by ST410, ST10, ST744, ST101, ST135, food isolates 39% (24/61) ST167, ST410, ST744, ST117. Overall, ST410, ST88, ST10 types were observed in all of our studied compartments. In a total of 58 different *fim*H allele and 126 serogroup had identified in our investigated isolates. From draft genome assembly, we categorized the pan (all the genes present in an organism), core (gene shared by all the isolates) and accessory (gene present specific set of isolates) genome. Our *in -silico* analysis indicate the large pan genome containg 36,365 genes of which approximately 14.5% (n=2488) constituted the core genome. Further analysis revealed 342 genes were soft-core genes (in 99% and <= 100%, in 95% and < 99 %), shell genes were 3045 (in 15% and < 95%) and cloud genes 30,493 (in 0% and < 15%). The core genome composed of a 2.49 mb where 4,46,553 polymorphic sites were identified. This large number of the accessory genome indicates a high degree of diversity of ESBL *E. coli* among different habitat. To investigate overall genomic similarity and relationships of all isolates we did SNP based core genome alignment and generated a maximum likelihood phylogenetic tree (Figure. 3.3). Core genome analysis revealed that all the isolates formed few lineages and each lineage composed of isolates from the different compartment.

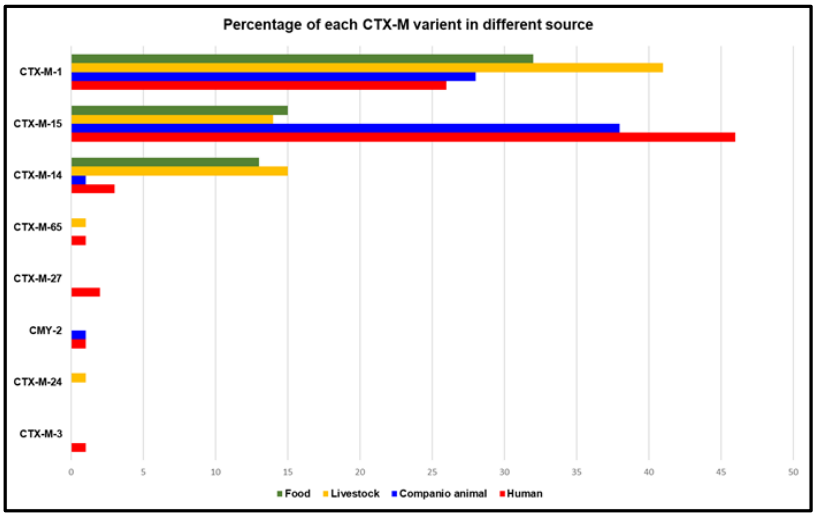


Figure 3.1: Percentage of each CTX-M variant in 281 ESBL-encoding *E.coli* isolates in different host.

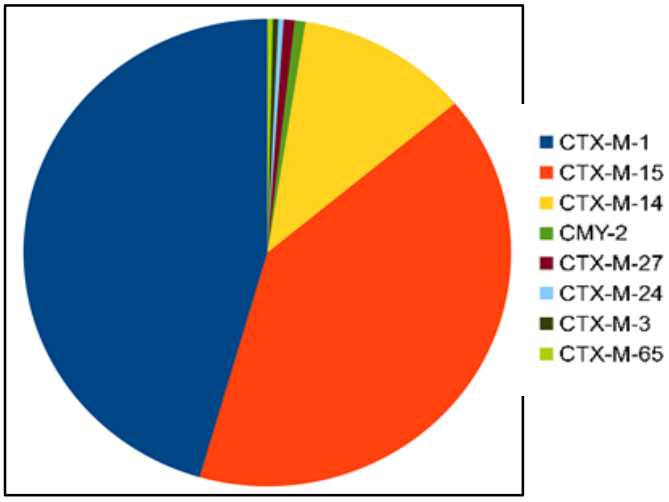


Figure 3.2: Prevalence of predominant ESBL genes in 281 ESBL-encoding *E. coli* studied isolates.

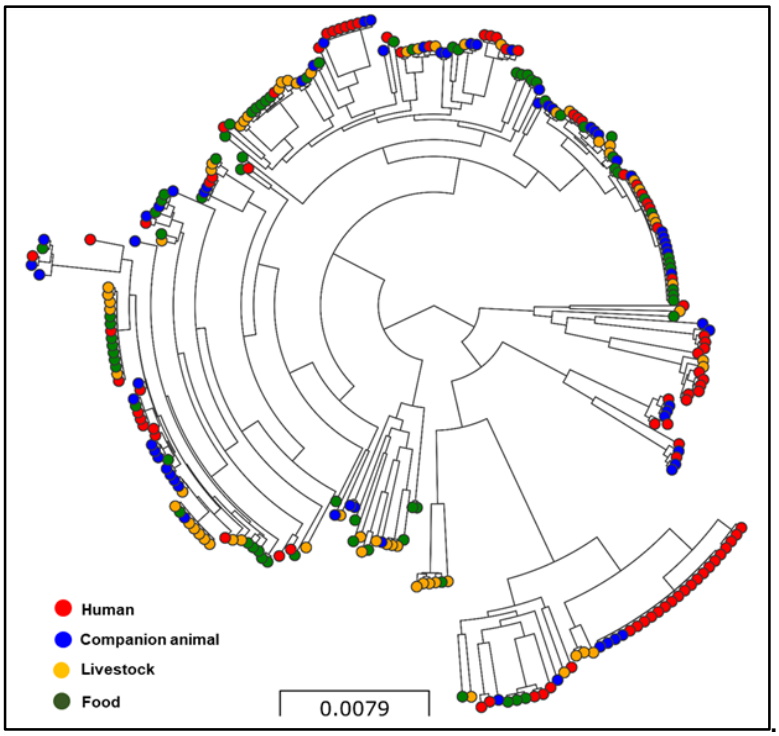


Figure 3 3: Genome phylogeny overview of 281 ESBL- encoding *E. coli*. The phylogeny was determined considering the core genome alignment and maximum likelihood reconstruction phylogeny by using Parsnp program implemented in Harvest suite [136].

3.6.2. Phylogenetic analysis

Next, we employed a combined analysis of core genome and the accessory genome of studied isolates. The pattern of accessory genome demonstrated a high level of concordance between the core genome tree and a tree built from the presence/absence of accessory genes. This allows us to separate the entire population into nine major clusters. This combined analysis suggests that each cluster has a

unique set of accessory genes, which could be generated through an early evolutionary process of each cluster (Figure 3.6) This analysis indicates that during evolutionary process each cluster acquired some accessory genome and is then evolving separately thereby forming separate lineages, which may have a role in niche adaptation. The pairwise SNP distance between isolates showed that the population was highly structured and composed of a closely related cluster with far more distantly related clones (Figure 3.5). According to SNP density cluster analysis revealed a total of 89 different clusters with a total of 446,553 polymorphic sites present in our studied population. All the clusters further classified into nine major clusters.

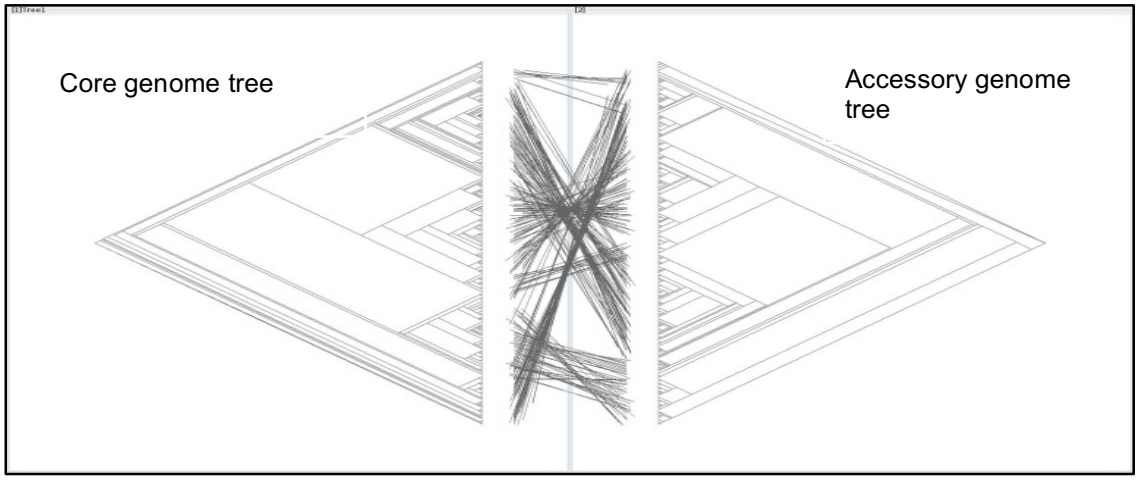


Figure 3.4: Comparison between the core genome (left) and the accessory genome tree (right). The accessory genome tree was constructed from the presence and absence pattern of genes and a core genome maximum likelihood tree were constructed from the variation in core genome. The connected line indicate the identical isolates of the two trees.

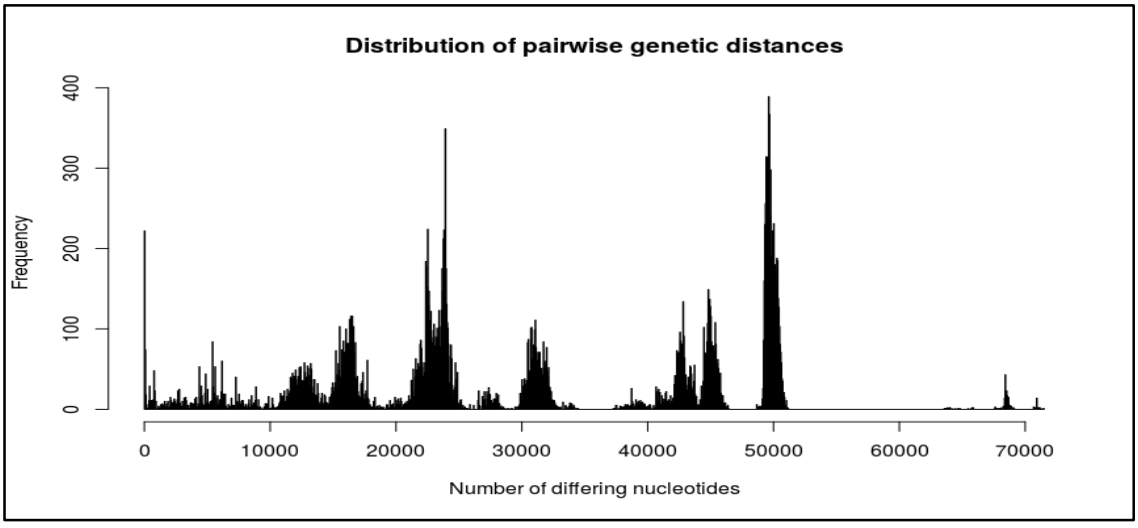


Figure 3.5: A histogram of pairwise SNP distances between all 281 isolates. The SNP distances were plotted based on the core genome alignment.

3.6.3. Cluster composition analysis

Our distance based clustering approach reveled the entire population to be comprised of nine predominant clusters, which harbor isolates from different compartments with a limited number of genetic distances (e.g. within the cluster three <100 SNPs, Figure 3.5). It supports one general hypothesis that ESBL *E. coli* is widely disseminated among different host types with closely related isolates. MLST composition of each cluster identified only few major ST types or similar clonal complex which are predominate in all clusters. We also identified one cluster (cluster IV) to be only composed of the isolates from human and companion animals. Which could be an evidence that certain ST types (e.g. ST131) can only be found in human and companion animal. We use the evolutionary time variation of each cluster to calculate the recent nucleotide substitution rate and the role of genetic recombination to form each cluster (not included in the thesis).

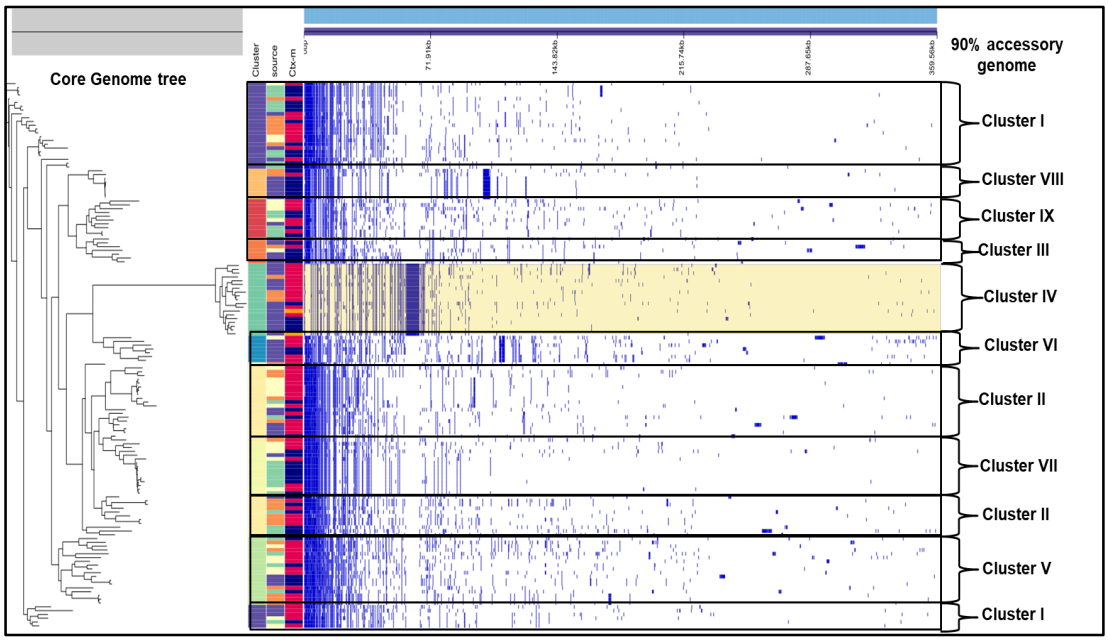


Figure 3.6: Maximum likelihood tree constructed from the core genome with gene presence (blue) absence (white) matrix of accessory genome elements.

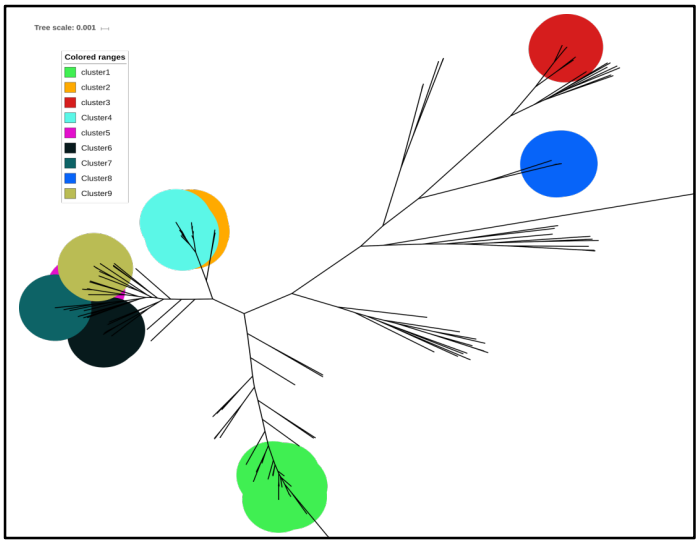


Figure 3.7: The SNP- distance based Neighbor Joining tree were generated using RaxML.

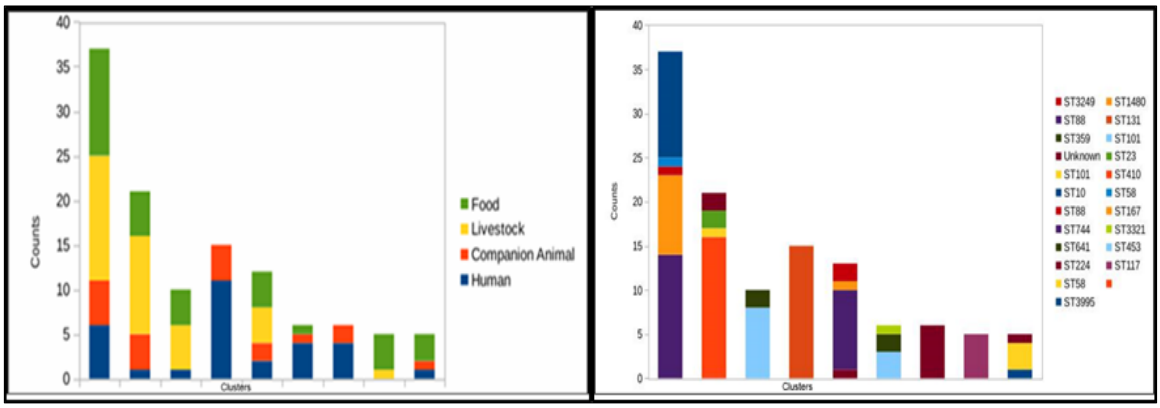


Figure 3.8. Cluster composition analysis Left side represent all the source of isolates of each cluster and right side represent MLST types of each cluster.

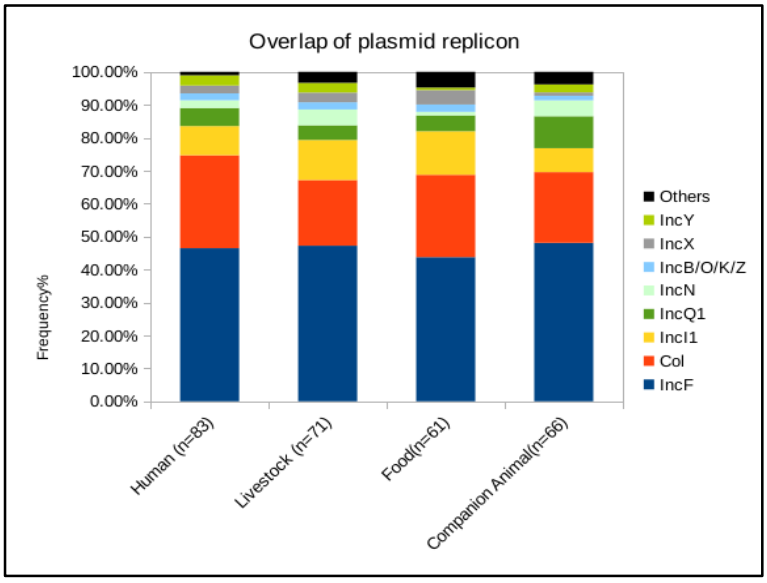


Figure 3.9. Frequency of overlapping plasmid replicon types of ESBL encoding *E coli.*

3.6.4. Plasmid analysis

Plasmids are important “vehicles” to spread of antibiotic resistance gene. The identification of complete structure of plasmid is always challenging from short reads sequencees. We therefore searched for Inc types of all the plasmids from plasmid finder database. plasmid Inc type analysis indicated that there are in a total thirty different replicon types present in our investigated dataset, which further clustered into nine major replicon types. This analysis shows there is a great diversity in plasmid types. In spite of huge diversity in plasmid Inc types, IncF plasmids ((IncFII /IncFIB, IncFIA/) were the most frequently observed replicons types in our investigated strains. Among all of those IncFII/ IncFIB predominant which indicates an intrinsic character with all those strains (Figure 3.9).

3.7 Conclusion

The current study describes an in-depth genome sequence analysis of a collection of ESBL encoding *E. coli* across Germany. Our genomic and phylogenetic approaches elucidated the population structure and dynamics of ESBL *E. coli*. Overall, the observed ESBL *E. coli* population in Germany,are highly diverse with a large accessory genome. Based on SNP and gene level the entire population clustered into nine major cluster and five minor clusters. Each cluster is predominated by either a single sequence type or a clonal complex type. There is always overlap between core genome as well as the accessory genome pattern also such as resistance gene and plasmid types. From plasmid analysis, we observed that multiple IncF replicons were present, which further classified into eight major combinations. Among all the observed Inc types, IncF plasmid predominate within all compartment and mostly associated with ESBL resistance gene. Our findings support the suggestion that ESBL –*E. coli* already well adapted in different ecological habitats. Our high-resolution whole genome sequencing data enable us to illustrate the population structure and dissemination of *ESBL E. coli* national wide.

.

4. Genomic portrait of *E. coli* ST131: The emergence of a globally disseminated clone

4.1 Publication

Manuscript: under preparation, (This study is the part of unpublished thesis work of HG).

4.2 Contribution

The author (H.G) was part of designing of the study under the supervision of TC. He performed the data analysis and carried out subsequent analysis such as developing and designing a bioinformatics workflow with BB and SD, for 3rd generation genome assembly and annotation.

4.3 Abstract

Since the initial emergence in 2008 *E. coli* sequence type 131 (*E. coli* - ST131) has spread explosively throughout the world and emerged as a major antimicrobial-resistant pathogen. *E. coli* ST131 often associated with fluoroquinolone-resistane (FQ-R) and causing bloodstream infections (BSIs) and urinary tract infections (UTIs). Recent population genomics studies elucidated the fine clonal structure of ST131, which comprises into three major lineages A, B and C, and multiple sub lineages (H30, H30-R, and H30-Rx) with the distinct resistance profile. The reasons behind the globally dissemination and expansion of ST131 sub clones in different ecological niche remain undefined. Here we analyzed 30 *E. coli* ST131 genomes (n=18: in-house, n=12: public database) of ESBL and non-ESBL ST131 by using PacBio RSII long-read sequencing platform. We determined the complete closed genome of 18 *E. coli* ST131 that shared ~64% of their core genome and a highly conserved IncF plasmid as a core component. On the other hand, ~36% genome content and non-IncF plasmid formed the accessory genome, which comprised of prophages, IS-associated elements, and acquired genetic islands that harbor metabolic genes for low energy substrates, type IV secretion systems, iron uptake, toxin anti-toxin systems and antibiotic resistance genes indicated high genomic plasticity. Those acquired genetic components could be accounted for the overall success of the ST131. The antibiotic resistance gene cassette (bla_{CTX-M}) was variably present on both IncF and non-IncF plasmid as well as

at different positions on the chromosome indicating that it has been acquired independently. Spread of the *bla*$_{CTX-M}$ cassette could be attributed to the plasmid-encoded IS26 dependent translocatable unit (TU). Our data provide a differentiated view of ST131 evolution, because the core elements in ESBL and non-ESBL genomes are conserved, we conclude that ST131 represented an already highly adapted clone that is ancestral to these habitats. Subsequent horizontal gene transfer (HGT)-based acquisition of antibiotic resistance and genomic island (GIs) together with niche specific bacteriophage uptake promoted the emergence and expansion of these lineages.

4.4 Background and Introduction

The emergence of antibiotic resistant *E. coli* ST131 presents a substantial clinical challenge in developed as well as in developing countries [137] [138]. *E. coli* ST131 is more frequently predominant in humans compared to other hosts. Earlier some low-resolution typing based studies (such as MLST, PFGE) suggested that most of the *E. coli* ST131 show highly similar PFGE, virulence, antibiotic resistance gene profile [139]. Population genetic studies indicate that *E. coli* ST131 is often associated with fluoroquinolone resistance and belongs to phylogenetic group B2, serogroup O25:H4 with type1 fimbriae FimH30 allele. Worryingly *E. coli* ST131 often associated with *bla*$_{CTX-M}$ alleles, which confer the resistance against several antibiotics such as cefotaxime, ceftriaxone, ceftazidime, cefepime and monobactams. Recent, NGS based studies suggested *E. coli* ST131 evolved from a single source which further separated into lineages (A, B and C) and several subsequent sub-lineages (A, B0, B1, B2, B3, B4, B5, C0, C1 and C2) of ST131 [140] [141]. Among all these lineages, isolates from lineage C, particularly sub-lineages C1/H30R (associated with FQ-R) and C2/H30-Rx (associated with the ESBL CTX-M-15) were rapidly expanding and associated with human infection while lineage A and B appeared to be sporadic. Lineage A is represented to contained the *fim*H 41, lineage B harbour *fim*H22 whereas lineage C mostly associated with *fim*H30 allele. Moreover ST131 also associated with several ExPEC-associated virulence genes such as adhesion, cell protection, iron uptake, toxin yersiniabactin receptor, serum resistance (*papA*, *iha*, *kps*MTII, *iut* and *sat, fyuA*, *iss*, *traT*). Along with all these features ST131, commonly harbour different plasmids such as IncF, IncI1, IncN and IncA/C [141]. The current view of expansion of the lineage of ST131 was promoted due to use of fluoroquinolone and the

introduction of third generation cephalosporin since late 1980s. In addition, homologous recombination events have accelerated the evolution process. Several other hypotheses highlights acquisition of several different plasmids and phages as playing a major role. Overall, the draft genome sequence studies provide an idea about the epidemiological aspects to a greater extend while offering clues for the perspectives involved in the evolution of ST131, such as, specific connotation with antibiotic resistance traits, obligatory association of IS and antibiotics genes, cumulative prophages as well as genomic islands and homologous recombination. To get a holistic view of evolution of *E. coli* ST131 that comprises the core genome and accessory genome is currently not available. In order to get genomic portrait in more detailed in this study, our first aim was to identify the genome dynamics of ST131.. Secondly, to identify the role of introduction of prophages, genomic islands, IS elements and horizontal gene transfer in the evolution of ST131. Thirdly, characterization of plasmids composition of the E. coli ST131. Finally, to identify the unique genomic composition of each clade that make each clades more successful than others. This study employed third generation single molecule real-time (SMRT) sequencing technology to analyze the genome plasticity of ESBL and non-ESBL *E. coli* ST131 strains (n=30) isolated from human and companion animals. Closed genome analysis, indicated that the core components of ST131 are highly conserved in both ESBL and non-ESBL isolates. A conserved IncF plasmid harboring $bla_{ctx\text{-}M}$ allele along with a number of toxin anti-toxin systems and metabolic gene cassettes was also observed. Genomic variability is mainly driven by the accessory genome such as prophage, acquisition, and loss of genomic islands as well as integrase-associated genetic segments. Our findings provide a differentiated view of ST131 evolution: because ESBL and non-ESBL genomes share, conserved core elements and it suggest that ST131 represents a highly adapted clone that is ancestral to these habitats. Subsequent acquisition of antibiotic resistance and genomic island (GIs) together with niche specific bacteriophage uptake promoted the worldwide expansion of several lineages.

4.5 Material and methods

Sample collection, preliminary characterization and whole genome short read sequencing are described earlier in chapter three. Long read resequencing was performed for the 18 *E. coli* ST131 isolates for current study by Single Molecule Real Time (SMRT) technique on a PacBio RSII (Pacific Biosciences, Menlo Park, CA, USA). SMRTbell™ template libraries were prepared according to the manufacturer's instructions using the 20 kb template preparation using BluePippin™ size-selection system. De novo long read genome assembly was carried out using the RS_HGAP_Assembly.3 protocol. Contigs obtained from each independent de novo assembly was error corrected by using illumina short reads, trimmed, circularized, and adjusted to *dna*A as first gene for all chromosomes. Chromosomes and plasmids were annotated using Prokka 1.11 [142] followed by manual curation. *In silico* MLST was performed, pan genome, genomic islands and prophage analysis performed by using PHAST [143]. Virulence genes were identified using, an in-house virulence gene database using LS-BSR package [144]. Plasmids were experimentally determined by S1-nuclease digestion followed by pulsed-field gel electrophoresis as described previously. From sequencing data plasmid, incompatibility groups were determined by PlasmidFinder [126].

4.6. Result and discussion:

4.6.1. The genome size and core genome of ST131 genome

Closed genome comparison identified that the entirely studied ST131 chromosomes were highly syntenic, with two small exceptions consisting in an inversion of 957 kb and 257 kb in H049 and H132. The genome size of ST131 considerably varied from 4,717,338 - 5,260,260 bp (Figure 4.1). All investigated isolates belonged to phylogroup B2 with serogroup O25b:H4 (n=15), O16:H5 (n=3) and O150:H5 (n=1). The whole-genome SNP phylogeny grouped to three major clonal lineages (A, B and C) of ST131. Core genome analysis identified in a total of 3,433,668 bp regions shared among all the investigated lineages, which included 3547 core genes (99% <= studied isolates <= 100%), 329 soft-core genes (95% <= studied isolates < 99%), 1181 shell genes (15% <= studied isolates strains < 95%) and in a total of 7460 (0% <= studied isolates <= 100%) genes were identified as pan genome. Lineage C observed more stable core genome compare to A and B, with 4323 core genes.

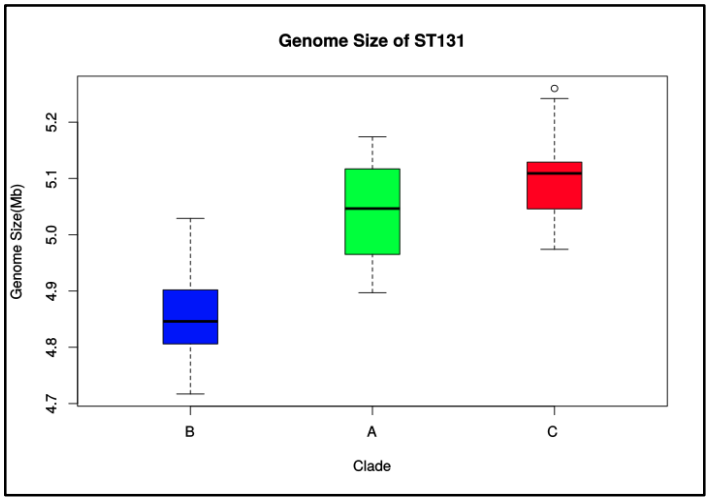


Figure 4.1: Box-plot comparisons of genome size estimates based on complete genome of *E. coli* ST131 three different lineages (lineage A, lineageB, lineage C). LineageC represent highest genome size.

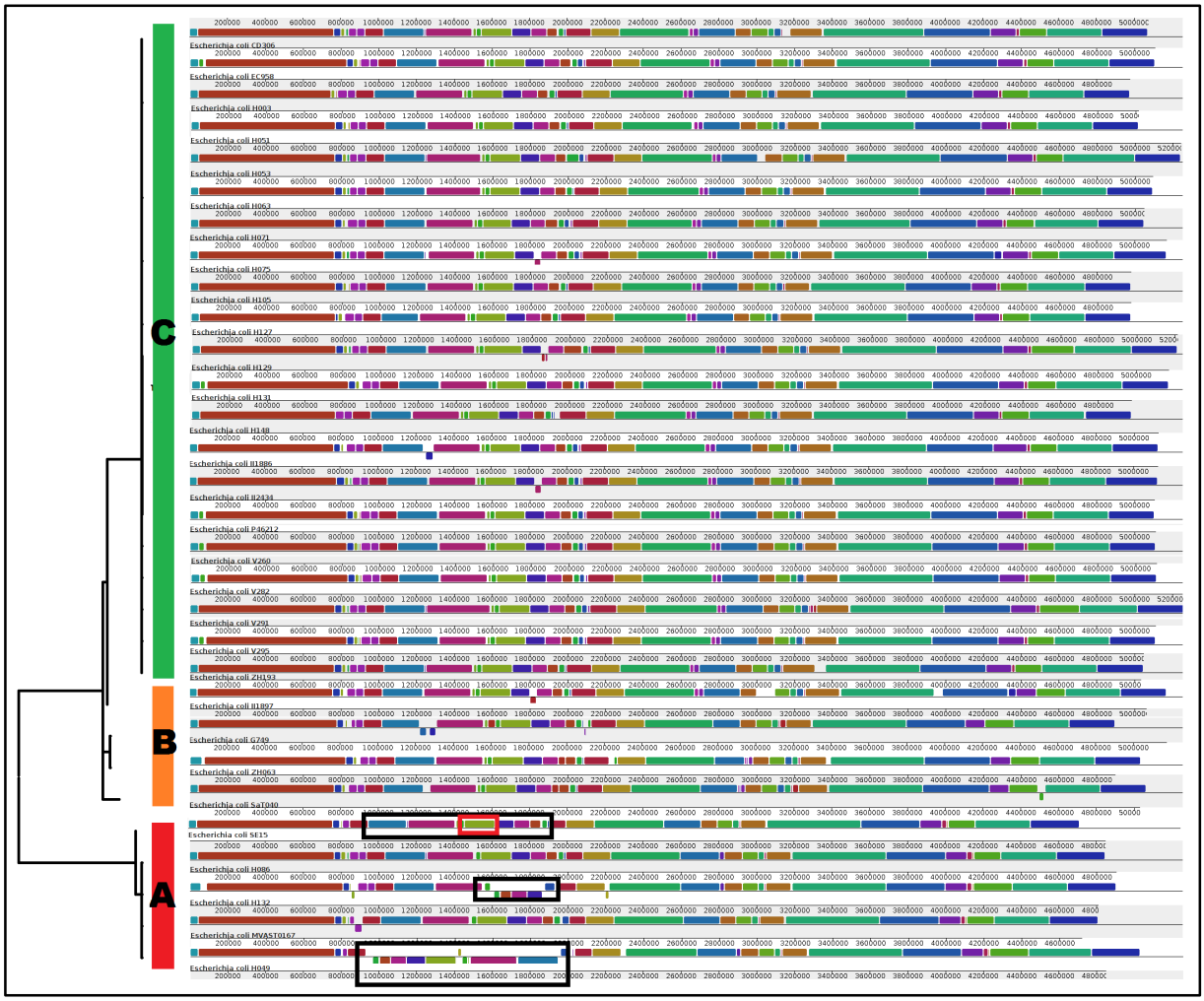


Figure 4.2: Chromosomal synteny of *E. coli* ST131 genomes. Pairwise alignments of genomes were generated using Mauve. All genomes were aligned to EC958 as a reference. The colored bars inside the blocks are representing similar homologous locally collinear blocks.

4.6.2. The chromosomal synteny of ST131 genome

A total of 525-545 chromosomal locations was identified to be as recombinant within the ST131 lineages that disturbed the genomic synteny. These chromosomal regions, mainly generated due mobile genetic elements (e.g. transposons, genomic island, IS elements) which mainly resulted from horizontal gene transfer mechanism. Notably, in case of lineage C there were ~544, variable regions (VRs) observed. We classified these VR according to their gene content (minor-VR: regions that contained less than five genes, medium VR: 6-15 gene and major-VR: more than 15 genes were assigned). The minor-VR contributed ~94% of VR and mainly consisted genes that due to point mutations, transposons, IS element or insertion of IS elements, whereas the medium and major VR (altogether ~6%) were observed generated due to horizontal gene transfer events such as prophages and GIs. A total of 39 major variable regions (assigned with serial alphabets from A-AL) were observed, which formed due to acquisition of prophage and genomic island (Figure 4.2).

4.6.3. Diversity within all the ST131 lineages

Comparing 30 closed ST131 genomes identified 13 VR that encode genes for functional proteins. Of these HGT segments, eight VR was associated with tRNA, thus identified to be GIs while others five were without tRNA and assigned as unclassified-VR (U-VR). All the GI and most of the U-VR flanked by integrase gene on either end, thus are the forms of integrative elements. Unlike prophages, different GI and U-VR encoded genes with functionally diverse categorical proteins and offered various new capabilities to the strains. Complete or partial metabolic gene operons for utilization of metabolite such as glutathione (*gsi* operon), glutamine (*gln*), galactitol (*gat*), galactarate (*gar)*, propanediol (*pdu*), dihydrogen-fumarate (*hyb*), ethanolamine (*eut*), glucitol (*srl*) and iron (*fec*) were observed on various GI and U-VR. Virulence factors such as aerobactin synthase (*iuc* operon), secreted auto-transporter toxin (*sat*), enterotoxin (*esp*C), afimbrial-adhesion (*afa*), antigen 43 (Ag43), flagella (*fli*, *flg*), fimbriae (*dra*), pili (*pap*), increased serum survival (*iss*), chemotaxis (*laf*); type II secretion system (*eps*), and capsule (*kps*) were also observed to be encoded on GIs and URs.

A careful observation indicates the uniqueness of 55 genes to lineage C. These genes were found to be present at a unified location of the chromosome and this location was nothing but the genomic island associated with tRNA-Leu. These genes revealed presence of five gene-clusters that were encoding- restriction modification (RM)

system, Fe (+3)-dicitrate transport system (*FecBCDE*), choline transport, while function of two gene cluster remained speculative. The presence of unique RM systems has been shown to be involved in the shaping lineages. The RM system present in lineage C was an unusual combination of Type IV (*mrr*) and Type I (*hsdM-hsdS-hsdR*) system forming 'immigration control region'. The Fe (+3)-dicitrate transport system known mainly for the siderophore based iron acquisition. Nevertheless, once the ferric citrate is taken up citrate is released along with iron, which is an important intermediate of TCA cycle. Low carbohydrates environment such as urine was invading strains has to survive on the limited resources, citrate can act as an important source of energy. Additionally, citrate and iron present in the human urine that forms a ready-made chelation complex can be captured by Fe (+3) dicitrate transport system. Thus, for bacterial strains invading the urinary tract, presence of Fe (+3)-dicitrate transport system is an advantage. Third gene cluster observed was responsible for the high affinity choline transporter (*BetT*), which is responsible for superior osmoprotection by Choline over Glycine or Betaine. Consequently, these uniquely occurring genes elucidate the formation of different lineages of ST131 as well as suggestive of host-adaptation (Figure 4.3).

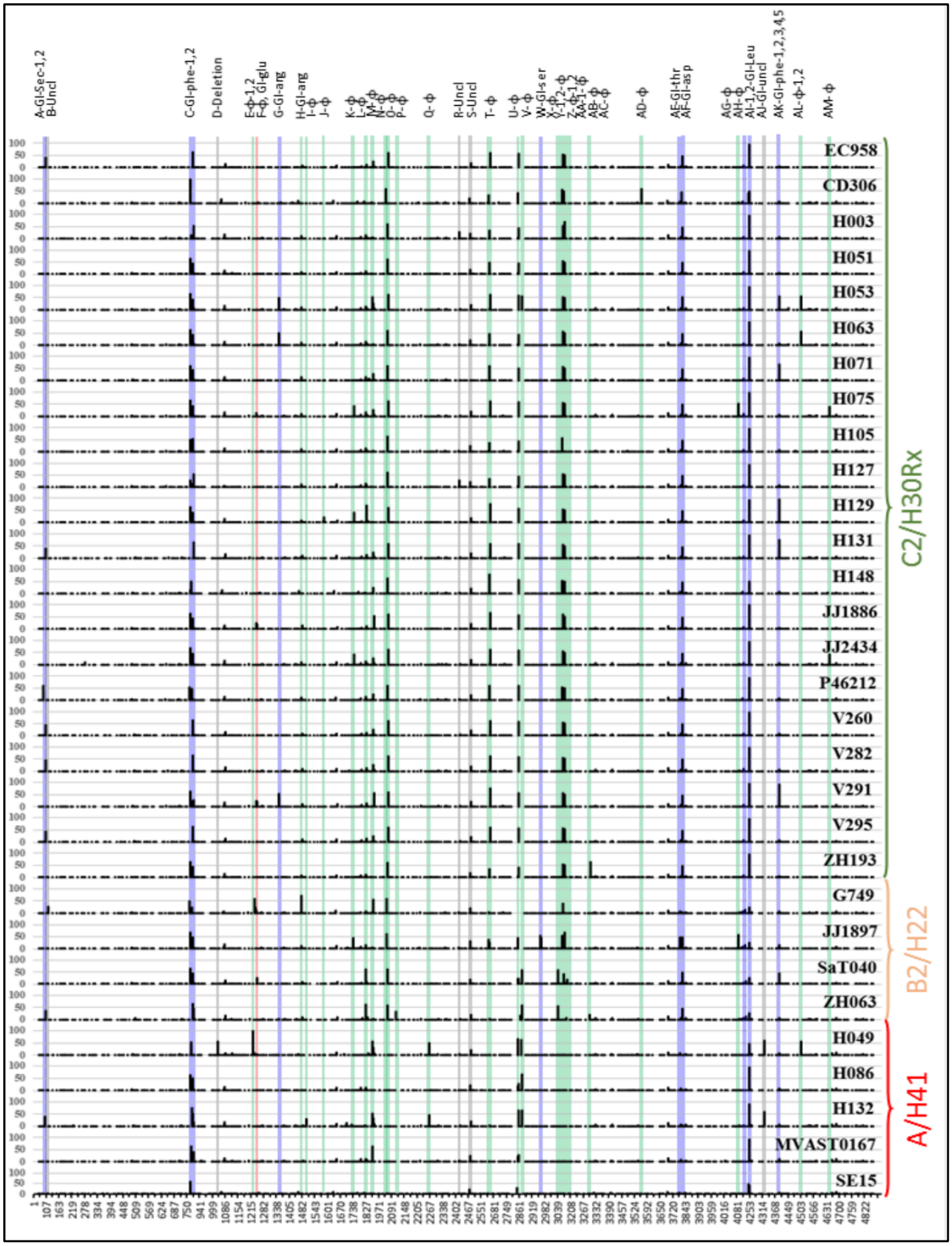


Figure 4.3: Variable regions of the ST131 strains. A plot was constructed based on the core genes (x-axis) (referenced to EC958 co-ordinates at the bottom) and the genes that disturbed the synteny (y-axis). The factors that caused disturbance in chromosomal synteny (or variable regions) were revealed to be prophages, genomic islands, unclassified region and regions containing more than one type of factors (red) are marked. The right side vertical labels denote the lineages. (Key: GI = genomic island, ϕ = prophage, Uncl = unclassified variable regions).

4.6.4. IncF plasmid replicon as a single conserved plasmid of ST131

All studied ST131 strains contained at least one plasmid. A total of 45 plasmids (> 7 Kb, circularized, 1-4/strain) was identified from 30 strains that varied from 24.4 - 209 Kb in size (avg size ~110kb). Plasmid analysis was made based on replicon types, backbone structure and phylogenetic approach (Figure 4.4). The metadata of plasmids consist of two major groups, such as IncF and non-IncF allele. Moreover, the IncF allele was found to be the only active allele. The plasmid backbone alignment showed similar collinear blocks shared across the IncF plasmids, suggesting the conserved backbone structure. The main variation observed among the commonly observed within IncF plasmids were associated with the IS- and integrase- associated elements (Figure 4.5). These elements carried cargo of toxin-antitoxin systems, sugar metabolism and antibiotic resistance genes.

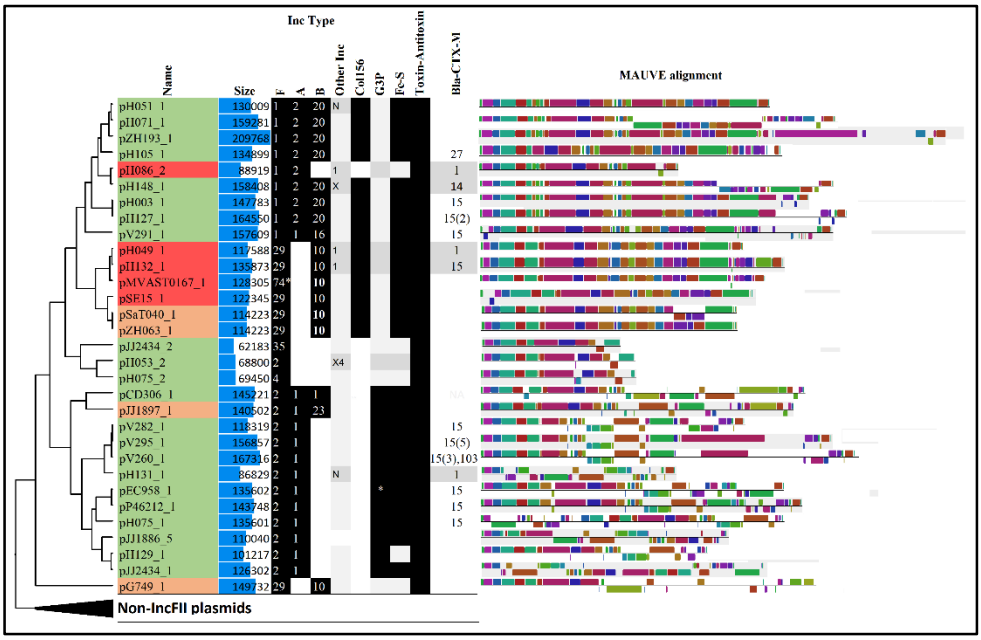


Figure 4.4: Comparative analysis of the IncF type plasmids of ST131 strains. IncF type plasmid showed similar backbone structure as demonstrated by MAUVE alignment on the right side. Dendrogram based on the DNA-DNA similarity of plasmids (left side). Name of the plasmids are highlighted as per their lineages A (red), B (orange) and C (green). Black squares represent the presence, while the number in it gives allele details. Grey square represents presence of a particular trait in non-IncF plasmid of the same strain. Key: G3P= IS-associated cassette of Glycerol-3-phosphate, Fe-S = IS-associated cassette of ron uptake system, *bla*$_{CTX-M}$ allele and their copy number in parentheses, * = variants

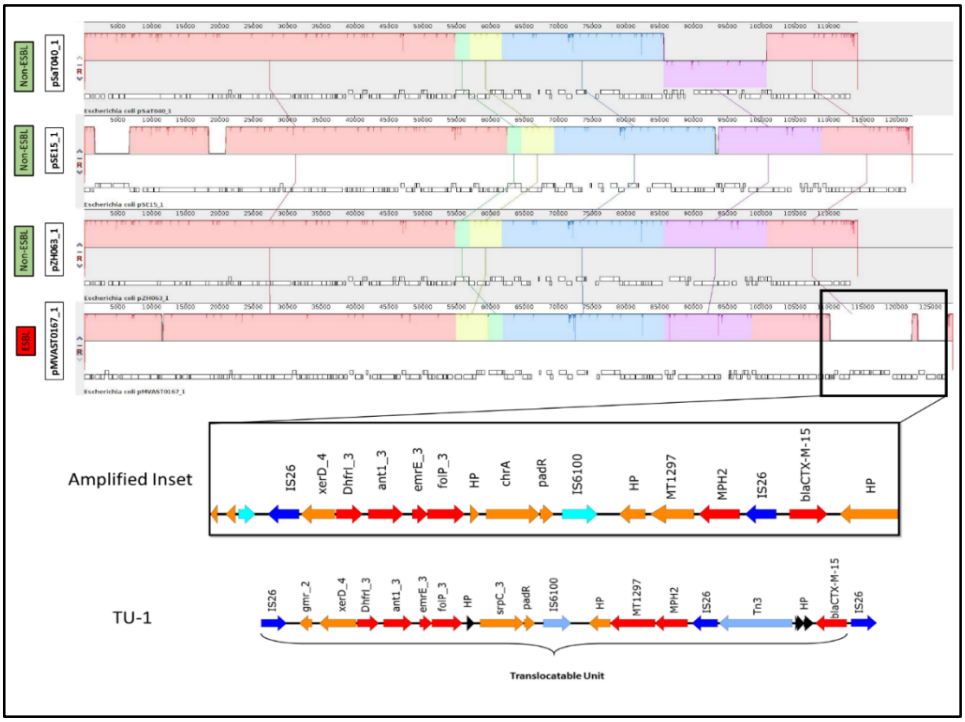


Figure 4.5: Comparative analysis of the non ESBL IncF type plasmids and ESBL IncF plasmids of ST131 isolates. Major difference observed in IS associate cassette region.

4.7 Conclusion

The results of this study provide compelling evidence that clonal expansion is the dominant mechanism for the proliferation of both ESBL encoding and non-ESBL also From the closed genome analysis, we observed that the core components of ST131 are highly conserved in both ESBL and non-ESBL isolates. In comparison with the earlier draft genome based study [145], where authors identified a large plasmid diversity, our closed genome study found a differentiated view about plasmid diversity. The current study revealed a conserved IncF plasmid harboring *bla*$_{\text{ctx-M}}$ allele along with a number of toxin-antitoxin systems and metabolic gene cassettes. Genomic variability is mainly driven by accessory genome such as prophage, acquisition and loss of genomic islands as well as integrase-associated genetic segments. Our findings provide a differentiated view of ST131 evolution: because ESBL and non-ESBL genomes share, conserved core elements and it suggest that ST131 represents a highly adapted clone that is ancestral to these habitats. Subsequent acquisition of antibiotic resistance and genomic island (GIs) together with niche specific bacteriophage uptake promoted the worldwide expansion of several lineages.

5. Emergence of *bla*$_{CTX-M-27}$-encoding *E coli* ST131 lineage C1/M27 clone in clinical isolates in Germany

5.1 Publication

Emergence of *bla*$_{CTX-M-27}$-encoding *E coli* ST131 lineage C1/M27 clone in clinical isolates in Germany

Ghosh H, Doijad S, Falgenhauer L, Fritzenwanker M, Imirzalioglu C and Chakraborty T*

Emerging Infectious Diseases EID-17-0938 DOI: 10.3201/eid 2310.170938

Complete Genome Sequence of *bla*$_{CTX-M-27}$ -Encoding *E coli* Strain H105 of ST131 Lineage C1/H30R

Ghosh H, Bunk B, Doijad S, Schmiedel J, Falgenhauer L, Spröer C, Imirzalioglu C, Overmann J, Chakraborty T*

Genome announcement: Volume 5 Issue 999 e00736-1

5.2 Contributions

The author (H.G) conceived and designed both the studies together with other authors. He processed data for both the publication, carried out all subsequent analysis, designed, and developed workflows of data analysis under the supervision of T.C. He did genome assembly, annotation and subsequent analysis. Furthermore, he drafting the manuscripts and responded reviewer comments with other authors.

5.3. Abstract

E. coli ST131 became high-risk pathogens causing infection humans and animal. Retrospective studies suggest strains, particularly one clade, C/H30R of ST131 are spreading extensively worldwide and frequently associated with urinary tract infections (UTIs) and bacteraemia. ESBL production is predominant with subgroup fluoroquinolone resistant *E. coli* ST131-C2/H30 which conferred by the CTX-M-15 allele. Recently another emerging subclade of *E. coli* ST131, C1/H30R1-M27 has been reported, which often associated with *bla*$_{CTX-M-27}$ allele and dramatically increased in Japan and France. Our study reported the emergence of ESBL producing *E. coli* ST131 C1/H30R1-M27 in Germany. We highlighted here that the incidences of *E. coli* ST131 C1/H30R-*bla*$_{CTX-M-27}$ isolate increased from zero to 45% in 2009 to 2016. The

whole genome comparative study suggested that genomic distance of all the German and Japan isolates separated though only with few SNPs, suggesting a clonal origin. Our data suggest an ongoing shift in ESBL allele of *E. coli* ST131 clone associated with *bla*$_{CTX-M-27}$ alleles, which could be a major warrant for further attention.

5.4. Background and Introduction

Over the past twenty years, *E. coli* ST131 have emerged as a prevalent vehicle for the spread of extended-spectrum β-lactamases (ESBL) worldwide. This is particularly true for isolates of the clade ST131 C/H30R that are associated frequently with urinary tract infections (UTIs) and bacteremia [138] [137]. While the ESBL-production of predominant subgroup ST131 C2/H30 is conferred by the CTX-M-15 allele (Rx), the emerging subgroup C1 is often associated with yet another CTX-M alleles such as -14 and -27. An increase of C1/H30R-bla CTX-M-27 ST131 isolates were first time reported for clinical isolates in Japan [146]. More recently, a dramatic rise from 0-65% in the incidence of ST131 C1/H30R-bla CTX-M-27 isolates in France between 2010 and 2015 has been reported [147]. In addition, there are sporadic reports of ST131 isolates harboring *bla*$_{CTX-M-27}$ from other countries [146]. To investigate the possible expansion of *bla*$_{CTX-M-27}$ encoding ST131 in Germany, we examined ESBL-producing isolates obtained from livestock, human, companion animals, food and environment between 2009 to 2016 (Figure 5.1). Here we report that *E. coli* ST131 C1/H30R-*bla*$_{CTX-M-27}$ is exclusively present in human populations and that its incidence increased from zero to 45%.

5.5. Material and Methods

In two nationwide research projects (DZIF, RESET) a total of 4,386 non-repetitive phenotypically ESBL-producing *E. coli* isolates were obtained from humans, livestock, and companion animals. For this study, a representative subset of 953 ESBL isolates were subjected to whole genome-sequenced using the Miseq/NextSeq (Illumina, The Netherlands). Briefly, genomic DNA was isolated from overnight cultures by using the Purelink Genomic DNA Mini kit (Invitrogen, Darmstadt, Germany). For short read whole genome sequencing, an Illumina Nextera XT library (Illumina Netherlands BV, Eindhoven, The Netherlands) was used for sequencing on an Illumina MiSeq or Illumina NextSeq with 2x300 or 2x150 read length, respectively. In the second study to investigate the genomic organization in more detailed one *bla*$_{CTX-M-27}$ encoding *E.*

coli ST131 isolate H105 (Accession numbers: CP021454) belonging to lineage C1/H30R was sequenced for its complete genome using PacBio RSII system (Pacific Biosciences, USA).

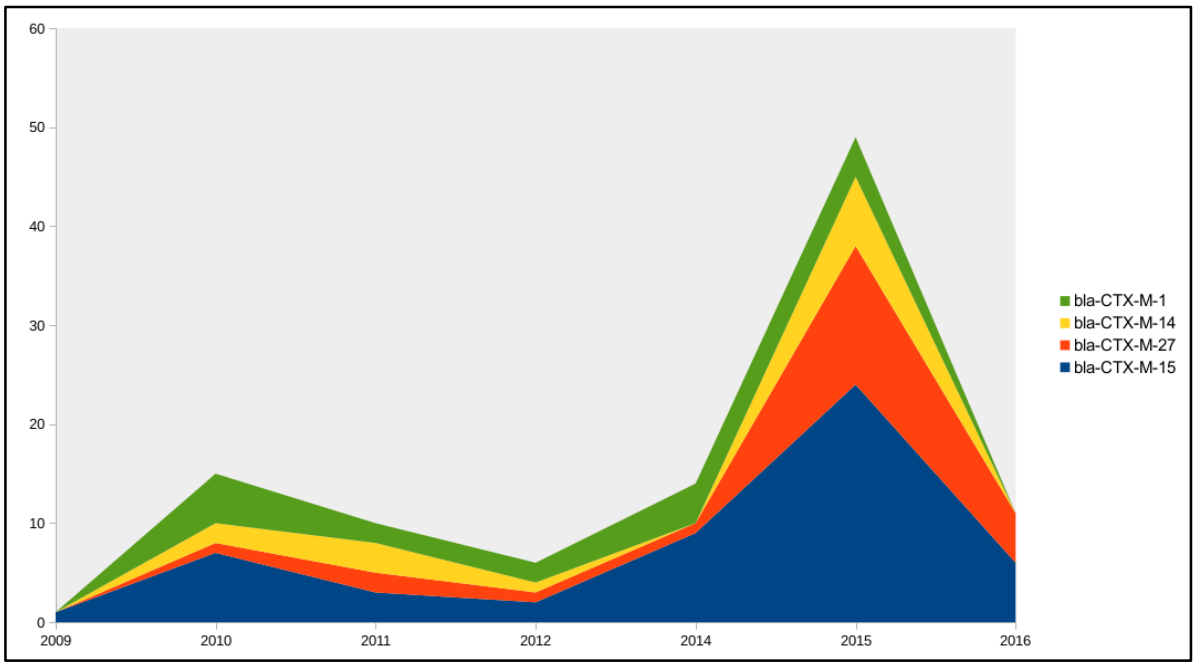
5.5. *In silico* analysis:

For genome assembly of Illumina reads we used Spades V.3.6 [135] and annotation were carried out using prokka V1.11 [142] with default parameter settings. Spades assembly yielded an average number of contig 149 N50 value 102438 bp and genome size ~5mb. PacBio data of H105 de novo assembled based on 59,447 PacBio long reads with an average read length of 10,355 bp using "RS_HGAP_Assembly.3", included in the SMRT Portal version 2.3.0. Later on Illumina short-reads were mapped onto the assembled sequences of H105 using BWA in order to obtain a highly accurate genome with QV60 final quality. Assembly quality was assessed through QUAST v2.3 [105], and contigs with >500 bp were considered for further analysis. Multilocus sequence typing (MLST) was carried out by 'mlst-package' (). The *bla*$_{CTX-M}$ profiles, FimH type, serotype and virulence gene were determined by Resfinder, FimTyper, SeroTypeFinder, VirulenceFinder respectively (https://cge.cbs.dtu.dk/services). Plasmid incompatibility and plasmid multilocus sequence typing were identified by PlasmidFinder and pMLST, respectively. The presence of M27PP1 region was confirmed by BLASTN. For core genome analysis, draft genomes were compared along with isolates from Japan (n=13) using Harvest Suite [136] with a default parameter by using EC958 [148] and H105 as a reference genome for in between and within the clade.

5.6 Result and discussion

In silico multi-locus sequence typing (MLST) identified 17% (159/953) isolates as of sequence type 131 (ST131). Of these, *bla*$_{CTX-M-15}$ was the most prevalent (46%, n=73) followed by *bla*$_{CTX-M-27}$ (15%, n=24), *bla*$_{CTX-M-1}$ (11%, n=18), *bla*$_{CTX-M-14}$(9%, n=15), and other *bla*$_{CTX-M-3/11/17/24/36/47}$ (6%, n=10). All ST131 isolates with *bla*$_{CTX-M-27}$ were of serogroup O25b and harbour a *fimH*30 allele with the exception of a single strain that was of serogroup O16 with a *fimH*41 allele. In 20 of 24 cases, the isolates were harboring contigs with F1:A2:B20 plasmid replicons. The F1:A2:B20 plasmid is highly conserved in ST131 and ancestral to the H30R/C1 clade, as it is present in all the ST131 *bla*$_{CTX-M-27}$ isolates regardless of whether they harbor antibiotic resistance

genes (5). In the remaining isolates, *bla*$_{CTX-M-27}$ was detected on other non-conserved F1:A6:B20, F1*:A2:B20, F1:A2: B- and F29: A-:B10 plasmids. Core genome phylogeny revealed that all the ST131 C1/H30R *bla*$_{CTX-M-27}$ isolates had an average of 294 SNPs when compared to *E. coli* ST131 EC958 (Accession number HG941719). In contrast, isolates within the C1-M27 clade are only separated by 68 SNPs (Figure 5.2). Comparative genomic analyses revealed that isolates from Germany and Japan shared ~85% of the genome with an average difference of SNPs between these genomes indicating clonal and possible evolution from a single common ancestor. Recently C1-M27' clade was defined by the presence M27PP1 prophage-like region.



*2013 data not available

Figure 5.1: Proportions of different *bla*$_{ctx-M}$ allele in ST131 studied population during the 7-yr sampling framework. The number of isolates has been plotted by year ordered by the frequency. The emergence of *bla*$_{ctx-M-27}$ encoding ST131 observed in 2014 and 2016, respectively.

5.7 Conclusion

Our results provide evidence for the recent emergence of ST131 subgroup *FimH*30-O25b, clade C1-M27, harbouring *bla*$_{CTX-M-27}$ in Germany and reinforce observations made elsewhere. Our findings suggest an ongoing shift in CTX-M alleles associated with ST131 infections worldwide that now warrants further attention.

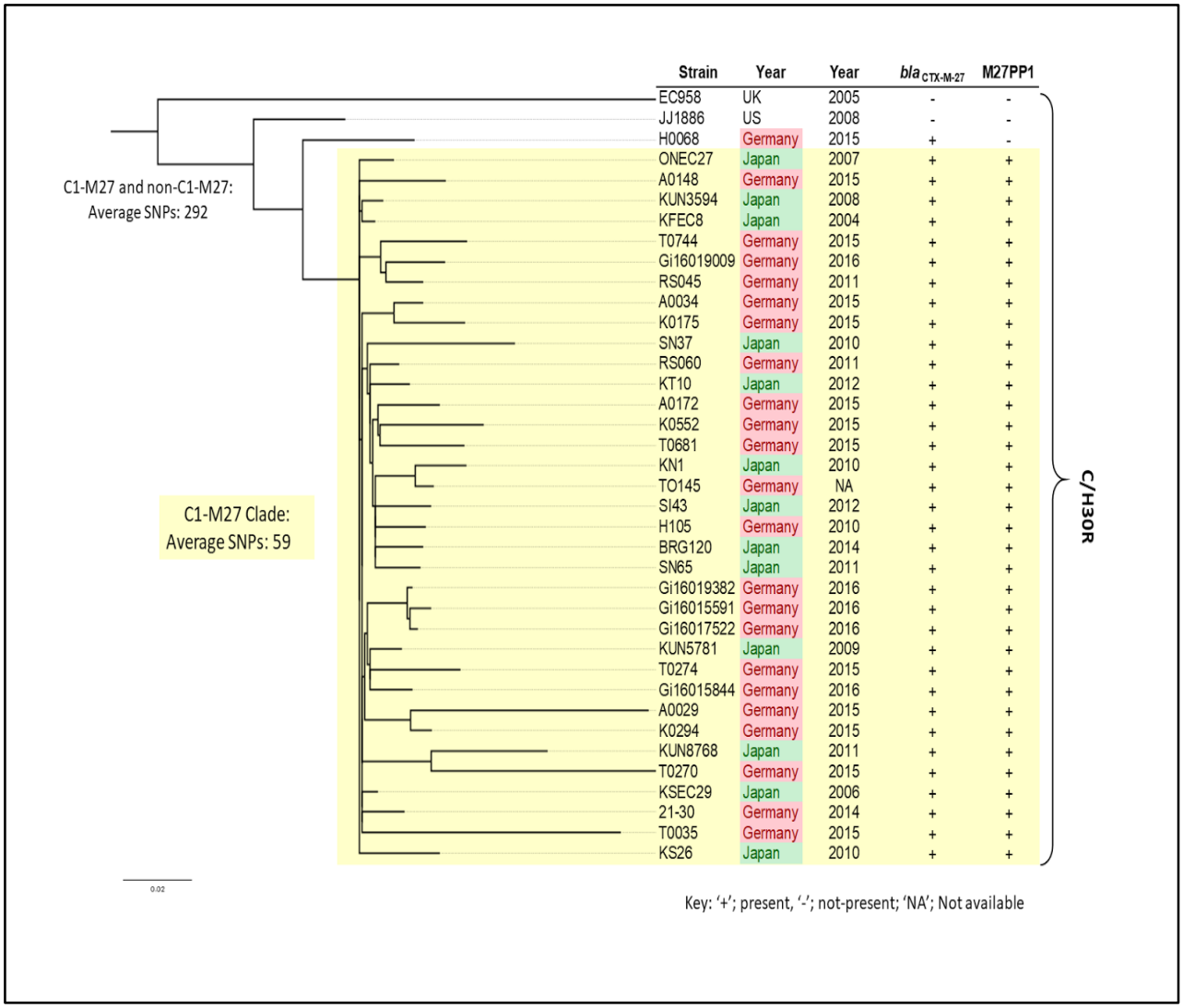


Figure 5.2: Core genome SNP based phylogenomic analysis of *bla*$_{CTX-M27}$-encoding *E. coli* ST131 from Germany and Japan (13). An average of 292 SNPs was identified in in between C1-M27 and non C1-M27.Whereas an average of 59 SNPs was observed within the C1-M27 clade. The isolates from Germany and Japan clustered together in a unified clade separated from non-C1-M27 clade (Source Ghosh, 2016).

6. Circulation of clonal populations of fluoroquinolone-resistant CTX-M-15-producing *Escherichia coli* ST410 in humans and animals in Germany

6.1 Publications

Circulation of clonal populations of fluoroquinolone-resistant CTX-M-15-producing *E.coli* ST410 in humans and animals in Germany

Falgenhauer L, Imirzalioglu C, **Ghosh H**, Gwozdzinski K, Schmiedel J, [...], Chakraborty T *

Int J Antimicrob Agents 2016 Jun; 47 (6):457-65.

Chromosomal Locations of mcr-1 and blaCTX-M-15 in Fluoroquinolone-Resistant *E. coli* ST410

Falgenhauer L, Waezsada S. E, Gwozdzinski K, **Ghosh H**, Doijad S, [..], Chakraborty T* *Emerging infectious diseases 22 (9), 1689-1691*

6.2 Contributions

The author (H.G) performed data analysis and drafting the manuscript on both the manuscript with LF and TC. Furthermore, H. G. carried out de novo assembly annotation and subsequent *in silico* analysis e.g. core genome SNPs analysis in between each clade by using Parsnp and virulence gene screening by using LS-BSR pipeline.

6.3 Abstract

The prevalence of ESBL producing multidrug-resistant *E. coli,* specially CTX-M-type ESBLs increasing in the community as well as in the hospital setting and became a major concern of one health issue. Here, we evaluated the relative abundance of different sequence type (STs) within the population of $bla_{CTX\text{-}M\text{-}15}$ encoding *E. coli,* which sampled from humans, companion animals, livestock and environments in between 2009-2014 from Germany. To get into further insights into the complex ecology we have chosen high-resolution whole-genome sequencing approach for our analysis. Our analysis revealed that the CTX-M-15 encoding *E. coli* highly diverse with 26 different sequence types (STs) were detected of which ST410 were predominant

within all the population. Core genome analysis revealed five (A-E) major clusters are present within ST410 population of which B and C shows a limited number of SNPs along with chromosomal insertion of $bla_{CTX\text{-}M\text{-}15}$ allele. Conclusively our analysis revealed the potential of clinically relevant multi-resistant bacteria *E. coli* ST410 found in different ecological niche, which could be a high-risk in interspecies transmission in coming future for one health approach.

6.4 Background and Introduction

The emergence of ESBLs encoding multidrug-resistant bacteria especially *E. coli* is increasing worldwide in the recent years [58]. Majority ESBL encoding genes are plasmid encoded and confers resistance to penicillins, cephalosporins and monobactams. Several types of CTX-M-type enzymes have been reported from different parts of world whereas majorities are CTX-M-15, CTX-M-1 predominant among different niche [149] [150]. To investigate the transmission and colonization of ESBL *E. coli* several epidemiological studies has performed in the UK, The Netherlands and Germany [151]. Most of these studies showed that isolates from different source share identical CTX-M alleles, common STs and have similar resistance plasmids. This suggests that they shared a common gene pool and few accessory genomes play a role in terms of adaptation in different ecological niches. Most of the studies are limited in terms of number or the low resolution analysis approach. Here we performed high resolution whole genome sequencing to identify the genetic relatedness of major STs within $bla_{ctx\text{-}m\text{-}15}$ encoding *E. coli* in animal and human populations from different locations in Germany. We highlighted majority of them are predominant by only few STs Types among them ST410 are predominant in all compartments. Core genome SNPs based phylogenetic analysis revealed five major clades within ST410. Few clades show limited number of SNPs with an overlap pattern of genetic content. Our findings suggest that the occurrence of frequency of *E. coli* ST410 is also high in the commensal niche, as well as hospital settings.

6.5 Material and Methods

In between 2009 -2013 a total of 111 phenotypically ESBL producer isolates from different sources (humans, animals, and environment) were collected under two national wide antimicrobial surveillance program (RESET) and the German Center for Infection Research, DZIF (http://www.dzif.de/). Preliminary ESBL production and

presence of CTX-M-15 resistance gene were identified using double disc synergy test and PCR. For whole genome sequencing were performed as mentioned proceeding chapters.

6.6 Analysis of Sequence Data

Raw illumina reads were assembled using Spades (version 3.0) [135] and contigs size over 500 bp were selected for further analysis. Core genome phylogeny and density of single nucleotide polymorphisms (SNPs) were analyzed using ParSnp [136] and Ginger under the Harvest Suite. *In silico* MLST and resistance gene, pathogenicity island and plasmid incompatibility groups were identified by web based MLST tool, resfinder [152] PAI DB and plasmid finder [126] respectively. For virulence gene, screening in house local virulence gene database were used and gene screening were performed by using LS-BSR pipeline [144].

6.7 Result and discussion

Phylogenetic group and multilocus sequence typing (MLST) analysis revealed all the studied isolates belongs to 4 major phylogroup A (n = 54), B1 (n = 16), B2 (n = 15) and D (n = 12) with a total of 26 different STs types. Among all isolates only four ST types were predominant within all compartments, ST410 (n=27), ST131 (n=15), followed by ST224 and ST648 (n= six isolates each). The all the remaining isolates comprised 22 different known STs. ST410 were most abundant cluster in all populations (humans, livestock, companion animals and farm environment) (Figure 6.1).

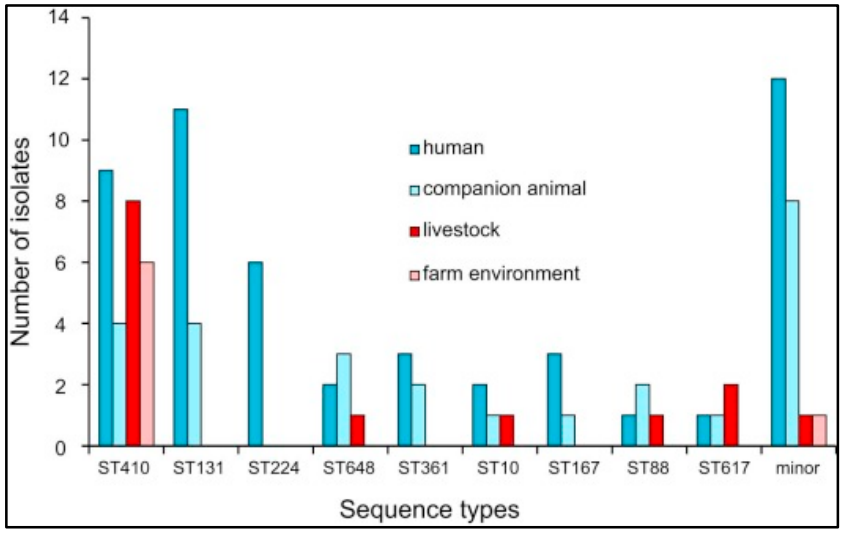


Figure 6.1: Distribution of different STs Types with four different compartments (Source Falgenhauer, 2016).

Core genome phylogenetic analysis of all ST410 isolates revealed that all isolates clustered into five clades designated as A–E. All the isolates share 86% of their core genome. Isolates of Clades A and E were exclusively human and companion animal. Clade D isolates were from a mixed population (humans and livestock) while it is true for clade C also. Clade B isolates were a mixed populations of all four different sources. *In silico* serotyping revealed clade A and B as O8:H21, isolates of clade C as O8:H9 and isolates of clades D and E as O-:H9 serotype. SNP analysis revealed clade B and clade C were tightly bound, the total number of SNPs were separated by between 31 - 70 SNPs. Comparative analysis revealed clade A and C differed from clade B by 121 and 146 SNPs. In a total of SNPs were detected in isolates of clades D and E in between 1528 and 1536 SNPs (Figure 6.2). Three different chromosomal insertion sites were identified with transposition unit "orf477- $bla_{CTX\text{-}M\text{-}15}$ - ISEcp1". In some case, the transposition unit was mediated by prophage. Chromosomal insertion was significantly different then plasmid-mediated transposition. In case of plasmid the role of IS26 mediate antibiotic resiatnce gene transfer were observed. Virulence gene screening revealed several virulence factors such as *lpfA* gene (codes for long polar fimbriae) *prfB* gene (coding for P-related fimbriae), serum survival gene (*iss*) microcins or colicins (*mcmA* or *cma*). Toxin genes (*astA* and *senB*). All isolates harboured the ferrichrome and ferrous iron-uptake operons (*fhuABCD*, *feoABCD*), the iron(III) dicitrate uptake operon (*fecRI-ABCDE*) as well as the enterobactin siderophore operon (*entABCDEFH*, *entS*, *fepABCDEG*, *fes*, *ybdZ*), *sit* operon (*sitABCD*, Fe2+ transport),aerobactin operon (*iucABCD*, *iutA*) salmochelin operon (*iroBCDEN*) yersiniabactin operon (*fyuA*, *irp1*, *irp2*, *ybtAEPQSTUX*) was present in isolates of clades D and E (Figure 6.3).

6.8 Conclusion

Our study revealed the genetic relationship of sampled CTX-M-15-producing *E. coli* isolates from the livestock, companion animals, farm environmental sources and humans from different regions in Germany. Majority of these isolates belongs to four major ST types, and separated by <100 SNPs, which support epidemiological linkage of these isolates. Earlier studies suggested that chromosomal locations of $bla_{CTX\text{-}M\text{-}15}$ are relatively uncommon. Our study showed chromosomal insertion insertions of $bla_{CTX\text{-}M\text{-}15}$ is a common phenomenon, especially in clade B. Plasmid analysis indicated that studied population harboured a commonly occurring FII, FIA and FIB plasmid replicon types with similar sizes. Future studies needed for the presence of ST410

clades which will provide information as to whether this is also the case in other countries worldwide.

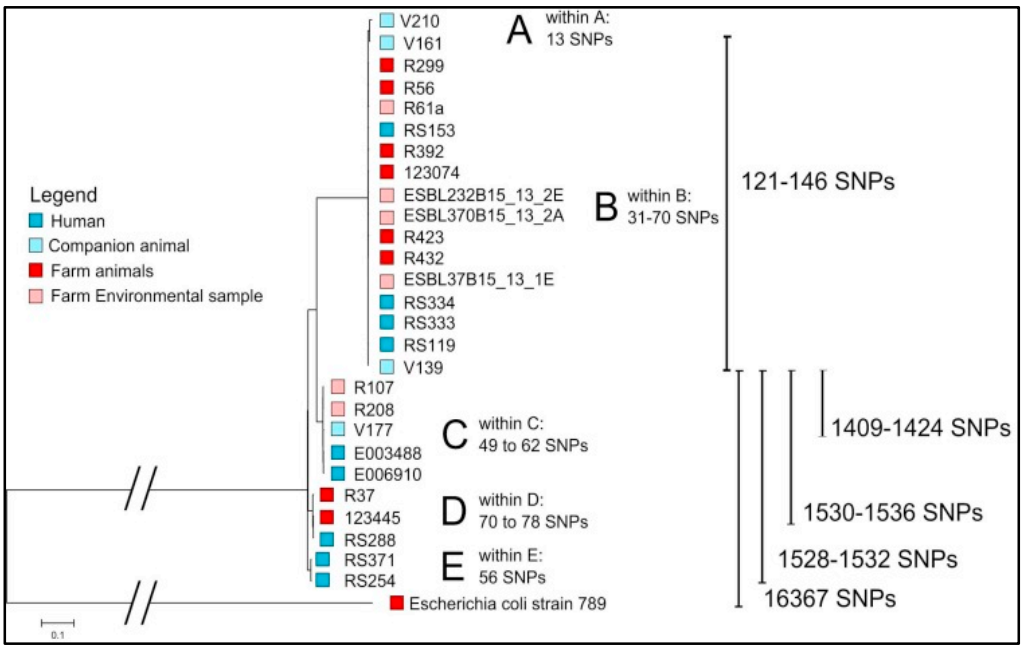


Figure 6.2: Phylogenetic analysis of the *E coli* sequence type 410 (ST410) isolates. The phylogenetic tree was generated by using Parsnp and visualized by Figtree (Source Falgenhauer, 2016).

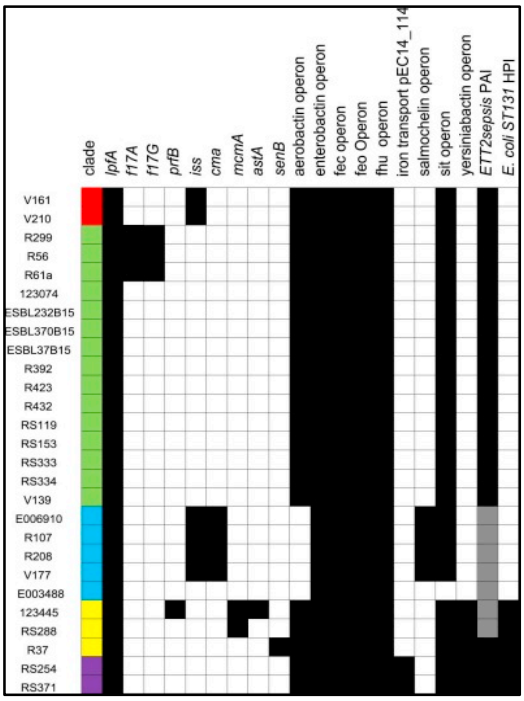


Figure 6.3: Presence of virulence genes heat map. Here black, grey and white indicate fully present, partially present and absent (Source Falgenhauer, 2016).

7. Detection of translocatable units in a blaCTX-M-15 extended-spectrum β-lactamase-producing ST131 E coli isolate using a hybrid sequencing approach

7.1 Publication

Detection of translocatable units in a blaCTX-M-15 extended-spectrum β-lactamase-producing ST131 *E coli* isolate using a hybrid sequencing approach

Ghosh H, Doijad S, Bunk B, Falgenhauer L, Yao Y, Spröer C, Gentil K, Schmiedel J, Imirzalioglu C, Overmann J and Chakraborty T

International Journal of Antimicrobial Agents, 2016-03-01, Volume 47, Issue 3

7.2 Contributions

The author (H.G) was part of designing of the study and drafting the manuscript under the supervision of T.C. He processed bioinformatics analysis steps to preprocessed data and carried out all subsequent analysis. He designed and developed workflows of data analysis of 3rd generation sequencing specially genome assembly and annotation under the supervision of BB and T.C. Furthermore, HG investigate reviewer comment with T.C and SD for the final publication.

7.3 Abstract

Microbes can exchange their genetic material between neighboring microbes by horizontal gene transfer (HGT). IS*26* plays a major role in the acquisition and dissemination of mobile genetic elements in gram-negative bacteria. In this study, we have investigated a novel mechanism of lateral gene transfer in bacteria through “Translocatable units” (TUs). TUs are direct flanking TSDs in IS26 copies and reflect a novel dynamic processes of bacterial genome alteration, involving transfer of antibiotic resistance gene via RecA-independent recombination. With the advent of hybrid sequencing approach, here we confirmed the complete nucleotide sequence and gene transfer mechanism of the two novel TUs. Both the TUs harbor several antibiotic resistance genes. One containing (TU-1) a circular DNA element carrying multiple copies of IS26 and antibiotic resistance genes such as *ant*1, $bla_{\text{CTX-M}}$-15, the major facilitator superfamily (MFS) efflux pump transporters *mphA*, *emrE*, *srpC* and *mdfA*, genes involved in folate biosynthesis (*folA* and *folP*), the phenolic acid stress response regulator *padR*, and cyclic-di-GMP phosphodiesterase *adrB* genes. Another one (TU-

2) additionally carries the antibiotic resistance genes *aacA*4, *catB*3 and *bla*$_{OXA-1}$. This is the first report of sequencing based study to identify not only on the 'static' genome information, but also captures dynamic changes mediated by mobile genetic elements.

7.4. Background and Introduction

Transposon mediated lateral gene transfer is a common phenomena for bacterial survival and adaptation in different niche [153]. It also plays a key role in the dissemination of antibiotic resistance and virulence gene to same bacterial species, as well as to bacteria in another genus or species. Recently, studies have demonstrated that intramolecular replicative transposition generates circular molecules, containing a single copy of the insertion sequence IS*26* with an adjacent DNA segment designated as a TUs [154][47]. TUs can move as a discrete unit inside bacterial cell and integrated at a new chromosomal location via replicative transposition with RecA-independent mechanism. The aim of this work is to analyze composite transposons and molecular characterization of TUs using third generation sequencing approach. As a specific example, we describe one of the most clinically important lineage *E. coli* sequence type (STs) ST131 "translocatable unit" and discuss its possible impact in the dissemination of antibiotic resistance gene. Extended-spectrum beta-lactamase (ESBL)-producing *E. coli* ST 131 (ST131) is a pandemic clonal group of strains and emerging global threat [155]. CTX-M-type beta-lactamases are the most common ESBL-type found in *E. coli* ST131 strains that are frequently associated with high virulence [65]. In particular, the CTX-M-15 enzyme has a worldwide distribution and is often encoded on conjugative plasmids, but also, less frequently, on the chromosome [156].

7.5 Material and Methods

An *E. coli* isolate V282 was obtained from a 12-year-old dog presenting with cystitis from the local veterinary hospital in Giessen, Germany. Preliminary characterization of the isolate revealed the sequence type ST131, harboring the ESBL gene *bla*$_{CTX-M-15}$. The isolate was resistant to ampicillin, trimethoprim/sulfamethoxazole, cefotaxime, chloramphenicol and tetracycline. For Genome sequencing, we used the protocol described in chapter III MiSeq platform (Illumina Netherlands BV, Eindhoven, The Netherlands) yielded 1,201,696 paired reads with a mean read length of 205 bp. Initially the raw reads were assembled by using Spades V.5 (7) after quality-filtering and preliminary analysis of the data indicated the presence of multiple copies of genetic

elements harboring antibiotic resistance genes such as *bla*$_{CTX-M-15}$. However, these genetic elements could neither be assigned to the plasmid nor to the chromosome. To resolve the issue, we performed long read Single Molecule Real Time (SMRT) sequencing (Pacific Biosciences, Menlo Park, USA) from the same DNA isolated from V282. Long-read single molecule real-time (SMRT) sequencing (Pacific Biosciences, Menlo Park, CA) was performed using the DNA isolated from V282. A total of 53,634 reads with a mean read length of 9,610 bp was obtained during PacBio sequencing. PacBio reads were first assembled by RS HGAP Assembly. protocol included in SMRT Portal 2.3.0, after which MiSeq reads were mapped onto the assembled sequence using Burrows–Wheeler Aligner (BWA) in order to obtain a highly accurate genome with QV60 final quality. Both the circular molecule finally annotated by using prokka 1.11 (8) and The genome and TUs sequences of V282 has been deposited in the National Center for Biotechnology Information (NCBI) (GenBank accession nos. KT988018–KT988020).

7.6 Result and discussion

Genome assembly resulted in a single circular contig of 5.11 Mb, identified as the chromosome and an additional circular contig of 118.3 kb designated as plasmid pECOV282. In addition, two discrete circular genetic elements of 23.7 and 16.3 kb were assembled, harboring the *bla*$_{CTX-M-15}$ allele. Hereby, the 23.7 kb element showed a two-fold, the 16.3 kb element even a three-fold increased coverage compared to that of the chromosome pECOV282 (Figure 7.1.) Further investigations revealed that these elements were identical to each other with the exception of a 6 kb insertion in either element. A BLAST search indicated no significant similarity to the chromosome, but ~99% sequence similarity to pECOV282, suggesting these elements originate from the plasmid (Figure 5.1) while absence of *rep*A gene confirms they are not plasmid. High-confidence assembly generated by PacBio followed by MiSeq read error correction eliminated the possibility of these segments being tandemly duplicated on pECOV282. All elements contain identical insertion sequences and can be mapped to a common region on pECOV282 (Figure 7.2.A). In addition, mapping to pECOV282 using MiSeq reads resulted in ~5-fold higher coverage for this region compared to the rest of the plasmid (Figure 7.2.B) supporting their presence as separate entities. The high coverage ensures the existence of these elements in high copy number, which is an attribute of the replicative transposition mechanism described earlier. Both elements

carry antibiotic resistance genes as *ant*1, $bla_{CTX\text{-}M\text{-}15}$, and a multidrug efflux transporter, the 23.7 kb element carries additional antibiotic resistance genes *aacA4*, $bla_{CTX\text{-}M\text{-}15,}$ *cat* and $bla_{OXA\text{-}1}$.

Based on our analysis, we conclude that these two elements represent dynamic regions of frequent transposition events within pECOV282 that are continuously being duplicated and excised. For further confirmation, our analysis was compared with another 30 ST131 strains, which showed the occurrence of such TUs in 12 cases. A careful observation of these clusters revealed these antibiotic resistance genes, and IS elements were occurring in a distinctive gene cassette where antibiotic resistance genes were flanked between IS26 elements (IS-ARC). Various combinations of IS-ARC were observed in the different clusters of the different plasmids, suggesting each IS-ARC as an independent unit. Thus, in-depth sequencing of a bacterial isolate provides information not only on the “static” genome but also captures dynamic changes, as indicated by higher copy numbers, such as duplication and insertion-excision events associated with mobile elements. Our data cautions against interpreting presence of additional copies of antibiotic coding resistance elements as multiple insertions, either singly or in tandem, within the chromosome or plasmid (Figure 7.3).

7.7 Conclusion

The hybrid sequencing approach used here by combining Illumina short reads and PacBio long reads allows the generation of an accurate and reliable finished and closed genome representing a snapshot of genome dynamics within the culture. In many cases, short reads mapped to the reference sequence with coverage values above the genomic average are considered as artifacts and are often neglected. We recommend careful observation of such high coverage regions as they may represent significant dynamic changes due to insertion elements with localized replicative activity.

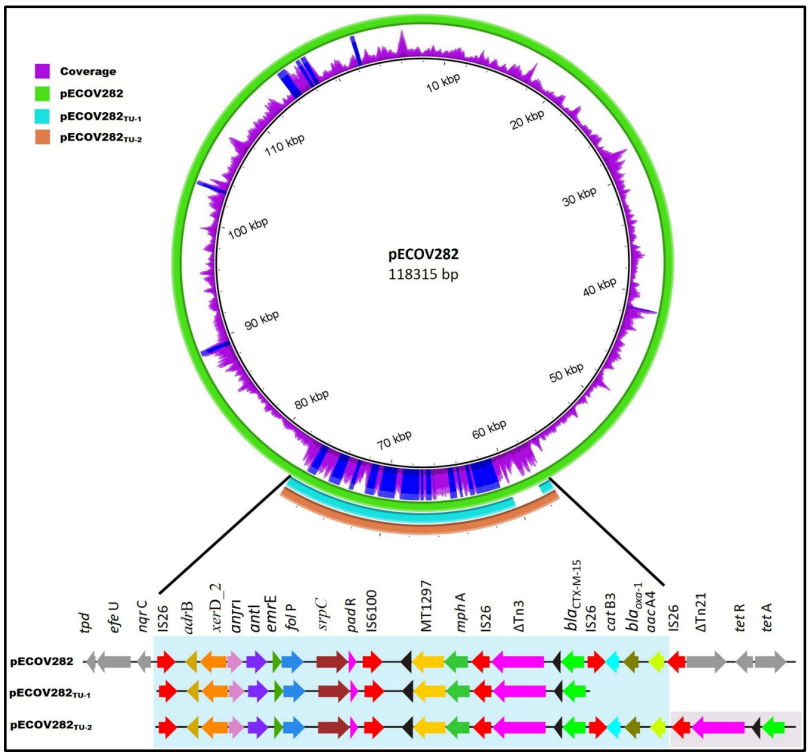


Figure 7.1: The representation of the two genetic elements and homologous regions on the closed plasmid from *E. coli* V282 (pV282). From inside out, ring 1: Read mapping indicating fold coverage (note comparative ~5x high coverage for genetic elements), ring 2: pV282, ring 3 and 4: 16.3 and 23.7 bp genetic element. The outermost circle represents transposon/Insertion sequence (red) and antibiotic resistance gene (blue). The image is generated by 'BLAST ring image generator' (BRIG) program (Source Ghosh, 2016).

Figure7.2.A: Depth of Sequence data coverage: Schematic diagram showing the coverage of illumina read mapping of pV282. Highlighted red region (< 12kb region) indicate the presence of TUs1 in this region.

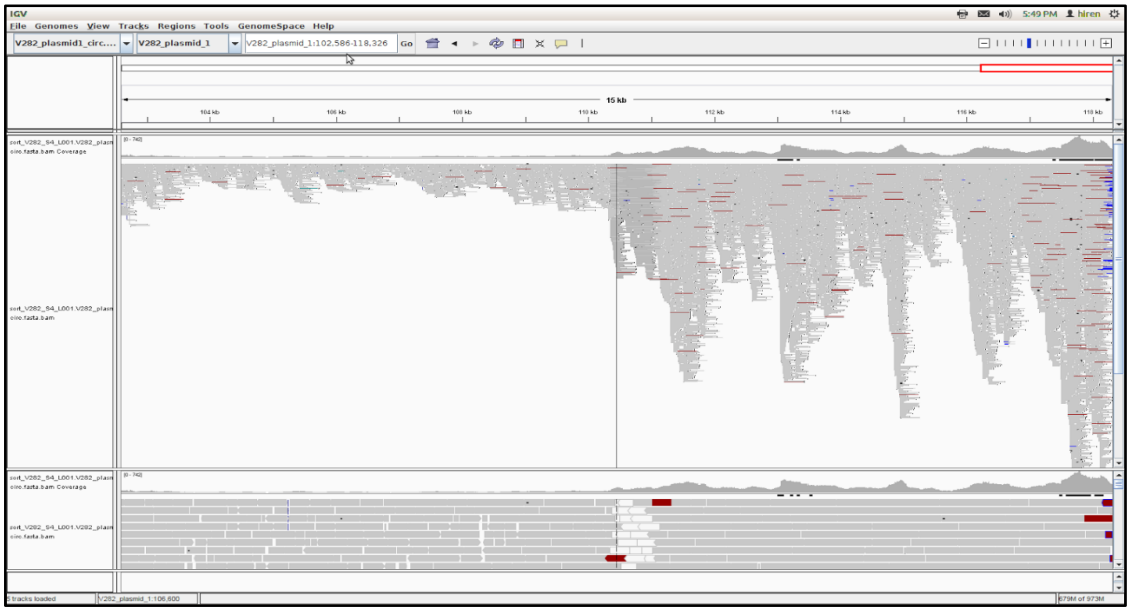


Figure 7.2.B: Depth of sequence data coverage: Schematic diagram showing the coverage of illumina read mapping of pV282. Highlighted red region (> 110kb) indicate the presence of TUs2 in this region.

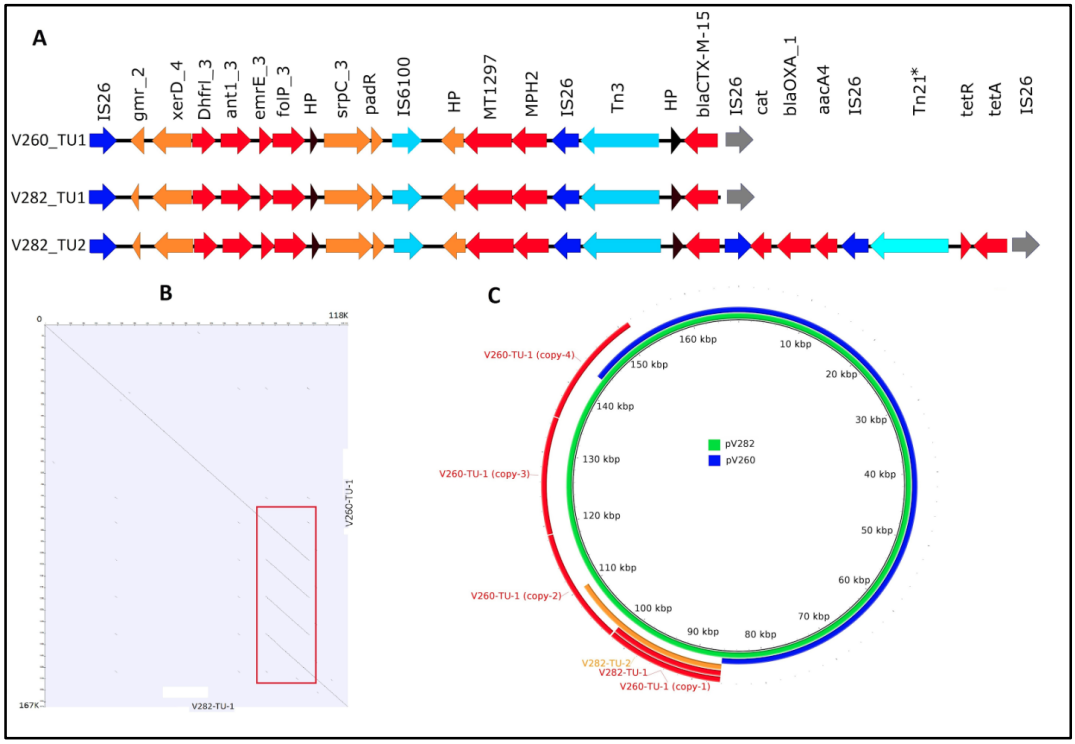


Figure 7.3: Translocatable Units (TU) of the ST131. (A) a typical cassette of TU observed in this study. A total of 3 homologues TUs were observed of which two were exactly identical (V282-TU-1, and V260-TU-1) while V282-TU-2 appended with additional antibiotic resistance genes cassette at one end. (B) An evidence for localized replication of TU (in this case V260-TU-1) was observed represented by dot plot matrix. (C) TUs mapped against the V282 and V260 (red; four copies of V260-TU-1 and a V282-TU-1, organe; V282-TU-2). * represented truncated version.

8. Summary

During recent decades, the emergence of multidrug-resistant bacteria has become a major concern to public healthcare settings worldwide [157–160]. Infections caused by drug-resistant bacteria, limiting the available treatment options with a high rate of mortality, morbidity and substantial economic burden Indeed, drug resistance phenomena occur due to the normal evolutionary process of bacteria, but it accelerates when bacteria gets any external selective pressure such as misuse or overuse of antibiotics [51]. However, in the last decades the problem of antimicrobial-resistance has been increasing due to the growing incidence of infections caused by antibiotic-resistant gram-negative bacteria. Particularly, the prevalence of extended-spectrum β-lactamase (ESBLs) and carbapenemase-producing *E. coli* in clinical settings is increasing at an alarming rate [25,161,162]. These ESBLs producers are resistant to penicillin, cephalosporin, and monobactam antibiotics and are often associated with urinary tract infections and life-threatening sepsis in immunocompromised hosts [133,163,164].

Overall, this thesis demonstrates the genomic characterization and diversity of antimicrobial resistant ESBL-*E. coli* that were isolated during two nationwide antimicrobial surveillance projects in between 2009 and 2016. Considering the advantages of NGS over the conventional sub genomic typing methods, we used whole genome sequencing to monitor the emergence and spread of antimicrobial resistance**.** To the best of our knowledge, this is the first study that considered the whole-genome sequence of ESBL-*E. coli* from humans, animal, environment, and food. The first part of this work demonstrates the molecular epidemiology and ecology of ESBL-*E. coil* isolated from a shared population. Our in-depth genomic and phylogenetic analysis elucidated a highly diverse population of ESBL-*E. coli* is circulating in the different hosts and ecology. In spite of the large genetic diversity, a limited number of clones that has emerged in the diverse background and highly associated with the spread of the antibiotic resistance gene. The predominant CTX-M allele, $bla_{\text{ctx-M-15/1/14}}$ that was mostly associated with the conjugative IncF plasmids replicon types. Previous studies from Germany, the Netherlands and UK have suggested that within ESBL-*E.coli* population, most commonly plasmid replicon types were IncI1-Iγ and followed by IncF [165]. Corresponding with proceeding report, our investigation revealed IncF and IncI replicons were also the most common replicon

types. Notable, IncFII IncFIB, IncFIA were predominant replicon types, which connoting transmission potential. Genome-wide association (GWAS) analysis illustrates that isolates from different ecological compartments share not only core genes, but also showed bidirectional gene flow in their accessory genomes in across different hosts. Moreover, we did not find any clear evidence for the presence of host specific resistance gene or virulence factor. Further, compared to conventional typing approaches, our analysis also highlighted the application of; single nucleotide polymorphism (SNP) based clustering methods could be an alternate approach to study a large scale closely related bacterial population. SNP clustering reveled nine major and five minor clusters in our studied population. Overall, in this part of the thesis, the existence of most common and niche specific, highly adapted antibiotic resistant clone in a diverse ecological niche were identified. E.g., *E. coli* ST131 was exclusively identified in human and companion animal, whereas ST10, ST410 were distributed in all different ecological niches and circulating across human, animal and the environment. Therefore, a significant difference in strain-specific transmission across diverse hosts was observed.

In the second part of the study, we studied the genomic portrait of *E.coli* ST131, which are commonly associated with urinary tract, kidney, bloodstream and other infections [166] [167]. We employed third generation, single-molecule real-time (SMRT) sequencing technology to analyze the genome plasticity of ESBL and non-ESBL *E. coli* ST131 strains (n=30) isolated from human and companion animals. As mentioned above, the major obstacle of the draft genome based study was challenging to understand the role of the contribution of core and accessory genome including the role of plasmids in evolution. Our closed genome analysis revealed that the core components of ST131 are highly conserved in both ESBL and non-ESBL isolates. Recent draft genome based studies indicated that a huge and diverse group of plasmids were associated with the rapid emergence and successful spread of *E. coli* ST131 [145]. Compared to the earlier study a limited plasmid diversity were observed in our studied population. In addition, a conserved IncF plasmid, harboring $bla_{ctx\text{-}M}$ allele along with a number of toxin- antitoxin systems and metabolic gene cassettes were identified. In the recent studies, it has been shown, that the onset success behind ST131 was mainly driven by accessory genome such as prophage, acquisition, and loss or gain of genomic islands as well as integrase-associated genetic segments [167] [168]. Our analysis also corresponding well together with preceding reports. Moreover,

our closed genome analysis provides evidence that the core component between ESBL and non-ESBL were conserved, which suggested a differentiated view of ST131 evolution that represents a highly adapted clone and were ancestral to these habitats. Subsequent acquisition of antibiotic resistance and genomic islands (GIs) together with niche specific bacteriophage uptake promoted the worldwide expansion of several lineages. In continuation with previous studies further, we identified the emergence of a major sub-lineage ST131 C1-M27 clade in Germany, which also observed in Japan and France [146] [147]. Our analysis illustrates that in Germany the incidence of ST131 C1-M27 has been increased from 0% in 2009 to 45% in 2016. Our closed genome data will help to advance the understanding of the genome plasticity and the role of the accessory genome in adaptation and evolution of ST131. Moreover, our high-quality data will serve as a reference genome in comparative genome analysis of *E. coli* ST131.

In the final part of the thesis, the probable mechanisms of the rapid emergence of antibiotic resistance were elucidated. Recently, studies have been reported that the IS26 plays a key role in the dissemination of antibiotic resistance genes via composite transposons [47] [154]. Similarly, in our approaches, the temporal dynamic of the bacterial genome mediated by IS26 was investigated. By using single molecule real time sequencing method a circular intramolecular replicative transposition comprising of only a single copy of IS26 along with an adjacent DNA segment has been identified, which is termed a “Translocatable Units” (TUs) [47,154,169]. TUs are novel genetic elements, capable of generating tandem arrays of antibiotic resistance genes by a mechanism that is distinct from transposition mechanism. Our Study found TUs plays a significant role in the mobilization of several antibiotic resistance genes (e.g. *ant1* and $bla_{\text{CTX-M-15)}}$ along with drug efflux pump (*mphA*, *emrE*, *srpC*) and stress response regulator genes (*padR* and *adrB)*. Our analysis concluded an accurate in-depth sequencing would provide information on not only the ‘static’ genome, but also captures dynamic events mediated by mobile genetic elements. In conclusion, the expansion of these multidrug-resistant clones suggests that the treatment of ESBL-*E. coli* infections will become increasingly difficult in the future.

Overall, the potential of whole-genome sequencing is still undisputed for clinical diagnostics and decision-making. Currently, complete microbiological diagnostics is not yet feasible by WGS due to time and cost reasons and proper standardization and bioinformatics analysis. In addition, other "omics" approaches, such as

transcriptomics, metabolomics, proteomics and sub proteomics, can be used. In the future, in addition to whole-genome analysis in the sense of an integrated, multi-dimensional omics approach would serve better to understand infectious diseases and for improved clinical microbiological diagnostics. Inherently, our study is limited to a few points: The numbers of investigating isolates was relatively small to get a clear epidemiological picture of antimicrobial resistance transmission. All the isolates obtained only from Germany and the collection time was limited. There is a clear need for a global collection of the dataset at least from the Eurozone to understand the population fluctuation in dominant lineages. In order to understand the clear picture of dissemination of antibiotic resistance, it is necessary to get a detailed plasmidome characterization. Finally, the draft genome analysis could always challenge due to chance, to lose functionally and ecologically important genetic material (e.g. genomic island), which could introduce bias into the subsequent comparative genome analysis.

9. Zusammenfassung

Das weltweite Auftreten multi-resistenter Bakterien in den öffentlichen Gesundheitseinrichtungen ist in den letzten Jahrzehnten ein wichtiges Thema geworden [155-158]. Für multi-resistente Bakterien-Infektionen gibt es Einschränkungen in der Auswahl von verfügbaren Behandlungsmöglichkeiten. Sie sind daher mit einer hohen Mortalitätsrate und Morbidität verbunden und verursachen beträchtliche wirtschaftliche Verluste. Aufgrund evolutionärer Prozesse in Bakterien können Arzneimittelresistenzphänomene auftreten; diese beschleunigen und verstärken sich aber, wenn ein Mikroorganismus unter externen Selektionsdruck gerät wie z.B. die unsachgemäße oder übermäßige Anwendung von Antibiotika [50]. Die zunehmende Häufigkeit der durch Antibiotika-resistente Gram-negative Bakterien verursachten Infektionen in den letzten Jahrzehnten verschärfte das Problem der antimikrobiellen Resistenz enorm. Insbesondere ist es alarmierend, dass die Prävalenz von β-Lactamasen mit erweitertem Spektrum (ESBLs) und Carbapenemase-produzierenden *Enterobacteriaceae* in Krankhäusern dramatisch angestiegen ist [25, 159, 160]. Diese ESBL-Produzenten sind resistent gegen Penicillin-, Cephalosporin- und Monobactam-Antibiotika und stehen häufig in Zusammenhang mit Harnwegsinfektionen und lebensbedrohlicher Sepsis bei Immunsupprimierten [131, 161, 162].

Die vorliegende Arbeit postuliert die genomische Charakterisierung und genetische Diversität von antimikrobiellen resistenten, ESBL produzierenden *E. coli* Isolaten, gewonnen aus zwei landesweite Überwachungsprojekten während der Zeit zwischen 2009 und 2016. Unter Berücksichtigung der Vorteile von next generation sequencing (NGS) gegenüber den herkömmlichen Typisierungsverfahren verwendete ich NGS für das Monitoring des Auftretens und der Ausbreitung der Antibiotika-Resistenz. Der erste Teil der Arbeit untersuchte die molekulare Epidemiologie und Ökologie von ESBL-kodierenden *E. coli* aus einer gemeinsamen Population. Die tiefergehende Genomanalyse und phylogenetische Untersuchung entdeckte eine sehr unterschiedliche Population von ESBL-positiven *E. coli*, die zwischen den verschiedenen Wirten und Ökosystemen zirkulierten. Trotz der großen genetischen Vielfalt gab es nur eine begrenzte Anzahl von Klonen, die in den verschiedenen Hintergrund entstanden und stark mit der Verbreitung des Antibiotikaresistenzgens verbunden sind. Die vorherrschenden ESBL-Allele waren $bla_{\text{CTX-M-15/1/14}}$ und

hauptsächlich mit konjugativen IncF-Plasmiden assoziiert. Frühere Studien aus Deutschland, den Niederlanden und dem Vereinigten Königreich deuten darauf hin, dass IncI1-Plasmide die am häufigsten vorkommenden Replikontypen waren, gefolgt von IncF [163]. Die vorliegenden Untersuchungen ergaben aber, dass IncF-Replikons, insbesondere IncFII, IncFIB und IncFIA die häufigsten Replikon-Typen waren, die gleichzeitig auf ein Übertragungspotential hindeuten. Die GWAS-Analyse zeigte, dass Isolate aus verschiedenen ökologischen Kompartimenten nicht nur ein gemeinsames Core-Genom (konservative Kerngene), sondern auch einen bidirektionalen Fluss von Genen in den akzessorischen Genomen der verschiedenen Wirtsspezies haben. Keinerlei Hinweise auf Wirts-spezifische Resistenz-Gene oder Virulenz-Faktoren wurden gefunden. Im Vergleich zu herkömmlichen Typisierungsverfahren wurde die Single-Nucleotide-Polymorphismus- (SNP)-Methode für die Analyse verwendet. SNP-Clustering könnte ein alternativer Weg sein, um eine umfangreiche Bakterien-Population zu gruppieren.

Die Ergebnisse nach dem SNP-Clustering zeigten neun große und fünf kleinere Cluster in der untersuchten Population. Die häufigsten und nischenspezifisch-adaptierten Antibiotika-resistenten Klone in vielfältigen ökologischen Nischen konnten identifiziert werden. Ich fand zum Beispiel, dass *E. coli* des Sequenztyps ST131 hauptsächlich in Menschen und Haustieren, und *E. coli* des ST410 in verschiedenen Nischen (in Mensch, Tier und Umwelt) zirkulieren. Daher wurde ein signifikanter Unterschied zwischen den stammspezifischen Übertragungen über verschiedene Wirte beobachtet.

Im zweiten Teil der vorliegenden Arbeit wurde *E. coli* des ST131, die häufig Infektionen in Harnwege, Nieren und Blutbahnen verursachtet, genauer untersucht. Die Einzelmolekül-Echtzeit-Sequenzierungstechnologie (SMRT), eine Dritt-Generationsmethode der Genomsequenzierung wurde zur Sequenzierung von 30 ST131 *E. coli* Isolate aus Menschen und Haustieren eingesetzt, um die Genomplastizität der ESBL- und Nicht-ESBL-Stämmen zu untersuchen. Analysen basiert auf Draft-Genom-Sequenz können nur sehr schwer erklären, welche Rolle das Kern- und akzessorische Genom einschließlich Plasmiden in der bakteriellen Evolution spielen. Mit geschlossenen Genomen können dagegen vielseitige Genomvergleiche durchgeführt werden. Es wurde festgestellt, dass die Kernkomponenten von ST131 sowohl in ESBL- als auch in Nicht-ESBL-Isolaten hochkonserviert waren. Nur wenige neuere genombasierte Studien wiesen darauf hin, dass eine große und vielfältige

Gruppe von Plasmiden mit dem schnellen Auftreten und der erfolgreichen Ausbreitung von *E. coli* ST131 assoziiert war [143]. Im Vergleich zu den früheren Studien fand ich eine begrenzte Plasmid-Diversität in den untersuchten Isolaten. Darüber hinaus wurde ein konserviertes IncF-Plasmid mit einem bla_{CTX-M}-Allel sowie einer Anzahl von Toxin-Antitoxin-Systemen und metabolischen Genkassetten identifiziert. Die jüngsten Studien haben gezeigt, dass sich hinter dem ST131 Erfolg hauptsächlich das akzessorische Genom wie Prophage, die Akquisition und der Verlust von genomischen Inseln sowie Integrase-assoziierten genetischen Segmenten verbergen. Dies wurde auch in dieser Studie beobachtet. Die Ergebnisse zeigten, dass die Kernkomponente zwischen ESBL und Nicht-ESBL konserviert war, was auf eine differenzierte Evolutionssichtweise der ST131 Stämme hindeutet, die einen hochadaptierten Klon repräsentieren und als Vorfahren in diesen Lebensräumen vorkamen. Der anschließende Erwerb von Antibiotika-Resistenz-Genen und genomischen Inseln (GIs), zusammen mit einer Aufnahme nischenspezifischer Bakteriophagen bildete mehrere Linien und förderte die weltweite Ausbreitung dieser Linien. In einem Unterkapitel wurde die Entstehung einer großen Sub-Abstammungslinie ST131 C1-M27 in Deutschland behandelt. Diese Sublinie wurde in Japan und Frankreich berichtet. Die hier durchgeführte Analyse zeigt einen Anstieg dieser Abstammungslinie von 0% in 2009 bis 45% in 2016.

Die Sequenzen geschlossener Genome ermöglichten uns, die Genome-Plastizität von ST131 Stämmen und die Rolle des accssory Genoms in der Adaption und Evolution zu verstehen. Die erzeugten hochqualitativen Daten können als Referenzgenome für vergleichende Genomanalyse von ST131 *E. coli* dienen.

Der letzte Teil dieser Arbeit sollte einen wahrscheinlichen Mechanismus des schnellen Auftretens von Antibiotikaresistenz aufklären. Neuerlich berichten einige Untersuchungen, dass IS26 einen entscheidende Rolle in der durch zusammengesetzten Transposons vermittelten Antibiotikaresistenzausbreitung spielt (47, 152). Ich untersuchte mit den oben genannten Methoden die durch IS26 vermittelte temporäre Dynamik in bakteriellen Genomen. Unter Verwendung von SMRT Sequenzierung wurde ein zirkuläres intramolekulares replikatives Transposon identifiziert, das eine einzige IS26 Kopie zusammen mit einem benachbarten DNA-Segment umfasst und als "Translocatable Units" (TUs) [47, 145, 164] bezeichnet ist. TUs sind neue genetische Elemente, die in der Lage sind, durch einen Mechanismus anders als die Transposition, Tandem-Arrays von Antibiotikaresistenzgenen zu

erzeugen. Die Analysen zeigten, dass TUs eine bedeutende Rolle in Mobilisierung von Antibiotikaresistenzgenen (z.B. *ant1* und $bla_{CTX\text{-}M\text{-}15}$) zusammen mit Antibiotika-Pumpen (*mphA*, *ermE*, *srpC*) und Stress-Response Regulator-Genen (*pad*R und *adr*B) haben [47]. Ich habe gezeigt, dass eine genauere Tiefsequenzierung nicht nur Informationen über das "statische" Genom liefert, sondern auch dynamische Ereignisse erfasst, die durch mobile genetische Elemente vermittelt werden. Zusammenfassend lässt die Expansion dieser multiresistenten Klone vermuten, dass die Behandlung von ESBL - *E. coli* - Infektionen in Zukunft immer schwieriger wird.
Insgesamt ist das Anwendungspotenzial der Gesamtgenomsequenzierung für die klinische Diagnostik und Entscheidungsfindung unbestritten. Allerdings ist zurzeit eine vollständige mikrobiologische Diagnostik aus Kosten- und Zeitgründen sowie durch eine entsprechende Standardisierung und bioinformatische Analyse noch nicht möglich. Darüber hinaus können andere "Omics" -Ansätze, wie Transkriptomik, Metabolomik, Proteomik und Subproteomik, verwendet werden. In Zukunft soll neben der Ganzgenomanalyse im Sinne eines integrierten, multidimensionalen Omics-Ansatzes ein besseres Verständnis von Infektionskrankheiten und eine verbesserte klinische mikrobiologische Diagnostik ermöglicht werden. Naturgemäß ist die vorgelegte Studie mit Einschränkung in folgenden Punkten behaftet: Die Anzahl der untersuchten Isolate war relativ gering, um ein klares epidemiologisches Bild der antimikrobiellen Übertragung zu erhalten. Alle Isolate stammten ausschließlich aus Deutschland und in einer kurzen Periode. Es besteht ein eindeutiger Bedarf für eine globale Sammlung des Datensatzes, zumindest aus der Eurozone, um die Populationsfluktuation der wichtigen Linien zu untersuchen. Darüber hinaus ist detaillierte Charakterisierung der Plasmidome notwendig, d.h. der Plasmidinhalte der Bakterien, um ein klares Bild für die Verbreitung von Antibiotika-Resistenzgenen zu verschaffen. Die Draft-Genome Analyse könnte manchmal unausreichend sein, aufgrund ihrer Unvollständigkeit, z.B. Verluste von funktionell und ökologisch wichtigen genetischen Sequenzen, und diese in dem anschließenden Genomvergleich zu Abweichung führen kann.

10. Abbreviations

AFLP	Amplified fragment length polymorphism
ABR	Antibacterial resistance
AMR	Antimicrobial resistance
AR	Antibiotic resistance
ARG	Antibiotic resistance gene
BLAST	Basic Local Alignment Search Tool
CARD	Comprehensive Antibiotic Resistance Database
CCD	Centers for Disease Control and Prevention
CCD	Charge-coupled device
EDR	Extensive drug resistance
ERIC	Enterobacterial repetitive intergenic consensus
ESBLs	Extended-Spectrum Beta- Lactamases
ESTs	Expressed sequence tags
EUCAST	European Committee on Antimicrobial Susceptibility Testing
FQ-R	Fluoroquinolone-resistant
GI	Genomic island
GWAS	Genome-wide association studies
HGAP	Hierarchical Genome Assembly Process
HGT	Horizontal gene transfer
ICEs	Integrative and Conjugative Elements
ICU	Intensive care unit
IS	Insertion sequences
MALDI	Matrix Assisted Laser Desorption/Ionization
MDR	Multi-drug-resistant
MDRGN	Multi-drug resistant gram-negative bacteria
MGE	Mobile genetic element
MGEs	Mobile genetic elements
MLEE	Multilocus enzyme electrophoresis
MLEE:	Multilocus enzyme electrophoresis

MLST	Multilocus sequence typing
MLVA	Multilocus variable number of tandem repeats analysis
NGS	Next-generation sequencing
OLC	Overlap layout consensus
ORFs	Open reading frames
PCR	Polymerase chain reaction
PFGE	Pulse field gel electrophoresis
RAPD	Random Amplified Polymorphic DNA
RMs	Restriction modification systems
rRNA	ribosomal RNA
SBS	Sequencing by synthesis
SGS	Second generation sequencing
SMRT	Single Molecule Real Time
SNPs	Single nucleotide polymorphism
TEs	Transposable elements
TGS	Third generation sequencing
TUs	Translocatable units
UTIs	Urinary tract infections
VFDB	Virulence factor database
VNTR	Variable number tandem repeat
VRs	Variable regions
WGS	Whole genome sequencing
WHO	World Health Organization
ZMW	Zero-mode waveguides

11. References

[1] Ligon BL. Penicillin: its discovery and early development. Semin Pediatr Infect Dis 2004;15:52–7. doi:10.1053/j.spid.2004.02.001.

[2] Lobanovska M, Pilla G. Penicillin's Discovery and Antibiotic Resistance: Lessons for the Future? Yale J Biol Med 2017;90:135–45.

[3] Davies J, Davies D. Origins and evolution of antibiotic resistance. Microbiol Mol Biol Rev 2010;74:417–33. doi:10.1128/MMBR.00016-10.

[4] Laxminarayan R, Duse A, Wattal C, Zaidi AKM, Wertheim HFL, Sumpradit N, et al. Antibiotic resistance—the need for global solutions. Lancet Infect Dis 2013;13:1057–98. doi:10.1016/S1473-3099(13)70318-9.

[5] Laxminarayan R, Heymann DL. Challenges of drug resistance in the developing world. Bmj 2012;344:e1567–e1567. doi:10.1136/bmj.e1567.

[6] Nessar R, Reyrat JM, Murray A, Gicquel B. Genetic analysis of new 16S rRNA mutations conferring aminoglycoside resistance in Mycobacterium abscessus. J Antimicrob Chemother 2011;66:1719–24. doi:10.1093/jac/dkr209.

[7] Gruger T, Nitiss JL, Maxwell A, Zechiedrich EL, Heisig P, Seeber S, et al. A mutation in Escherichia coli DNA gyrase conferring quinolone resistance results in sensitivity to drugs targeting eukaryotic topoisomerase II. Antimicrob Agents Chemother 2004;48:4495–504. doi:10.1128/AAC.48.12.4495-4504.2004.

[8] Fàbrega A, Madurga S, Giralt E, Vila J. Mechanism of action of and resistance to quinolones. Microb Biotechnol 2009;2:40–61. doi:10.1111/j.1751-7915.2008.00063.x.

[9] Kong K-F, Schneper L, Mathee K. Beta-lactam antibiotics: from antibiosis to resistance and bacteriology. APMIS 2010;118:1–36. doi:10.1111/j.1600-0463.2009.02563.x.

[10] AMIN N EL, GISKE CG, JALAL S, KEIJSER B, KRONVALL G, WRETLIND B. Carbapenem resistance mechanisms in Pseudomonas aeruginosa: alterations of porin OprD and efflux proteins do not fully explain resistance patterns observed in clinical isolates. APMIS 2005;113:187–96. doi:10.1111/j.1600-0463.2005.apm1130306.x.

[11] Ruppé É, Woerther P-L, Barbier F. Mechanisms of antimicrobial resistance in Gram-negative bacilli. Ann Intensive Care 2015;5:61. doi:10.1186/s13613-015-0061-0.

[12] Ogwang S, Nguyen HT, Sherman M, Bajaksouzian S, Jacobs MR, Boom WH, et al. Bacterial conversion of folinic acid is required for antifolate resistance. J Biol Chem 2011;286:15377–90. doi:10.1074/jbc.M111.231076.

[13] Blair JMA, Webber MA, Baylay AJ, Ogbolu DO, Piddock LJ V. Molecular mechanisms of antibiotic resistance. Nat Rev Microbiol 2015;13:42–51. doi:10.1038/nrmicro3380.

[14] Lowe J. Mechanisms of Antibiotic Resistance. Annu Rep Med Chem 1982;17:119–27. doi:10.1016/S0065-7743(08)60495-9.

[15] Von Wintersdorff CJH, Penders J, Van Niekerk JM, Mills ND, Majumder S, Van Alphen LB, et al. Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer. Front Microbiol 2016;7:1–10. doi:10.3389/fmicb.2016.00173.

[16] Lederberg J, Tatum EL. Gene recombination in Escherichia coli. Nature 1946;158:558. doi:10.1038/158558a0.

[17] Griffith F. Classics in infectious diseases: The significance of pneumococcal types. Rev Infect Dis 1981;3:372–95. doi:10.1093/clinids/3.2.372.

[18] Lorenz MG, Wackernagel W. Bacterial gene transfer by natural genetic transformation in the environment. Microbiol Rev 1994;58:563–602.

[19] Jeon B, Muraoka W, Sahin O, Zhang Q. Role of Cj1211 in natural transformation and transfer of antibiotic resistance determinants in Campylobacter jejuni. Antimicrob Agents Chemother 2008;52:2699–708. doi:10.1128/AAC.01607-07.

[20] ZINDER ND, LEDERBERG J. Genetic exchange in Salmonella. J Bacteriol 1952;64:679–99.

[21] Global risks 2014. Global risks 2014: Insight report. 2014.

[22] Neill JO'. Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations The Review on Antimicrobial Resistance Chaired 2014.

[23] WHO. Antimicrobial resistance. Bull World Health Organ 2014;61:383–94. doi:10.1007/s13312-014-0374-3.

[24] Cantón R, Novais A, Valverde A, Machado E, Peixe L, Baquero F, et al. Prevalence and spread of extended-spectrum β-lactamase-producing Enterobacteriaceae in Europe. Clin Microbiol Infect 2008;14:144–53. doi:10.1111/j.1469-0691.2007.01850.x.

[25] Pitout JDD, Nordmann P, Laupland KB, Poirel L. Emergence of Enterobacteriaceae producing extended-spectrum β-lactamases (ESBLs) in the community. J Antimicrob Chemother 2005;56:52–9. doi:10.1093/jac/dki166.

[26] Baroud M, Dandache I, Araj GF, Wakim R, Kanj S, Kanafani Z, et al. Underlying mechanisms of carbapenem resistance in extended-spectrum β-lactamase-producing Klebsiella pneumoniae and Escherichia coli isolates at a tertiary care centre in Lebanon: role of OXA-48 and NDM-1 carbapenemases. Int J Antimicrob Agents 2013;41:75–9. doi:10.1016/j.ijantimicag.2012.08.010.

[27] Cantón R, Horcajada JP, Oliver A, Garbajosa PR, Vila J. Inappropriate use of antibiotics in hospitals: The complex relationship between antibiotic use and antimicrobial resistance. Enferm Infecc Microbiol Clin 2013;31:3–11. doi:10.1016/S0213-005X(13)70126-5.

[28] Baquero F, Alvarez-Ortega C, Martinez JL. Ecology and evolution of antibiotic resistance. Environ Microbiol Rep 2009;1:469–76. doi:10.1111/j.1758-2229.2009.00053.x.

[29] Guinane CM, Cotter PD. Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ. Therap Adv Gastroenterol 2013;6:295–308. doi:10.1177/1756283X13482996.

[30] Becattini S, Taur Y, Pamer EG. Antibiotic-Induced Changes in the Intestinal Microbiota and Disease. Trends Mol Med 2016;22:458–78. doi:10.1016/j.molmed.2016.04.003.

[31] Fair RJ, Tor Y. Antibiotics and bacterial resistance in the 21st century. Perspect Medicin Chem 2014;6:25–64. doi:10.4137/PMC.S14459.

[32] Santajit S, Indrawattana N. Mechanisms of Antimicrobial Resistance in ESKAPE Pathogens. Biomed Res Int 2016;2016:2475067. doi:10.1155/2016/2475067.

[33] Pan S-C, Tien K-L, Hung I-C, Lin Y-J, Sheng W-H, Wang M-J, et al. Compliance of Health Care Workers with Hand Hygiene Practices: Independent Advantages of Overt and Covert Observers. PLoS One 2013;8:e53746. doi:10.1371/journal.pone.0053746.

[34] Bonten MJ. Colonization pressure: a critical parameter in the epidemiology of antibiotic-resistant bacteria. Crit Care 2012 164 2012;16:142. doi:10.1186/cc11417.

[35] Manaia CM. Assessing the Risk of Antibiotic Resistance Transmission from the Environment to Humans: Non-Direct Proportionality between Abundance and Risk. Trends Microbiol 2017;25:173–81. doi:10.1016/j.tim.2016.11.014.

[36] Berendonk TU, Manaia CM, Merlin C, Fatta-Kassinos D, Cytryn E, Walsh F, et al. Tackling antibiotic resistance: the environmental framework. Nat Rev Microbiol 2015;13:310–7. doi:10.1038/nrmicro3439.

[37] Rizzo L, Manaia C, Merlin C, Schwartz T, Dagot C, Ploy MC, et al. Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: A review. Sci Total Environ 2013;447:345–60. doi:10.1016/J.SCITOTENV.2013.01.032.

[38] Clemente JC, Pehrsson EC, Blaser MJ, Sandhu K, Gao Z, Wang B, et al. The microbiome of uncontacted Amerindians. Sci Adv 2015;1:e1500183–e1500183. doi:10.1126/sciadv.1500183.

[39] Santiago-Rodriguez TM, Fornaciari G, Luciani S, Dowd SE, Toranzos GA, Marota I, et al. Gut Microbiome of an 11th Century A.D. Pre-Columbian Andean Mummy. PLoS One 2015;10:e0138135. doi:10.1371/journal.pone.0138135.

[40] Bennett PM. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. Br J Pharmacol 2008;153 Suppl 1:S347-57. doi:10.1038/sj.bjp.0707607.

[41] Orlek A, Stoesser N, Anjum MF, Doumith M, Ellington MJ, Peto T, et al. Plasmid Classification in an Era of Whole-Genome Sequencing: Application in Studies of

Antibiotic Resistance Epidemiology. Front Microbiol 2017;8:182. doi:10.3389/fmicb.2017.00182.

[42] McClintock B. The origin and behavior of mutable loci in maize. Proc Natl Acad Sci 1950;36:344–55. doi:10.1073/pnas.36.6.344.

[43] Berg DE, Berg CM. The prokaryotic transposable element Tn5. Bio/Technology 1983;1:417–35. doi:10.1038/nbt0783-417.

[44] Martínez-Rubio R, Quiles-Puchalt N, Martí M, Humphrey S, Ram G, Smyth D, et al. Phage-inducible islands in the Gram-positive cocci. ISME J 2017;11:1029–42. doi:10.1038/ismej.2016.163.

[45] Sentchilo V, Mayer AP, Guy L, Miyazaki R, Tringe SG, Barry K, et al. Community-wide plasmid gene mobilization and selection. ISME J 2013;7:1173–86. doi:10.1038/ismej.2013.13.

[46] Johnson CM, Grossman AD. Integrative and Conjugative Elements (ICEs): What They Do and How They Work. Annu Rev Genet 2015;49:577–601. doi:10.1146/annurev-genet-112414-055018.

[47] Harmer CJ, Moran RA, Hall RM. Movement of IS26-associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS26 and preferentially inserts adjacent to another IS26. MBio 2014;5:e01801-14. doi:10.1128/mBio.01801-14.

[48] Tansirichaiya S, Mullany P, Roberts AP. PCR-based detection of composite transposons and translocatable units from oral metagenomic DNA. FEMS Microbiol Lett 2016;363:fnw195. doi:10.1093/femsle/fnw195.

[49] Simpkin VL, Renwick MJ, Kelly R, Mossialos E. Incentivising innovation in antibiotic drug discovery and development: progress, challenges and next steps. J Antibiot (Tokyo) 2017. doi:10.1038/ja.2017.124.

[50] Laxminarayan R. Antibiotic effectiveness: Balancing conservation against innovation. Science (80-) 2014;345:1299–301. doi:10.1126/science.1254163.

[51] Ventola CL. The antibiotic resistance crisis: part 1: causes and threats. P T A Peer-Reviewed J Formul Manag 2015;40:277–83. doi:Article.

[52] Buckland D. Antimicrobial resistance and the race to find new antibiotics. Prescriber 2017;28:12–5. doi:10.1002/psb.1528.

[53] M.C. Chung E, Dean SN, Propst CN, Bishop BM, van Hoek ML. Komodo dragon-inspired synthetic peptide DRGN-1 promotes wound-healing of a mixed-biofilm infected wound. Npj Biofilms Microbiomes 2017;3:9. doi:10.1038/s41522-017-0017-2.

[54] Samuels RI, Mattoso TC, Moreira DDO. Chemical warfare: Leaf-cutting ants defend themselves and their gardens against parasite attack by deploying antibiotic secreting bacteria. Commun Integr Biol 2013;6:1–7. doi:10.4161/cib.23095.

[55] Chellat MF, Riedl R. Pseudouridimycin: The First Nucleoside Analogue That Selectively Inhibits Bacterial RNA Polymerase. Angew Chemie Int Ed 2017;56:13184–6. doi:10.1002/anie.201708133.

[56] Zipperer A, Konnerth MC, Laux C, Berscheid A, Janek D, Weidenmaier C, et al. Human commensals producing a novel antibiotic impair pathogen colonization. Nature 2016;535:511–6. doi:10.1038/nature18634.

[57] Rawat D, Nair D. Extended-spectrum β-lactamases in Gram Negative Bacteria. J Glob Infect Dis 2010;2:263–74. doi:10.4103/0974-777X.68531.

[58] Gniadkowski M. Evolution and epidemiology of extended-spectrum β-lactamases (ESBLs) and ESBL-producing microorganisms. Clin Microbiol Infect 2001;7:597–608. doi:10.1046/j.1198-743x.2001.00330.x.

[59] Drawz SM, Bonomo RA. Three decades of ??-lactamase inhibitors. Clin Microbiol Rev 2010;23:160–201. doi:10.1128/CMR.00037-09.

[60] Shaikh S, Fatima J, Shakil S, Rizvi SMD, Kamal MA. Antibiotic resistance and extended spectrum beta-lactamases: Types, epidemiology and treatment. Saudi J Biol Sci 2015;22:90–101. doi:10.1016/j.sjbs.2014.08.002.

[61] Bradford P. Extended spectrum betalactamase in the 21 century: characterization, epidemiology, and detection of this important resistant threat. Clin Microbiol Rev 2001;14:933–51. doi:10.1128/CMR.14.4.933.

[62] Paterson DL, Bonomo RA. Extended-Spectrum beta-Lactamases : a Clinical Update. Clin Microbiol Rev 2005;18:657–86. doi:10.1128/CMR.18.4.657.

[63] Wellington EMH, Boxall ABA, Cross P, Feil EJ, Gaze WH, Hawkey PM, et al. The role of the natural environment in the emergence of antibiotic resistance in Gram-negative bacteria. Lancet Infect Dis 2013;13:155–65. doi:10.1016/S1473-3099(12)70317-1.

[64] Carattoli A. Resistance plasmid families in Enterobacteriaceae. Antimicrob Agents Chemother 2009;53:2227–38. doi:10.1128/AAC.01707-08.

[65] Cantón R, González-Alba JM, Galán JC. CTX-M enzymes: Origin and diffusion. Front Microbiol 2012;3. doi:10.3389/fmicb.2012.00110.

[66] Melin P, Boreux R, Mol P De, Studies MP, Sciences B. Prevalence and spread of extended- spectrum b -lactamase-producing Enterobacteriaceae in Ngaoundere , 2013.

[67] Seiler C, Berendonk TU. Heavy metal driven co-selection of antibiotic resistance in soil and water bodies impacted by agriculture and aquaculture. Front Microbiol 2012;3:1–10. doi:10.3389/fmicb.2012.00399.

[68] Ranjbar R, Karami A, Farshad S, Giammanco GM, Mammina C. Typing methods used in the molecular epidemiology of microbial pathogens: a how-to guide. New Microbiol 2014;37:1–15.

[69] Pérez-Losada M, Arenas M, Castro-Nallar E. Microbial sequence typing in the genomic era. Infect Genet Evol 2017:0–1. doi:10.1016/j.meegid.2017.09.022.

[70] Schwartz DC, Cantor CR. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell 1984;37:67–75. doi:10.1016/0092-8674(84)90301-5.

[71] Larsen M V., Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, et al. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol 2012;50:1355–61. doi:10.1128/JCM.06094-11.

[72] Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A 1998;95:3140–5. doi:10.1073/pnas.95.6.3140.

[73] Tien YY, Ushijima H, Mizuguchi M, Liang SY, Chiou CS. Use of multilocus variable-number tandem repeat analysis in molecular subtyping of Salmonella enterica serovar Typhi isolates. J Med Microbiol 2012;61:223–32. doi:10.1099/jmm.0.037291-0.

[74] Bouchet V, Huot H, Goldstein R. Molecular genetic basis of ribotyping. Clin Microbiol Rev 2008;21:262–73. doi:10.1128/CMR.00026-07.

[75] Pcr R, Healy M, Huong J, Bittner T, Lising M, Raza S, et al. Microbial DNA Typing by Automated Microbial DNA Typing by Automated 2005;43:199–207. doi:10.1128/JCM.43.1.199.

[76] Bumgarner R. DNA microarrays: Types, Applications and their future. Curr Protoc Mol Biol 2013;6137:1–17. doi:10.1002/0471142727.mb2201s101.DNA.

[77] Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 2016;17:333–51. doi:10.1038/nrg.2016.49.

[78] Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. J Appl Genet 2011;52:413–35. doi:10.1007/s13353-011-0057-x.

[79] Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 1975;94:441–8. doi:10.1016/0022-2836(75)90213-2.

[80] Sanger F, Nicklen S, Coulson a R. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 1977;74:5463–7. doi:10.1073/pnas.74.12.5463.

[81] Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, et al. The complete genome of an individual by massively parallel DNA sequencing. Nature 2008;452:872–6. doi:10.1038/nature06884.

[82] Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome Sequencing in Open Microfabricated High Density Picoliter Reactors. Nat Biotechnol 2006;437:376–80. doi:10.1038/nature03959.Copyright.

[83] Shao K, Ding W, Wang F, Li H, Ma D, Wang H. Emulsion PCR: A high efficient way of PCR amplification of random DNA libraries in aptamer selection. PLoS One 2011;6:1–7. doi:10.1371/journal.pone.0024910.

[84] Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of Next-Generation Sequencing Systems. J Biomed Biotechnol 2012;2012:1–11. doi:10.1155/2012/251364.

[85] Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008;456:53–9. doi:10.1038/nature07517.

[86] Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-Time DNA Sequencing from Single Polymerase Molecules. Science (80-) 2009;323:133–8. doi:10.1126/science.1162986.

[87] Levene M, Korlach J, Turner S, Foquet M, Craighead H, Webb WW. Zero-mode waveguides for single-molecule analysis at high concentrations. Sci 2003;299:682–6.

[88] Kasianowicz JJ, Brandin E, Branton D, Deamer DW. Characterization of individual polynucleotide molecules using a membrane channel. Proc Natl Acad Sci U S A 1996;93:13770–3. doi:10.1073/pnas.93.24.13770.

[89] Wiggin M, Tropini C, Tabard-Cossa V, Jetha NN, Marziali A. Nonexponential kinetics of DNA escape from alpha-hemolysin nanopores. Biophys J 2008;95:5317–23. doi:10.1529/biophysj.108.137760.

[90] Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 2016;32:3047–8. doi:10.1093/bioinformatics/btw354.

[91] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–20. doi:10.1093/bioinformatics/btu170.

[92] Edwards DJ, Holt KE. Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. Microb Inform Exp 2013;3:2. doi:10.1186/2042-5783-3-2.

[93] Nagarajan N, Pop M. Sequence assembly demystified. Nat Rev Genet 2013;14:157–67. doi:10.1038/nrg3367.

[94] Commins J, Toft C, Fares MA. Computational biology methods and their application to the comparative genomics of endocellular symbiotic bacteria of insects. Biol Proced Online 2009;11:52–78. doi:10.1007/s12575-009-9004-1.

[95] Compeau PEC, Pevzner PA, Tesler G. How to apply de Bruijn graphs to genome assembly. Nat Biotechnol 2011;29:987–91. doi:10.1038/nbt.2023.

[96] Miller JR, Koren S, Sutton G. Assembly algorithm for next-generation sequencing data. Genomics 2010;95:315–27. doi:10.1016/j.ygeno.2010.03.001.Assembly.

[97] Pevzner PA, Tang H, Waterman MS. An Eulerian path approach to DNA fragment assembly. Proc Natl Acad Sci U S A 2001;98:9748–53. doi:10.1073/pnas.171285098.

[98] Salzberg SL, Phillippy AM, Zimin A, Puiu D, Magoc T, Koren S, et al. GAGE: A critical evaluation of genome assemblies and assembly algorithms. Genome Res 2012;22:557–67. doi:10.1101/gr.131383.111.

[99] Magoc T, Pabinger S, Canzar S, Liu X, Su Q, Puiu D, et al. GAGE-B: An evaluation of genome assemblers for bacterial organisms. Bioinformatics 2013;29:1718–25. doi:10.1093/bioinformatics/btt273.

[100] Gnerre S, Lander ES, Lindblad-Toh K, Jaffe DB. Assisted assembly: how to improve a de novo genome assembly by using related species. Genome Biol 2009;10:R88. doi:10.1186/gb-2009-10-8-r88.

[101] Li R, Fan W, Tian G, Zhu H, He L, Cai J, et al. The sequence and de novo assembly of the giant panda genome. Nature 2010;463:311–7. doi:10.1038/nature08696.

[102] Analysis NB. Canu: scalable and accurate long---read assembly via adaptive 2016:1–35.

[103] Sović I, Šikić M, Wilm A, Fenlon SN, Chen S, Nagarajan N. Fast and sensitive mapping of nanopore sequencing reads with GraphMap. Nat Commun 2016;7:11307. doi:10.1038/ncomms11307.

[104] Berlin K, Koren S, Chin C-S, Drake JP, Landolin JM, Phillippy AM. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. Nat Biotechnol 2015;33:623–30. doi:10.1038/nbt.3238.

[105] Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: Quality assessment tool for genome assemblies. Bioinformatics 2013;29:1072–5. doi:10.1093/bioinformatics/btt086.

[106] Zhu X, Leung HCM, Wang R, Chin FYL, Yiu SM, Quan G, et al. misFinder: identify mis-assemblies in an unbiased manner using reference and paired-end reads. BMC Bioinformatics 2015;16:386. doi:10.1186/s12859-015-0818-3.

[107] Swain MT, Tsai IJ, Assefa SA, Newbold C, Berriman M, Otto TD. A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. Nat Protoc 2012;7:1260–84. doi:10.1038/nprot.2012.068.

[108] Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 2014;9. doi:10.1371/journal.pone.0112963.

[109] Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014;30:2068–9. doi:10.1093/bioinformatics/btu153.

[110] Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics 2008;9:75. doi:10.1186/1471-2164-9-75.

[111] Otto TD, Dillon GP, Degrave WS, Berriman M. RATT: Rapid Annotation Transfer Tool. Nucleic Acids Res 2011;39:1–7. doi:10.1093/nar/gkq1268.

[112] Van Domselaar GH, Stothard P, Shrivastava S, Cruz JA, Guo AC, Dong X, et al. BASys: A web server for automated bacterial genome annotation. Nucleic Acids Res 2005;33:455–9. doi:10.1093/nar/gki593.

[113] Kalkatawi M, Alam I, Bajic VB. BEACON: automated tool for Bacterial GEnome Annotation ComparisON. BMC Genomics 2015;16:616. doi:10.1186/s12864-015-1826-4.

[114] Guindon S, Lethiec F, Duroux P, Gascuel O. PHYML Online - A web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 2005;33:557–9. doi:10.1093/nar/gki352.

[115] Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014;30:1312–3. doi:10.1093/bioinformatics/btu033.

[116] Pettengill JB, Luo Y, Davis S, Chen Y, Gonzalez-Escalona N, Ottesen A, et al. An evaluation of alternative methods for constructing phylogenies from whole genome sequence data: a case study with *Salmonella*. PeerJ 2014;2:e620. doi:10.7717/peerj.620.

[117] He Z, Zhang H, Gao S, Lercher MJ, Chen W-H, Hu S. Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. Nucleic Acids Res 2016;44:W236-41. doi:10.1093/nar/gkw370.

[118] Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, Harris SR. Phandango: an interactive viewer for bacterial population genomics. Bioinformatics 2018;34:292–3. doi:10.1093/bioinformatics/btx610.

[119] Laing C, Buchanan C, Taboada EN, Zhang Y, Kropinski A, Villegas A, et al. Pan-genome sequence analysis using Panseq: an online tool for the rapid analysis of core and accessory genomic regions. BMC Bioinformatics 2010;11:461. doi:10.1186/1471-2105-11-461.

[120] Zhao Y, Wu J, Yang J, Sun S, Xiao J, Yu J. PGAP: Pan-genomes analysis pipeline. Bioinformatics 2012;28:416–8. doi:10.1093/bioinformatics/btr655.

[121] Chaudhari NM, Gupta VK, Dutta C. BPGA-an ultra-fast pan-genome analysis pipeline. Sci Rep 2016;6:1–10. doi:10.1038/srep24373.

[122] Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: Rapid large-scale prokaryote pan genome analysis. Bioinformatics 2015;31:3691–3. doi:10.1093/bioinformatics/btv421.

[123] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–75. doi:10.1086/519795.

[124] Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol 2016;17:238. doi:10.1186/s13059-016-1108-8.

[125] Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, et al. VFDB: A reference database for bacterial virulence factors. Nucleic Acids Res 2005;33:325–8. doi:10.1093/nar/gki008.

[126] Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid

multilocus sequence typing. Antimicrob Agents Chemother 2014;58:3895–903. doi:10.1128/AAC.02412-14.

[127] Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: A Fast Phage Search Tool. Nucleic Acids Res 2011;39:W347–52. doi:10.1093/nar/gkr485.

[128] Al-Hasan MN, Eckel-Passow JE, Baddour LM. Bacteremia complicating gram-negative urinary tract infections: a population-based study. J Infect 2010;60:278–85. doi:10.1016/j.jinf.2010.01.007.

[129] Marschall J, Zhang L, Foxman B, Warren DK, Henderson JP. Both Host and Pathogen Factors Predispose to Escherichia coli Urinary-Source Bacteremia in Hospitalized Patients. Clin Infect Dis 2012;54:1692–8. doi:10.1093/cid/cis252.

[130] Schmiemann G, Kniehl E, Gebhardt K, Matejczyk MM, Hummers-Pradier E. The diagnosis of urinary tract infection: a systematic review. Dtsch Arztebl Int 2010;107:361–7. doi:10.3238/arztebl.2010.0361.

[131] Rankin DJ, Rocha EPC, Brown SP. What traits are carried on mobile genetic elements and why? Heredity (Edinb) 2011;106:1–10. doi:10.1038/hdy.2010.24.

[132] Melzer M, Petersen I. Mortality following bacteraemic infection caused by extended spectrum beta-lactamase (ESBL) producing E . coli compared to non-ESBL producing E . coli * 2007:254–9. doi:10.1016/j.jinf.2007.04.007.

[133] Pyelonephritis E, Ender PT, Gajanana D, Johnston B, Clabots C, Tamarkin FJ, et al. Transmission of an Extended-Spectrum-Beta-Lactamase-Producing Escherichia coli (Sequence Type ST131) Strain between a Father and Daughter Resulting in Septic Shock and 2009;47:3780–2. doi:10.1128/JCM.01361-09.

[134] Schmiedel J, Falgenhauer L, Domann E, Bauerfeind R, Prenger-Berninghoff E, Imirzalioglu C, et al. Multiresistant extended-spectrum β-lactamase-producing Enterobacteriaceae from humans, companion animals and horses in central Hesse, Germany. BMC Microbiol 2014;14:187. doi:10.1186/1471-2180-14-187.

[135] Bankevich A, Nurk S, Antipov D, Gurevich A a, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 2012;19:455–77. doi:10.1089/cmb.2012.0021.

[136] Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol 2014;15:524. doi:10.1186/s13059-014-0524-x.

[137] Petty NK, Ben Zakour NL, Stanton-Cook M, Skippington E, Totsika M, Forde BM, et al. Global dissemination of a multidrug resistant Escherichia coli clone. Proc Natl Acad Sci 2014;111:5694–9. doi:10.1073/pnas.1322678111.

[138] Nicolas-Chanoine M-H, Bertrand X, Madec J-Y. Escherichia coli ST131, an intriguing clonal group. Clin Microbiol Rev 2014;27:543–74. doi:10.1128/CMR.00125-13.

[139] Johnson JR, Johnston B, Kuskowski MA, Sokurenko E V, Tchesnokova V. Intensity and Mechanisms of Fluoroquinolone Resistance within the H30 and H30Rx Subclones of Escherichia coli Sequence Type 131 Compared with Other

Fluoroquinolone-Resistant *E. coli*. Antimicrob Agents Chemother 2015;59:4471–80. doi:10.1128/AAC.00673-15.

[140] Price LB, Johnson JR, Aziz M. The Epidemic of Extended-Spectrum- $^{N}_{L}$ -Lactamase-Producing Subclone , H 30-Rx 2013. doi:10.1128/mBio.00377-13.Editor.

[141] Johnson TJ, Danzeisen JL, Youmans B, Case K, Llop K, Munoz-Aguayo J, et al. Separate F-Type Plasmids Have Shaped the Evolution of the H 30 Subclone of Escherichia coli Sequence Type 131. mSphere 2016;1:e00121-16. doi:10.1128/mSphere.00121-16.

[142] Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014;30:2068–9. doi:10.1093/bioinformatics/btu153.

[143] Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res 2011;39:W347-52. doi:10.1093/nar/gkr485.

[144] Sahl JW, Caporaso JG, Rasko DA, Keim P. The large-scale blast score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. PeerJ 2014;2:e332. doi:10.7717/peerj.332.

[145] Lanza VF, de Toro M, Garcillán-Barcia MP, Mora A, Blanco J, Coque TM, et al. Plasmid flux in Escherichia coli ST131 sublineages, analyzed by plasmid constellation network (PLACNET), a new method for plasmid reconstruction from whole genome sequences. PLoS Genet 2014;10:e1004766. doi:10.1371/journal.pgen.1004766.

[146] Matsumura Y, Johnson JR, Yamamoto M, Nagao M, Tanaka M, Takakura S, et al. CTX-M-27- and CTX-M-14-producing, ciprofloxacin-resistant Escherichia coli of the H30 subclonal group within ST131 drive a Japanese regional ESBL epidemic. J Antimicrob Chemother 2015;58:4997–5004. doi:10.1093/jac/dkv017.

[147] Birgy A, Bidet P, Levy C, Sobral E, Cohen R, Bonacorsi S. CTX-M-27-Producing Escherichia coli of Sequence Type 131 and Clade C1-M27, France. Emerg Infect Dis 2017;23:885. doi:10.3201/eid2305.161865.

[148] Phan M-D, Peters KM, Sarkar S, Lukowski SW, Allsopp LP, Gomes Moriel D, et al. The serum resistome of a globally disseminated multidrug resistant uropathogenic Escherichia coli clone. PLoS Genet 2013;9:e1003834. doi:10.1371/journal.pgen.1003834.

[149] Moremi N, Manda E V, Falgenhauer L, Ghosh H, Imirzalioglu C, Matee M, et al. Predominance of CTX-M-15 among ESBL Producers from Environment and Fish Gut from the Shores of Lake Victoria in Mwanza, Tanzania. Front Microbiol 2016;7:1862. doi:10.3389/fmicb.2016.01862.

[150] Johnson JR, Johnston B, Clabots C, Kuskowski M a, Castanheira M. Escherichia coli sequence type ST131 as the major cause of serious multidrug-resistant *E. coli* infections in the United States. Clin Infect Dis 2010;51:286–94. doi:10.1086/653932.

[151] Wu G, Day MJ, Mafura MT, Nunez-Garcia J, Fenner JJ, Sharma M, et al. Comparative analysis of ESBL-positive Escherichia coli isolates from animals and

humans from the UK, The Netherlands and Germany. PLoS One 2013;8:e75392. doi:10.1371/journal.pone.0075392.

[152] Kleinheinz KA, Joensen KG, Larsen MV. Applying the ResFinder and VirulenceFinder web-services for easy identification of acquired antibiotic resistance and E . coli virulence genes in bacteriophage and prophage nucleotide sequences 2014:1–7.

[153] Skippington E, Ragan MA. Lateral genetic transfer and the construction of genetic exchange communities. FEMS Microbiol Rev 2011;35:707–35. doi:10.1111/j.1574-6976.2010.00261.x.

[154] He S, Hickman AB, Varani AM, Siguier P, Chandler M, Dekker JP, et al. Insertion Sequence IS26 Reorganizes Plasmids in Clinically Isolated Multidrug-Resistant Bacteria by Replicative Transposition. MBio 2015;6:e00762. doi:10.1128/mBio.00762-15.

[155] Johnson JR, Johnston B, Clabots C, Kuskowski MA, Castanheira M. Escherichia coli sequence type ST131 as the major cause of serious multidrug-resistant *E. coli* infections in the United States. Clin Infect Dis 2010;51:286–94. doi:10.1086/653932.

[156] Falgenhauer L, Imirzalioglu C, Ghosh H, Gwozdzinski K, Schmiedel J, Gentil K, et al. Circulation of clonal populations of fluoroquinolone-resistant CTX-M-15-producing Escherichia coli ST410 in humans and animals in Germany. Int J Antimicrob Agents 2016;47:457–65. doi:10.1016/j.ijantimicag.2016.03.019.

[157] Tanwar J, Das S, Fatima Z, Hameed S. Multidrug resistance: an emerging crisis. Interdiscip Perspect Infect Dis 2014;2014:541340. doi:10.1155/2014/541340.

[158] Roca I, Akova M, Baquero F, Carlet J, Cavaleri M, Coenen S, et al. The global threat of antimicrobial resistance: science for intervention. New Microbes New Infect 2015;6:22–9. doi:10.1016/J.NMNI.2015.02.007.

[159] Nathan C, Cars O. Antibiotic Resistance — Problems, Progress, and Prospects. N Engl J Med 2014;371:1761–3. doi:10.1056/NEJMp1408040.

[160] Levy SB. Factors impacting on the problem of antibiotic resistance. J Antimicrob Chemother 2002;49:25–30. doi:10.1093/jac/49.1.25.

[161] Brolund A. Overview of ESBL-producing Enterobacteriaceae from a Nordic perspective. Infect Ecol Epidemiol 2014;4. doi:10.3402/iee.v4.24555.

[162] Guenther S, Ewers C, Wieler LH. Extended-Spectrum Beta-Lactamases Producing *E. coli* in Wildlife, yet Another Form of Environmental Pollution? Front Microbiol 2011;2:246. doi:10.3389/fmicb.2011.00246.

[163] Thomson KS. Extended-spectrum-beta-lactamase, AmpC, and Carbapenemase issues. J Clin Microbiol 2010;48:1019–25. doi:10.1128/JCM.00219-10.

[164] Dhillon RH-P, Clark J. ESBLs: A Clear and Present Danger? Crit Care Res Pract 2012;2012:625170. doi:10.1155/2012/625170.

[165] Cerqueira GC, Earl AM, Ernst CM, Grad YH, Dekker JP, Feldgarden M, et al. Multi-institute analysis of carbapenem resistance reveals remarkable diversity, unexplained mechanisms, and limited clonal outbreaks. Proc Natl Acad Sci 2017;114:1135–40. doi:10.1073/pnas.1616248114.

[166] Clark G, Paszkiewicz K, Hale J, Weston V, Constantinidou C, Penn C, et al. Genomic analysis uncovers a phenotypically diverse but genetically homogeneous Escherichia coli ST131 clone circulating in unrelated urinary tract infections. J Antimicrob Chemother 2012;67:868–77. doi:10.1093/jac/dkr585.

[167] Petty NK, Ben Zakour NL, Stanton-Cook M, Skippington E, Totsika M, Forde BM, et al. Global dissemination of a multidrug resistant Escherichia coli clone. Proc Natl Acad Sci U S A 2014;111:5694–9. doi:10.1073/pnas.1322678111.

[168] Price LB, Johnson JR, Aziz M, Clabots C, Johnston B, Tchesnokova V, et al. The epidemic of extended-spectrum-β-lactamase-producing Escherichia coli ST131 is driven by a single highly pathogenic subclone, H30-Rx. MBio 2013. doi:10.1128/mBio.00377-13.

[169] Ghosh H, Doijad S, Bunk B, Falgenhauer L, Yao Y, Spröer C, et al. Detection of translocatable units in a blaCTX-M-15 extended-spectrum β-lactamase-producing ST131 Escherichia coli isolate using a hybrid sequencing approach. Int J Antimicrob Agents 2016;47:245–7. doi:10.1016/j.ijantimicag.2016.01.003.

12. Acknowledgments

Firstly, I would like to express my sincerest gratitude to my Ph.D. supervisor, Prof. Dr. Trinad Chakraborty, to undertake my doctoral studies with his excellent guidance, constant support, freedom, innumerable ideas and valuable discussions throughout my doctoral work.

Prof. Dr. Christoph Lämmlerfor being second supervisor, his encouragement and valuable comments on my thesis.

My sincere thanks to Dr. Can Imirzalioglu for his encouragement, insightful comments, and hard questions.

I would like to express my gratitude to Dr. Linda.Falgenhauer, Dr. Moritz Fritzenwanker, Dr. Yancheng Yao and Dr. Torsten Hain for their moral support guidance throughout my doctoral thesis. I would have been lost without their constant support.

I would like thank Prof. Dr. Jörg Overmann, and Dr. Boyke Bunk, for hosting me into their laboratory of two weeks and taking care of me.

I wish to thank my wonderful colleagues Swapnil, Goapala, Niladri, Kondrad, Sanjeeb, Tilman, Judith for their friendship, help, encouragement, and great support. I thank all others member of the Institute of Medical Microbiology who directly and indirectly help me for creating a nice work environment, good co-operation. My special thanks to Martina, Alex, Natalia and Christina for their for excellent technical assistance.
Most importantly, I wish to thank my dear parents: Arabinda Ghosh and Nirmala Ghosh for supporting me spiritually throughout my life and my brother Sisir Ghosh, Ajoy Ghosh, Sanjoy Ghosh and Sujay Ghosh, deceased. They raised me, supported me, taught me, and loved me. To them I dedicate this thesis.

13. Declaration

I declare that I have completed this dissertation single-handedly without the unauthorized help of a second party and only with the assistance acknowledged therein. I have appropriately acknowledged and referenced all text passages that are derived literally from or are based on the content of published or unpublished work of others, and all information that relates to verbal communications. I have abided by the principles of good scientific conduct laid down in the charter of the Justus Liebig University of Giessen in carrying out the investigations described in the dissertation.

---------- ----------

Place, Date Sign