Universiteit
Antwerpen

APPLICATION OF NOVEL GENOMIC TECHNIQUES TO STUDY MULTI-DRUG RESISTANT BACTERIAL PATHOGENS

TOEPASSING VAN NIEUWE GENOMISCHE TECHNIEKEN VOOR HET BESTUDEREN VAN MULTI-DRUGRESISTENTE BACTERIËLE PATHOGENEN

Proefschrift voorgelegd tot het behalen van de graad
van doctor in de Medische Wetenschappen

aan de Universiteit Antwerpen

te verdedigen door

BASIL BRITTO XAVIER

Promotors

Prof. dr. S. Malhotra-Kumar and Prof. dr. H. Goossens

Promotors

Prof. dr. Surbhi Malhotra-Kumar, University of Antwerp

Prof. dr. Herman Goossens, University of Antwerp

Internal Jury

Prof. dr. Geert Mortier, University of Antwerp
Prof. dr. Jurgen Del-Favero, University of Antwerp

External Jury

Prof. dr. João André Carriço, University of Lisbon, Portugal
Prof. dr. Jan Kluytmans, UMC Utrecht, The Netherlands

Application of novel genomic techniques to study multi-drug resistant bacterial pathogens

Laboratory of Medical Microbiology
Vaccine & Infectious Disease Institute
University of Antwerp

Dedicated to

my mother

&

my little princess

(Benita Samara Basil)

TABLE OF CONTENTS

ABBREVIATIONS & GLOSSARY

N50: Parameter to define the quality of the genome assembly by the size and a number of contigs or scaffolds produced by the assembler

Genome assembly: Raw sequencing reads are stitched into larger contiguous sequences known as ʹcontigsʹ and extended contigs called ʹscaffolds'

Hybrid assembly: Raw sequencing reads from second generation (short read), and third generation (long read) technologies are used to make larger contiguous sequences like contigs and scaffolds is called a hybrid assembly

***k mer*:** *k*-mer is a subset of a sequence length of k

Genome annotation: Demarcation of a gene or protein coding sequences, and other genetic features such as tRNA, and rRNA in a raw DNA sequence of genome

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats, which are the indication of a bacterial defense system.

Insertion sequence: Insertion sequence is a short DNA sequence flanked by inverted repeats and act as a transposable element

WGS: Whole genome sequencing

WGM: Whole genome mapping

MLST: Multi locus sequence typing

PFGE: Pulsed field gel electrophoresis

NGS: Next generation sequencing

PCR: Polymerase chain reaction

FMT: Faecal microbial transplantation

AR: Antibiotic resistance

AMR: Antimicrobial resistance

SNPs: Single nucleotide polymorphisms

HGT: Horizontal Gene transfer

MDR: Multi-drug resistant

XDR: Xtreme drug resistant

CPE: Carbapenemase-producing enterobacteriaceae

IS: Insertion sequence

KPC: Klebsiella pneumoniae carbapenemase

MALDI-TOF: matrix-assisted laser desorption ionization time-of-flight

ESBL: Extended-spectrum β-lactamase

RNA: Ribonucleic acid

***bla*:** β-lactamase gene

DNA: Deoxyribonucleic acid

WHO: World Health Organization

ECDC: European Centre for Disease Prevention and Control

NDM: New Delhi metallo-beta-lactamase

AAC: Aminoglycoside N-acetyltransferase

DHFR: Dihydrofolate reductase

GDPR: General Data Protection Regulation

SMRT: Single Molecule Real Time

BLAST: Basic local alignment search tool

MIC: Minimum Inhibitory concentration

CHAPTER I

INTRODUCTION AND AIMS

1.1 GENERAL INTRODUCTION

1.1.1 Antimicrobial resistance

Antibiotics are naturally produced secondary metabolites of microorganisms or modified chemical structures that are synthesised semi-synthetically (Davies and Davies 2010). Antibiotics selectively inhibit the growth or kill other bacteria and are widely used to treat infectious diseases. One of the major public health threats is antibiotic resistance (WHO, 2012). However, antibiotic resistance existed even before the first antibiotic was discovered in 1928 and, even before Antonie van Leeuwenhoek observed the first bacteria in 1676 that belonged to the genus *Selenomonas*, a crescent-shaped bacterium from the oral microbiome (Bardell 1983). Antibiotic resistance existed in nature, even before the therapeutic use of antibiotics. As an indication, the structural components of non-ribosomal peptides of antibiotics existed in space for billions of years (Davies and Davies 2010).

Antibiotics such as erythromycin, a polyketide molecule, evolved as early as 800 million years ago. Similarly, the streptomycin pathway is as old as 600 million years (Davies and Davies 2010). Antibiotic resistance is associated with specific bacterial biochemical and physiological processes. Four major resistance mechanisms based on the antibiotic targets have been described: (i) Modifications of the antibiotic, (ii) Prevention to reach the antibiotic target (by decreasing penetration or actively extruding the antimicrobial compound), (iii) Changes in target sites (for example rifampicin resistance), and (iv) Resistance due to global cell adaptive processes (example daptomycin resistance) (Figure 1.1) (Munita and Arias 2016). Bacteria can adapt to abnormal temperatures and extreme environments, rich in metals and highly toxic compounds such as antibiotics, by modifying their genetic content (Arsène-Ploetze, Koechler et al. 2010). Human gut bacteria play a significant role in human health, but antibiotic intake will destroy the composition of resident bacteria (Kau, Ahern et al. 2011). The antibiotic resistance also very much depends on another genetic background of bacteria; for instance, co-harbouring metal resistance genes play a significant role in the transfer of antibiotic resistance genes among the bacterial population (Davies and Davies 2010).

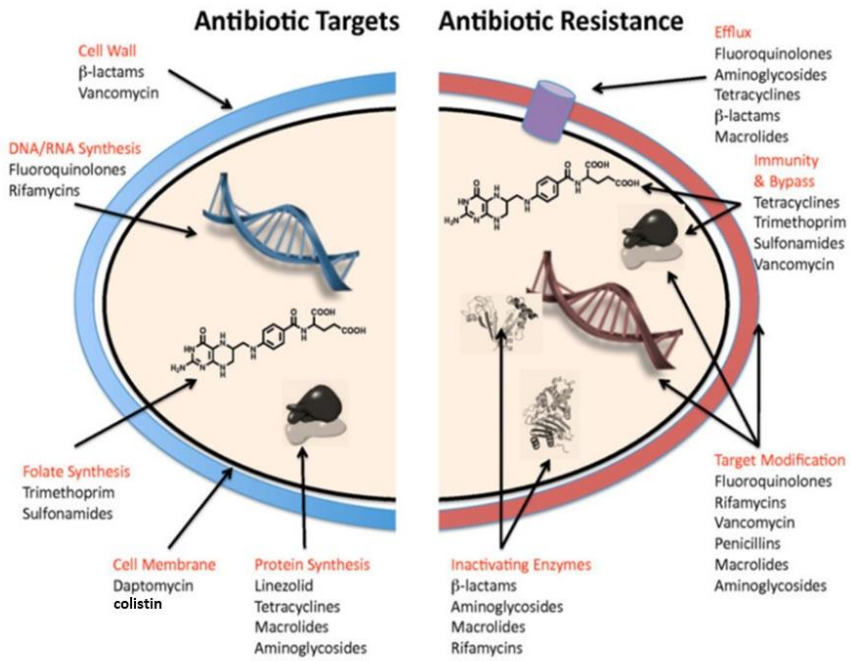


Figure 1.1 The major antibiotic targets and antibiotic resistance mechanisms (Figure is reproduced/updated from Wright GD, 2010 (Wright 2010))

Infections caused by multidrug-resistant (MDR) pathogens, such as extended-spectrum β-lactamase-harbouring Enterobacteriaceae (ESBL-EN), *Escherichia coli* and *Klebsiella pneumoniae* or methicillin-resistant *Staphylococcus aureus* (MRSA) are a worldwide threat to public health (Stefani and Varaldo 2003, Pitout and Laupland 2008, Stefani, Chung et al. 2012). Notably, the spread of MDR *E. coli* and *K. pneumoniae* strains constitutes a pressing global health problem. The increasing prevalence of plasmid-encoded carbapenem-hydrolysing enzymes in *K. pneumoniae* is of particular concern due to their ability to hydrolyse almost all β-lactam antibiotics, as well as their genetic association with transferable multidrug-resistance (Mathers, Cox et al. 2011). Infections due to carbapenem-resistant *K. pneumoniae* are challenging to treat due to limited therapeutic options, and clones such as ST258, ST147, ST101 and ST383 could potentially cause hospital epidemics if not promptly detected and contained (Saidel-Odes and Borer 2014).

The monumental increase in antibiotic resistance among important bacterial pathogens, driven by the inappropriate and appropriate use of ineffective drugs (Littmann and Viens 2015). Recently, WHO called for new antibiotics that are needed to control the priority pathogens such as the carbapenem-resistant *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and ESBL and carbapenem-resistant Enterobacteriaceae (including *Klebsiella*, *E. coli*, *Serratia* and *Proteus*). These pathogens cause the majority of hospital-acquired infections worldwide and are mostly MDR and extremely drug-resistant (XDR) gram-negative pathogens (Walsh and Toleman 2012) (Li,

Nation et al. 2006, WHO 2014).

1.1.2 Colistin

Colistin (polymyxin E), a cyclic polypeptide, is one of the oldest and last line antibiotics available to treat infections caused by MDR bacteria. It belonged to the polymyxin group (cationic) and was isolated from *Bacillus polymyxa var. colistinus* in 1949 in Japan. Colistin was first introduced in 1952 and is sold commercially as colistin sulfate and colistin methanesulfonate or colistimethate sodium (CMS) (Biswas, Brunel et al. 2012). Colistin has attracted considerable interest for use in both human and veterinary medicine. However, it was withdrawn from clinical use in the 1980s due to its neuro- and nephrotoxicities and other pharmacodynamics related issues (Biswas, Brunel et al. 2012). Since its introduction, colistin is used for various purposes, as a prophylactic drug in humans, and to treat diarrhea in weaning piglets (Katsunuma, Hanazumi et al. 2007, Rhouma, Beaudry et al. 2016). Its use persisted for decades in swine farms and resulted in a significant increase in colistin resistance rates of *Enterobacteriaceae* isolated from pigs (Harada, Asai et al. 2005). Due to the scarcity of novel antibiotics to treat infections in humans and the reduced efficiency of existing antibiotics to treat infections caused by MDR and XDR bacteria, colistin was reintroduced in humans to treat these recently emerged antibiotic resistant bacteria. However, antibiotics of the polymyxin group, such as colistin, are sometimes the only drug to which these bacteria show sensitivity and therefore, any emergence of resistance to polymyxin antibiotics is of paramount importance (Falagas and Kasiakou 2005, Walsh and Toleman 2012). Considering the importance of these compounds, WHO and the National Institution of Canada have reclassified colistin as relevant to human medicine (WHO, 2011; Government of Canada, 2014).

1.1.3 Colistin resistance

According to the European Antimicrobial Resistance Surveillance Network (EARS-Net) data of 2013, 8.8% of *K. pneumoniae* isolates were found to be resistant to colistin.The majority of strains were derived from Greece, Italy, Romania, and Hungary, with 32% of carbapenem-resistant *K. pneumoniae* strains that were also resistant to polymyxins (ECDC, EARS-Net 2014 report). As expected, high-level carbapenem resistance in these countries led to excessive use of polymyxins. As a consequence, in Greece, the 6-fold increase of use of colistin between 2009 and 2013 favoured the emergence of colistin resistance rates among *K. pneumoniae* in intensive care units during the

same period from 0% to 21% (Meletis, Oustas et al. 2015).

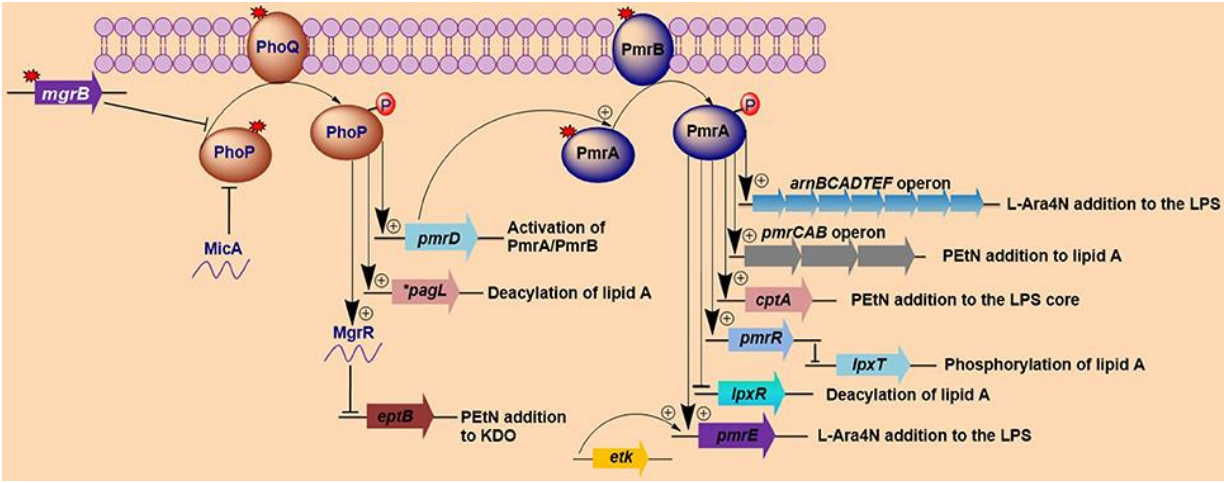


Figure 1.2 Polymyxin resistance in Gram-negative bacteria (adapted from Olaitan et al., 2014(Olaitan, Morand et al. 2014))

In Belgium, the increase and spread of Carbapenemase-producing Enterobacteriaceae (CPE) has been observed countrywide and has also driven the use of colistin in the hospital sector. For the last decade, the consumption of colistin in hospitals has increased from 0.0010 DDD per 1000 inhabitants per day (DID) in 2005 to 0.0078 DID in 2014 (ECDC, ESAC-Net data 2015) page 9.

Colistin is a cationic antibiotic, which causes severe damage to both the Gram-negative cytoplasmic and outer membranes. Colistin acts by displacing divalent cations from the phosphate groups of the outer membrane lipids during interaction with the lipopolysaccharide (LPS) and phospholipids membranes, which have net negative charge and are the primary target for colistin (Olaitan, Morand et al. 2014). Colistin resistance mechanisms have been studied in *A. baumannii* (Figure 1.2). The loss of LPS (Moffatt, Harper et al. 2010) resulted due to non-synonymous mutations in *pmrB* that lead to constitutive expression of the two-component signalling system, *pmrAB* that senses the environmental pH, Fe^{3+}, and Mg^{2+} levels (Beceiro, Llobet et al. 2011).

The exact mode of colistin action is unknown, but it is supposed that the mechanisms are very similar to another polymyxin B (Li, Nation et al. 2006). The cationic cyclic peptide acts by binding to the anionic LPS present in the lipid A molecule (Nation and Li 2009). The Mg^{2+} and Ca^{2+} present in the membrane are replaced, which increases cell permeability, followed by leakage of cell-contents, lysis of cell, and finally, bacterial cell death (Figure 1.3).

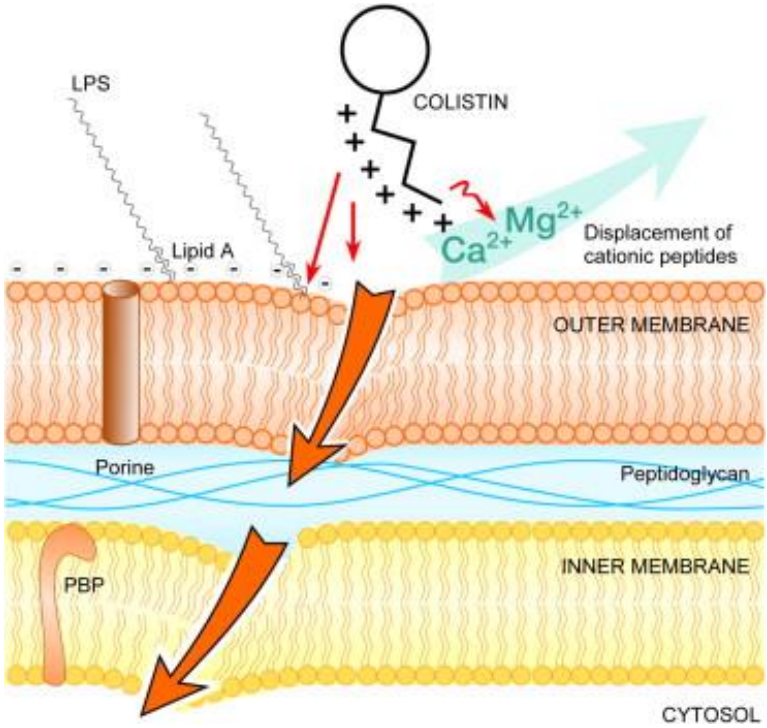


Figure 1.3 Action of colistin on bacterial membrane. The cationic cyclic decapeptide structure of colistin binds with the anionic lipopolysaccharide (LPS) molecules by displacing calcium and magnesium from the outer cell membrane of Gram-negative bacteria, leading to permeability changes in the cell envelope and leakage of cell contents (Figure is reproduced from Martis et al. 2014) (Martis, Leroy et al.).

In *K. pneumoniae,* non-synonymous mutations in *pmrA, pmrB, phoQ, pmrC* genes have shown a role in the development of colistin resistance. Another mechanism often reported in the emergence of colistin resistance in *K.pneumoniae* is the modification of the conserved *mgrB* gene, which is a small transmembrane protein interrupted by insertion elements (IS5-like elements), that exhibits negative feedback to *PhoP/PhoQ* regulatory systems (Cannatelli, Giani et al. 2014, Poirel, Jayol et al. 2015, Olaitan and Li 2016). Additionally, the capsular genes of *K. pneumoniae,* as well as an efflux pump *kpnEF,* are implicated in colistin resistance in *K. pneumoniae* (Srinivasan and Rajamohan 2013). All colistin resistance mechanisms are the result of the activation of the two-component systems (TCSs) *PhoP*/*PhoQ* and *PmrA*/*PmrB* by specific mutations or environmental stimuli leading to an overexpression of LPS-modifying genes in Gram-negative bacteria (Figure 1.2) (Olaitan, Morand et al. 2014). In *A. baumannii* colistin resistance was shown to result from the

loss of LPS, the non-synonymous mutation in the *pmrB gene* (Wand, Bock et al. 2015). Very interestingly, recent data point towards involvement of an Hfq-dependent sRNA, *mgrR*, that is one of the regulons regulated by the PhoP/PhoQ regulatory systems in colistin resistant *E. coli* (Reynolds, Kalb et al. 2005, Moon and Gottesman 2009).

The extensive use of colistin in food-producing animals has resulted in increased CR levels among *E. coli* of animal origin, and the reservoir of novel genes acquired through horizontal gene transfer from other species have raised concern about the use of this drug of last resort in humans (Harada, Asai et al. 2005, Wang, Wu et al. 2008). The discovery of a plasmid-mediated *mcr* gene and its variants adds further complexity to the colistin resistance mechanism (Liu, Wang et al. 2015, Rebelo, Bortolaia et al. 2018), and identification of horizontally transferred gene conferring colistin resistance has raised global alarm. Moreover, the link between animals and humans regarding colistin-resistant *E. coli* strain transfer following direct contact has recently been confirmed (Olaitan, Thongmalayvong et al. 2015). These findings have led to a severe fear about the possible loss of colistin effectiveness in the treatment of MDR Gram-negative bacilli in humans.

Adding further complexity to CR mechanistic studies is the fact that there is no standardised method as yet to determine susceptibility or resistance to colistin. The lack of such a standardised method has been attributed to colistin physical characteristics, such as binding to plastics, inability to hydrolyse entirely and uneven diffusion through agar (Sader, Rhomberg et al. 2012). Most studies have used the Epsilometer E-test gradient strips, which is a an easy method to test the susceptibility to colistin (Galani, Kontopidou et al. 2008, Behera, Mathur et al. 2010, Lee, Shin et al. 2013). However, the method shows 2-fold higher minimum inhibitory concentrations (MICs) to colistin as well as irreproducible results in comparison to other methods such as the macro broth dilution or agar dilution (Turlej-Rogacka, Xavier et al. 2017).

1.1.4 Heteroresistance

"Heteroresistance" describes a phenomenon where subpopulations of seemingly isogenic bacteria exhibit a range of susceptibilities to a particular antibiotic (El-Halfawy, Valvano 2015). Studies suggest that resistant cells selected from growth in the presence of lower antibiotic concentrations are much fitter than the resistant cells that developed during growth in the presence of well above MICs of the strains (Bengtsson-Palme and Larsson 2016). Heteroresistance has been described in Gram-negative pathogens, especially in *K. pneumoniae* (Moffatt, Harper et al. 2010, Meletis, Tzampaz et al. 2011, Jayol, Nordmann et al. 2015, Nodari, Ribeiro et al. 2015) and has been

considered responsible for treatment failure (Band, Satola et al. 2018). Studies have shown that differences in gene expression profile and genetic variation might play a role in underlying heteroresistance to the antibiotic (Hjort, Nicoloff et al. 2016). As colistin is currently used to treat infections due to MDR or XDR pathogens, the clinical implications of *in vivo* emergence of heteroresistant populations under colistin therapy pose a threat to the clinical outcome (Band, Satola et al. 2018).

The use of colistin in the agricultural sector and farming animals is likely to reach other natural reservoirs such as soil and water systems like lakes and rivers, and through water systems to sewage and manure. Thus, as a consequence of the constant presence of colistin in the environment, selective pressure may be created that will eventually lead to the emergence of colistin resistance, by mutation-selection or bacterial plasmid transfer through different routes in the ecosystem (Berendonk, Manaia et al. 2015, Hughes and Andersson 2017) (Figure 1.4).

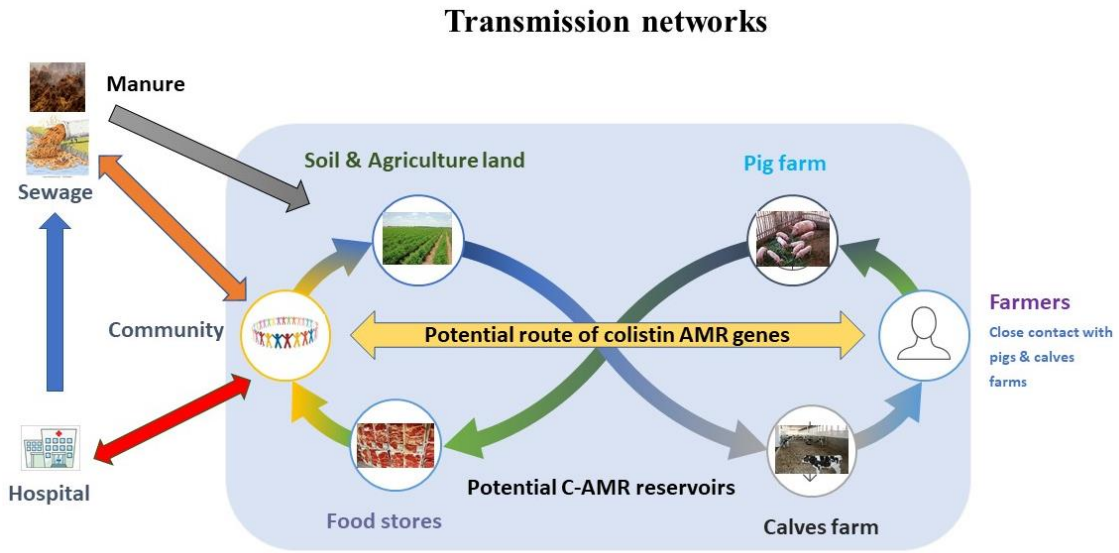


Figure 1. 4 Potential reservoirs and transmission network of colistin (C) resistance genes in the ecosystem.

1.2 TECHNOLOGY BACKGROUND

1.2.1 Molecular methods used in the detection of antibiotic resistance

The most conventional method used to detect antimicrobial resistance is the Kirby-Bauer test (commonly called the disc diffusion test). In this method, the inhibitory zone area is determined, and it is one of the oldest methods used to detect antimicrobial resistance. This method was first used in the 1950's and was recommended by WHO in 1961. However, this method is time-

consuming and labour intensive, so most of the labs have switched to the automated methods. Other methods, such as the dilution method (broth and agar dilution method) and E-test (Epsilometer test) are also used. Automated systems, such as Vitek System (bioMerieux, France), Walk-Away System (Dade International, Sacramento, Calif.), Sensititre ARIS (Trek Diagnostic Systems, East Grinstead, UK), Advantage Test System (Abbott Laboratories, Irving, Texas), Micronaut (Merlin, Bornheim-Hesel, Germany), and Phoenix (BD Biosciences, Maryland) are widely used because they reduce the time and manpower for measuring of MICs. All the methods mentioned above are used for phenotypic detection of antimicrobial resistance.

Currently, polymerase chain reaction (PCR), DNA microarray, whole-genome sequencing, metagenomics, and matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF-MS) are used to detect antibiotic resistance genes (Anjum, Zankari et al 2017). Molecular typing methods to determine genetic diversity among resistant strains and the genomic characterisation of these bacteria play a vital role in studying antimicrobial resistance and molecular epideomiology. Methods commonly used are: amplification of marker genes followed by sequencing (multilocus sequence typing, MLST), or targeting entire genomes by restriction fragments analysed by pulsed-field gel electrophoresis (PFGE) and whole genome mapping (WGM). WGM is new technology, with the restriction fragments analysed on a microfluidic device (Miller 2013). Aforementioned molecular typing methods are used to track, trace and monitor outbreaks of antimicrobial resistance in public health environment.

1.2.1 DNA sequencing

Next-generation sequencing revolutionized biological research in the 21st century. It has completely changed the view of characterising microbes or any living cell molecular information. The first sequencing method was introduced by Gilbert and Maxam for the sequence of 24 base pair sequencing by wandering spot analysis (Gilbert and Maxam 1973)) shortly after Sanger introduced chain termination or dideoxy technique which utilize modified bases of the dideoxyribonucleotides (ddNTPs) which lack 3’ hydroxyl group which is required for the DNA strand synthesis. By mixing of DNA polymerase, and dNTPs (dATP, dTTP, dGTP, and dCTP) along with modified ddNTPs (dideoxynucleotide triphosphates) until ddNTPs get and halt the extension of new strand synthesis (Sanger, Nicklen et al. 1977). In the past two decades, the Sanger sequencing method, referred to as the first generation of sequencing, underwent a lot of changes in automation and faster sequencing process (Figure 1.5A). However, the real fundamental change in the field of sequencing happened in the late 1990’s: the luminescent method by measuring pyrophosphate after base incorporation.

With this method, the polymerase reaction is modified to emit light as each base gets incorporated and this was monitored, and the image captured. This process was developed in parallel sequencing reactions on a micrometre scale. The pyrophosphate-based sequencing (pyrosequencing) was called 454 sequencing, and Roche was the first company to commercialise this technology (Margulies, Egholm et al. 2005). But, the main disadvantage is insertions and deletions of nucleotides in the homopolymer regions. This is due to light intensity emitted by pyrosequencing. Similarily, pyrosequencing was introduced by Life Technologies as Ion torrent semi-conductor technology with modifications instead detecting flurescent labelled nucleotides it uses to detect the release of hydrogen ion during the sequencing process. But, this technology had same homopolymers insertions issue. The revolution of sequencing came from Solexa with new method called sequencing by synthesis. Later this technology was commercialized by Illumina the sequencing-by-synthesis approach works, by using modified dNTPs with a terminator which controls polymerization with a single base addition in DNA strand synthesis. This sequencing reaction takes place at a large scale by attaching fragmented DNA molecules randomly to the optically transparent solid surface (flow cell) with known adapter sequences and followed by bridge amplification (Figure 1.5B) The second-generation sequencing technology has dominated the field for some time now. A significant part of the next generation sequencing data in this thesis is generated from this technology. The third-generation sequencing technology such as Pacbio and Oxford Nanopore aims to sequence DNA rapidly by watching an array of single DNA molecules being replicated in real time (Figure 1.5C) (Heather and Chain 2016).

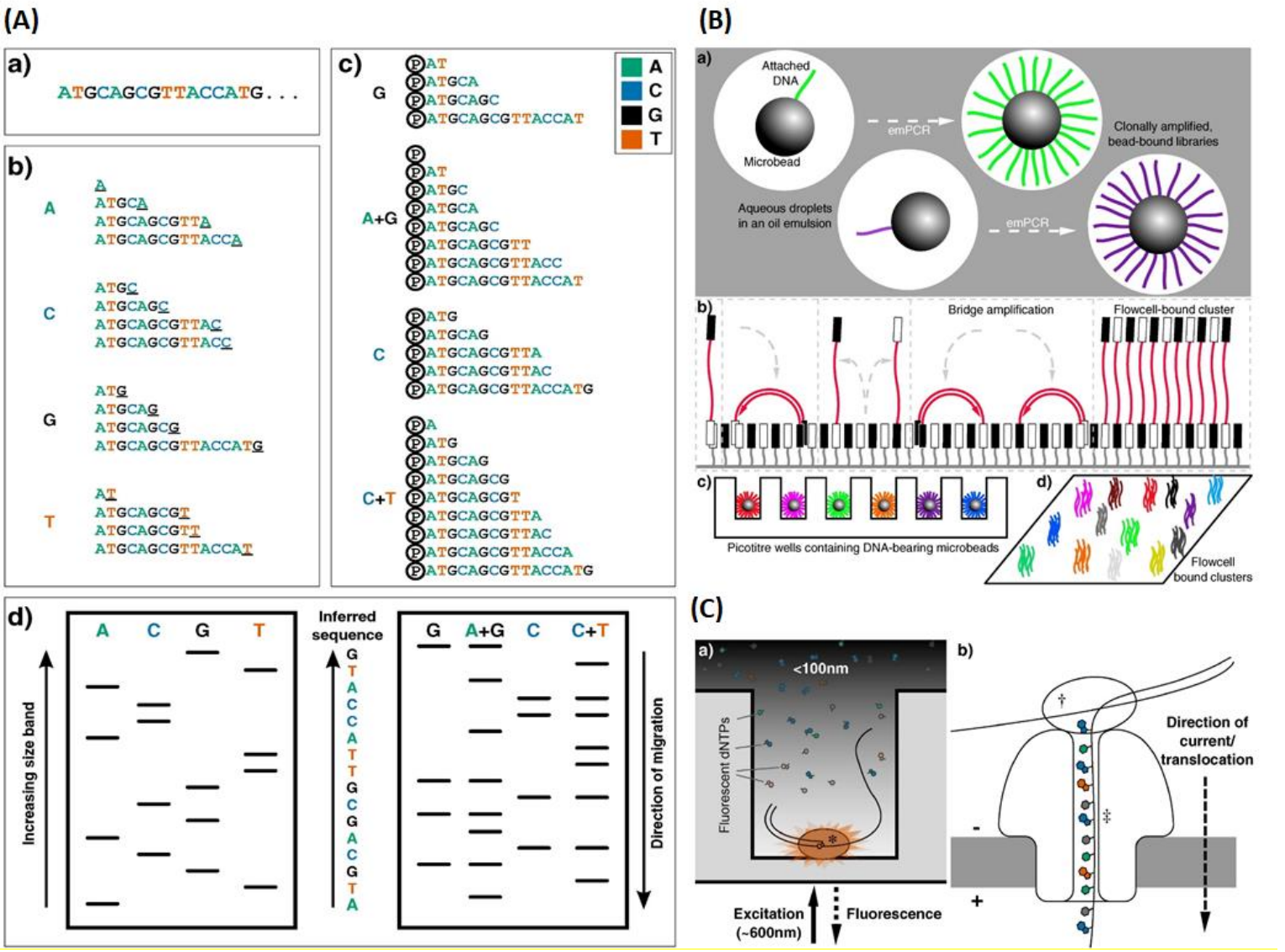

(A)
a)
ATGCAGCGTTACCATG...
b)
A
ATGCA
ATGCAGCGTTA
ATGCAGCGTTACCA
C
ATGC
ATGCAGC
ATGCAGCGTTAC
ATGCAGCGTTACC
G
ATG
ATGCAG
ATGCAGCG
ATGCAGCGTTACCATG
T
AT
ATGCAGCGT
ATGCAGCGTT
ATGCAGCGTTACCAT
c)
G
ATGCAGCGTTACCAT
A+G
ATGCAGCGTTACCAT
C
ATGCAGCGTTACCATG
C+T
ATGCAGCGTTACCATG
A
C
G
T
d)
A C G T
Increasing size band
Inferred sequence
G T A C C A T T G C G A C G T A
G A+G C C+T
Direction of migration
(B)
a)
Attached DNA
Microbead
emPCR
Clonally amplified, bead-bound libraries
Aqueous droplets in an oil emulsion
emPCR
b)
Bridge amplification
Flowcell-bound cluster
c)
Picotitre wells containing DNA-bearing microbeads
d)
Flowcell bound clusters
(C)
a)
<100nm
Fluorescent dNTPs
Excitation (~600nm)
Fluorescence
b)
Direction of current/ translocation
-
+

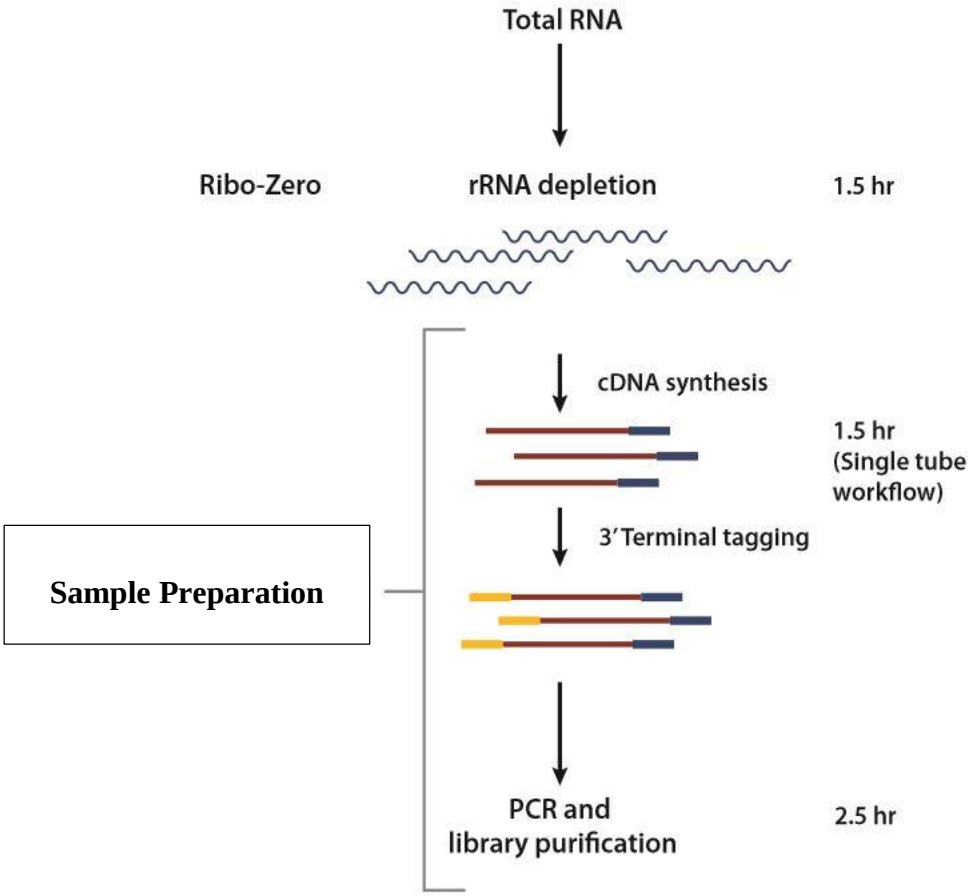
Figure 1.5 A. First-generation DNA sequencing technologies. Example DNA to be sequenced **(**a) is illustrated undergoing either Sanger (b) or Maxam–Gilbert (c) sequencing. (b): Sanger's 'chain-termination' sequencing. Radio- or fluorescently-labelled ddNTP nucleotides of a given type - which once incorporated, prevent further extension - are included in DNA polymerisation reactions at low concentrations (primed off a 5′ sequence, not shown). Therefore, in each of the four reactions, sequence fragments are generated with 3′ truncations as a ddNTP is randomly incorporated at a particular instance of that base (underlined 3′ terminal characters). (c): Maxam and Gilbert's 'chemical sequencing' method. DNA must first be labelled, typically by the inclusion of radioactive P^{32} in its 5′ phosphate moiety (shown here by Ⓟ). Different chemical treatments are then used to selectively remove the base from a small proportion of DNA sites. Hydrazine removes bases from pyrimidines (cytosine and thymine), while hydrazine in the presence of high salt concentrations can only remove those from cytosine. Acid can then be used to remove the bases from purines (adenine and guanine), with dimethyl sulphate being used to attack guanines (although adenine will also be affected to a much lesser extent). Piperidine is then used to cleave the phosphodiester backbone at the basic site, yielding fragments of variable length. (d): Fragments generated from either methodology can then be visualised via electrophoresis on a high-resolution polyacrylamide gel: sequences are then inferred by reading 'up' the gel, as the shorter DNA fragments migrate fastest. In Sanger sequencing (left) the sequence is inferred by finding the lane in which the band is present for a given site, as the 3′ terminating labelled ddNTP corresponds to the base at that position. Maxam–Gilbert sequencing requires a small additional logical step: Ts and As can be directly inferred from a band in the pyrimidine or purine lanes respectively, while G and C are indicated by the presence of dual bands in the G and A + G lanes, or C and C + T lanes respectively.

B. Second-generation DNA sequencing parallelised amplification. (a): DNA molecules being clonally amplified in an emulsion PCR (emPCR). Adapter ligation and PCR produces DNA libraries with appropriate 5′ and 3′ ends, which can then be made single-stranded and immobilised onto individual suitably oligonucleotide-tagged microbeads. Bead-DNA conjugates can then be emulsified using aqueous amplification reagents in oil, ideally producing emulsion droplets containing only one bead (illustrated in the two leftmost droplets, with different molecules indicated in different colours). Clonal amplification then occurs during the emPCR as each template DNA is physically separate from all others, with daughter molecules remaining bound to the microbeads. This is the conceptual basis underlying sequencing in 454, Ion Torrent and polony sequencing protocols. (b): Bridge amplification to produce clusters of clonal DNA populations in a planar solid-phase PCR reaction, as occurs in Solexa/Illumina sequencing. Single-stranded DNA with terminating sequences complementary to the two lawn-oligos will anneal when washed over the flow-cell, and during isothermal PCR will replicate in a confined area, bending over to prime at neighbouring sites, producing a local cluster of identical molecules. (c) and (d) demonstrate how these two different forms of clonally-amplified sequences can then be read in a highly parallelized manner: emPCR-produced microbeads can be washed over a picotiter plate, containing wells large enough to fit only one bead (c). DNA polymerase can then be added to the wells, and each nucleotide can be washed over in turn, and dNTP incorporation monitored (e.g. via pyrophosphate or hydrogen ion

release). Flow-cell bound clusters produced via bridge amplification (d) can be visualized by detecting fluorescent reversible-terminator nucleotides at the ends of a proceeding extension reaction, requiring cycle-by-cycle measurements and removal of terminators. C. Third-generation DNA sequencing nucleotide detection. (a): Nucleotide detection in a zero-mode waveguide (ZMW), as featured in PacBio sequencers. DNA polymerase molecules are attached to the bottom of each ZMW (*), and target DNA and fluorescent nucleotides are added. As the diameter is narrower than the excitation light's wavelength, illumination rapidly decays travelling up the ZMW: nucleotides being incorporated during polymerisation at the base of the ZMW provide real-time bursts of fluorescent signal, without undue interference from other labelled dNTPs in solution. (b): Nanopore DNA sequencing as employed in ONT's MinION sequencer. Double-stranded DNA gets denatured by a processive enzyme (†) which ratchets one of the strands through a biological nanopore (‡) embedded in a synthetic membrane, across which a voltage is applied. As the ssDNA passes through the nanopore, the different bases prevent ionic flow distinctively, allowing the sequence of the molecule to be inferred by monitoring the current at each channel. (Reproduced from Heather, J. M., & Chain, B. (2016)).

1.2.3 RNA sequencing

The transcriptome is the complete set of transcripts present in the genome for a specific developmental stage or physiological condition. The analysis of the transcriptome can give unique insights into the functional elements of the genome (Wang, Gerstein et al. 2009). Analysis of the transcriptome can be done using different techniques such as microarrays and sequence-based approaches, RNA-Seq. For RNA-Seq, cDNA libraries are prepared from the isolated total RNA after removal of ribosomal RNAs, followed by sequencing of the cDNA (Mortazavi, Williams et al. 2008). Further, *de novo* information can be derived from sequencing because there is no involvement of predefined probes (Raz, Kapranov et al. 2011). Still, the most significant challenge of RNA-Seq is downstream analysis (Wang, Gerstein et al. 2009) (Figure 1.6). Attempts were made to sequence RNA molecules directly but with only limited success so far with Helicos single molecule sequencing. Currently, the main disadvantage is that we follow the same steps as DNA sequencing, except cDNA conversion and rRNA removal.



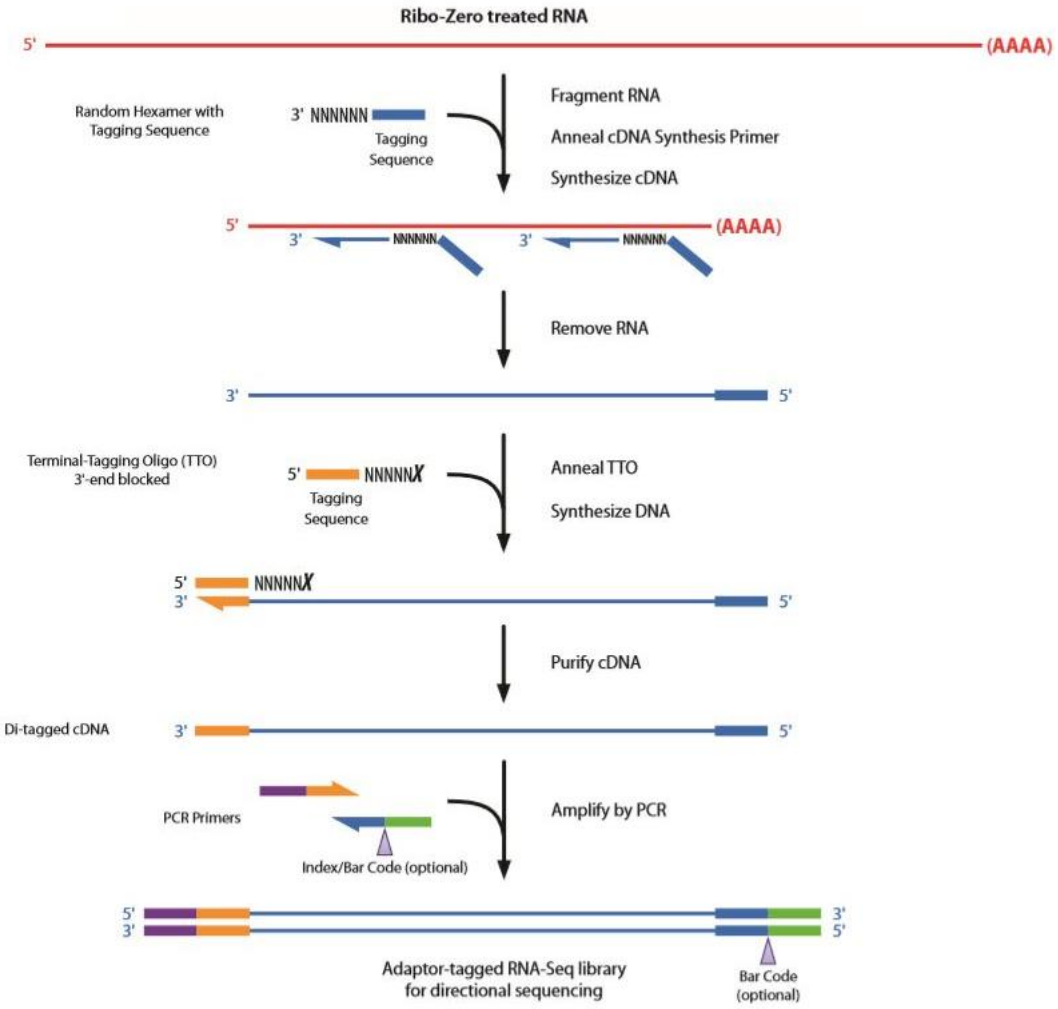


Figure 1. 6 An overview of the procedure for the RNA-Seq Library Preparation (reproduced from Illumina Protocol).

1.3 METAGENOMICS

Metagenomics helps us to directly sequence clinical samples and identify the bacterial composition using either targeted 16S rDNA amplicon sequencing directly from sample, and using random shot gun genomics to identify 'antibiotic resistome', which is the collection of all antibiotic resistance genes in sample, from both pathogenic and non-pathogenic bacteria. With metagenomics, we can define the functional potential, and microbial dynamics, as indigenous microbes also act as a reservoir for antimicrobial resistance genes. Healthy human microbial communities constitute reservoirs of antibiotic resistance genes that can be accessed by potentially pathogenic bacteria

(Dethlefsen, McFall-Ngai et al. 2007, Sommer, Dantas et al. 2009). Selective pressure exerted by extensive antibiotic use in human medicine and agriculture presumably sustains these reservoirs. Resistant organisms can persist for months after antibiotic therapy. In contrast, maintenance of antibiotic resistance gene reservoirs cannot entirely be ascribed to antibiotic use, since resistant organisms are also found in non-antibiotics-exposed healthy individuals (Bartoloni, Pallecchi et al. , Grenet, Guillemot et al. 2004). Better knowledge of antibiotic resistance gene reservoirs in healthy individuals can serve as a source to fully understand the worldwide emergence and spread of multidrug-resistant pathogens. For this, a metagenomics approach can be applied.

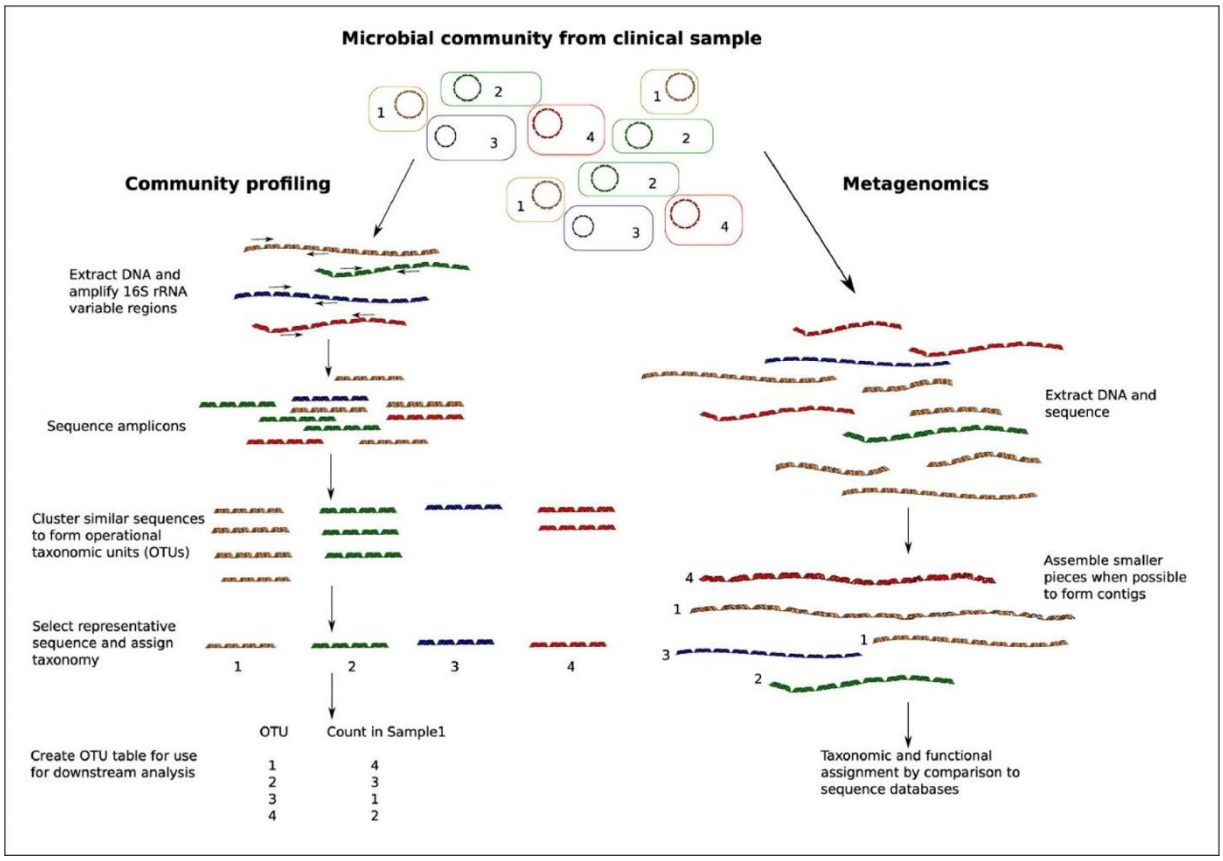


Figure 1. 7 The workflow of metagenomics for clinical samples (Figure adapted from Willner et al 2013 (Willner and Hugenholtz 2013)).

1.4 References

1. Anjum M, Zankari E, Hasman H. 2017. Molecular Methods for Detection of Antimicrobial Resistance. Microbiol Spectrum 5(6): ARBA-0011-2017. doi: 10.1128/microbiolspec.ARBA-0011-2017.
2. Arsène-Ploetze, F., S. Koechler, M. Marchal, J.-Y. Coppée, M. Chandler, V. Bonnefoy, C. Brochier-Armanet, M. Barakat, V. Barbe, F. Battaglia-Brunet, O. Bruneel, C. G. Bryan, J. Cleiss-Arnold, S. Cruveiller, M. Erhardt, A. Heinrich-Salmeron, F. Hommais, C. Joulian, E. Krin, A. Lieutaud, D. Lièvremont, C. Michel, D. Muller, P. Ortet, C. Proux, P. Siguier, D. Roche, Z. Rouy, G. Salvignol, D. Slyemi, E. Talla, S. Weiss, J. Weissenbach, C. Médigue and P. N. Bertin (2010). "Structure, Function, and Evolution of the Thiomonas spp. Genome." PLOS Genetics **6**(2): e1000859.
3. Band, V. I., S. W. Satola, E. M. Burd, M. M. Farley, J. T. Jacob and D. S. Weiss (2018). "Carbapenem-Resistant Klebsiella pneumoniae Exhibiting Clinically Undetected Colistin Heteroresistance Leads to Treatment Failure in a Murine Model of Infection." mBio **9**(2).
4. Bardell, D. (1983). "The roles of the sense of taste and clean teeth in the discovery of bacteria by Antoni van Leeuwenhoek." Microbiological Reviews **47**(1): 121-126.
5. Bartoloni, A., L. Pallecchi, H. Rodríguez, C. Fernandez, A. Mantella, F. Bartalesi, M. Strohmeyer, C. Kristiansson, E. Gotuzzo, F. Paradisi and G. M. Rossolini "Antibiotic resistance in a very remote Amazonas community." International Journal of Antimicrobial Agents **33**(2): 125-129.
6. Beceiro, A., E. Llobet, J. Aranda, J. A. Bengoechea, M. Doumith, M. Hornsey, H. Dhanji, H. Chart, G. Bou, D. M. Livermore and N. Woodford (2011). "Phosphoethanolamine Modification of Lipid A in Colistin-Resistant Variants of Acinetobacter baumannii Mediated by the pmrAB Two-Component Regulatory System." Antimicrobial Agents and Chemotherapy **55**(7): 3370-3379.
7. Behera, B., P. Mathur, A. Das, A. Kapil, B. Gupta, S. Bhoi, K. Farooque, V. Sharma and M. C. Misra (2010). "Evaluation of susceptibility testing methods for polymyxin." International Journal of Infectious Diseases **14**(7): e596-e601.
8. Bengtsson-Palme, J. and D. G. J. Larsson (2016). "Concentrations of antibiotics predicted to select for resistant bacteria: Proposed limits for environmental regulation." Environment International **86**: 140-149.
9. Berendonk, T. U., C. M. Manaia, C. Merlin, D. Fatta-Kassinos, E. Cytryn, F. Walsh, H. Burgmann, H. Sorum, M. Norstrom, M. N. Pons, N. Kreuzinger, P. Huovinen, S. Stefani, T. Schwartz, V. Kisand, F. Baquero and J. L. Martinez (2015). "Tackling antibiotic resistance: the environmental framework." Nat Rev Microbiol **13**(5): 310-317.
10. Biswas, S., J. M. Brunel, J. C. Dubus, M. Reynaud-Gaubert and J. M. Rolain (2012). "Colistin: an update on the antibiotic of the 21st century." Expert Review of Anti-Infective Therapy **10**(8): 917-934.
11. Cannatelli, A., T. Giani, M. M. D'Andrea, V. Di Pilato, F. Arena, V. Conte, K. Tryfinopoulou, A. Vatopoulos and G. M. Rossolini (2014). "MgrB Inactivation Is a

Common Mechanism of Colistin Resistance in KPC-Producing Klebsiella pneumoniae of Clinical Origin." Antimicrobial Agents and Chemotherapy **58**(10): 5696-5703.
12. Davies, J. and D. Davies (2010). "Origins and Evolution of Antibiotic Resistance." Microbiology and Molecular Biology Reviews : MMBR **74**(3): 417-433.
13. Dethlefsen, L., M. McFall-Ngai and D. A. Relman (2007). "An ecological and evolutionary perspective on human–microbe mutualism and disease." Nature **449**: 811.
14. El-Halfawy OM, Valvano MA. Clin Microbiol Rev. 2015 Jan;28(1):191-207. doi: 10.1128/CMR.00058-14. Review.
15. Falagas, M. E. and S. K. Kasiakou (2005). "Colistin: the revival of polymyxins for the management of multidrug-resistant gram-negative bacterial infections." Clin Infect Dis **40**(9): 1333-1341.
16. Galani, I., F. Kontopidou, M. Souli, P.-D. Rekatsina, E. Koratzanis, J. Deliolanis and H. Giamarellou (2008). "Colistin susceptibility testing by Etest and disk diffusion methods." International Journal of Antimicrobial Agents **31**(5): 434-439.
17. Gilbert, W. and A. Maxam (1973). "The Nucleotide Sequence of the lac Operator." Proceedings of the National Academy of Sciences of the United States of America **70**(12 Pt 1-2): 3581-3584.
18. Grenet, K., D. Guillemot, V. Jarlier, B. Moreau, S. Dubourdieu, R. Ruimy, L. Armand-Lefevre, P. Bau and A. Andremont (2004). "Antibacterial Resistance, Wayampis Amerindians, French Guyana." Emerging Infectious Diseases **10**(6): 1150-1153.
19. Harada, K., T. Asai, A. Kojima, C. Oda, K. Ishihara and T. Takahashi (2005). "Antimicrobial susceptibility of pathogenic Escherichia coli isolated from sick cattle and pigs in Japan." J Vet Med Sci **67**(10): 999-1003.
20. Heather, J. M. and B. Chain (2016). "The sequence of sequencers: The history of sequencing DNA." Genomics **107**(1): 1-8.
21. Hjort, K., H. Nicoloff and D. I. Andersson (2016). "Unstable tandem gene amplification generates heteroresistance (variation in resistance within a population) to colistin in Salmonella enterica." Molecular Microbiology **102**(2): 274-289.
22. Hughes, D. and D. I. Andersson (2017). "Environmental and genetic modulation of the phenotypic expression of antibiotic resistance." FEMS Microbiology Reviews **41**(3): 374-391.
23. Jayol, A., P. Nordmann, A. Brink and L. Poirel (2015). "Heteroresistance to Colistin in Klebsiella pneumoniae Associated with Alterations in the PhoPQ Regulatory System." Antimicrobial Agents and Chemotherapy **59**(5): 2780-2784.
24. Katsunuma, Y., M. Hanazumi, H. Fujisaki, H. Minato, Y. Hashimoto and C. Yonemochi (2007). "Associations between the use of antimicrobial agents for growth promotion and the occurrence of antimicrobial-resistant Escherichia coli and enterococci in the feces of livestock and livestock farmers in Japan." Journal of General and Applied Microbiology **53**(5): 273-279.
25. Kau, A. L., P. P. Ahern, N. W. Griffin, A. L. Goodman and J. I. Gordon (2011). "Human nutrition, the gut microbiome, and immune system: envisioning the future." Nature **474**(7351): 327-336.

26. Lee, S. Y., J. H. Shin, K. Lee, M. Y. Joo, K. H. Park, M. G. Shin, S. P. Suh, D. W. Ryang and S. H. Kim (2013). "Comparison of the Vitek 2, MicroScan, and Etest Methods with the Agar Dilution Method in Assessing Colistin Susceptibility of Bloodstream Isolates of Acinetobacter Species from a Korean University Hospital." Journal of Clinical Microbiology **51**(6): 1924-1926.
27. Li, J., R. L. Nation, J. D. Turnidge, R. W. Milne, K. Coulthard, C. R. Rayner and D. L. Paterson (2006). "Colistin: the re-emerging antibiotic for multidrug-resistant Gram-negative bacterial infections." Lancet Infect Dis **6**(9): 589-601.
28. Littmann, J. and A. M. Viens (2015). "The Ethical Significance of Antimicrobial Resistance." Public Health Ethics **8**(3): 209-224.
29. Liu, Y.-Y., Y. Wang, T. R. Walsh, L.-X. Yi, R. Zhang, J. Spencer, Y. Doi, G. Tian, B. Dong, X. Huang, L.-F. Yu, D. Gu, H. Ren, X. Chen, L. Lv, D. He, H. Zhou, Z. Liang, J.-H. Liu and J. Shen (2015). "Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study." The Lancet Infectious Diseases **16**(2): 161-168.
30. Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y.-J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J.-B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley and J. M. Rothberg (2005). "Genome sequencing in microfabricated high-density picolitre reactors." Nature **437**: 376.
31. Martis, N., S. Leroy and V. Blanc "Colistin in multi-drug resistant Pseudomonas aeruginosa blood-stream infections." Journal of Infection **69**(1): 1-12.
32. Mathers, A. J., H. L. Cox, B. Kitchel, H. Bonatti, A. K. C. Brassinga, J. Carroll, W. M. Scheld, K. C. Hazen and C. D. Sifri (2011). "Molecular Dissection of an Outbreak of Carbapenem-Resistant Enterobacteriaceae Reveals Intergenus KPC Carbapenemase Transmission through a Promiscuous Plasmid." mBio **2**(6): e00204-00211.
33. Meletis, G., E. Oustas, C. Botziori, E. Kakasi and A. Koteli (2015). "Containment of carbapenem resistance rates of Klebsiella pneumoniae and Acinetobacter baumannii in a Greek hospital with a concomitant increase in colistin, gentamicin and tigecycline resistance." New Microbiol **38**(3): 417-421.
34. Meletis, G., E. Tzampaz, E. Sianou, I. Tzavaras and D. Sofianou (2011). "Colistin heteroresistance in carbapenemase-producing Klebsiella pneumoniae." Journal of Antimicrobial Chemotherapy **66**(4): 946-947.
35. Miller, J. M. (2013). "Whole-genome mapping: a new paradigm in strain-typing technology." J Clin Microbiol **51**(4): 1066-1070.
36. Moffatt, J. H., M. Harper, P. Harrison, J. D. F. Hale, E. Vinogradov, T. Seemann, R. Henry, B. Crane, F. S. Michael, A. D. Cox, B. Adler, R. L. Nation, J. Li and J. D. Boyce (2010). "Colistin Resistance in Acinetobacter baumannii Is Mediated by Complete Loss of Lipopolysaccharide Production." Antimicrobial Agents and Chemotherapy **54**(12): 4971-

4977.
37. Moon, K. and S. Gottesman (2009). "A PhoQ/P-regulated small RNA regulates sensitivity of Escherichia coli to antimicrobial peptides." Molecular Microbiology **74**(6): 1314-1330.
38. Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer and B. Wold (2008). "Mapping and quantifying mammalian transcriptomes by RNA-Seq." Nature Methods **5**: 621.
39. Munita, J. M. and C. A. Arias (2016). "Mechanisms of Antibiotic Resistance." Microbiology spectrum **4**(2): 10.1128/microbiolspec.VMBF-0016-2015.
40. Nation, R. L. and J. Li (2009). "Colistin in the 21(st) Century." Current opinion in infectious diseases **22**(6): 535-543.
41. Nodari, C. S., V. B. Ribeiro and A. L. Barth (2015). "Imipenem heteroresistance: high prevalence among EnterobacteriaceaeKlebsiella pneumoniae carbapenemase producers." Journal of Medical Microbiology **64**(1): 124-126.
42. Olaitan, A. O. and J. Li (2016). "Emergence of polymyxin resistance in Gram-negative bacteria." International Journal of Antimicrobial Agents **48**(6): 581-582.
43. Olaitan, A. O., S. Morand and J. M. Rolain (2014). "Mechanisms of polymyxin resistance: acquired and intrinsic resistance in bacteria." Front Microbiol **5**: 643.
44. Olaitan, A. O., B. Thongmalayvong, K. Akkhavong, S. Somphavong, P. Paboriboune, S. Khounsy, S. Morand and J.-M. Rolain (2015). "Clonal transmission of a colistin-resistant Escherichia coli from a domesticated pig to a human in Laos." Journal of Antimicrobial Chemotherapy **70**(12): 3402-3404.
45. Pitout, J. D. D. and K. B. Laupland (2008). "Extended-spectrum β-lactamase-producing Enterobacteriaceae: an emerging public-health concern." The Lancet Infectious Diseases **8**(3): 159-166.
46. Poirel, L., A. Jayol, S. Bontron, M.-V. Villegas, M. Ozdamar, S. Türkoglu and P. Nordmann (2015). "The mgrB gene as a key target for acquired resistance to colistin in Klebsiella pneumoniae." Journal of Antimicrobial Chemotherapy **70**(1): 75-80.
47. Raz, T., P. Kapranov, D. Lipson, S. Letovsky, P. M. Milos and J. F. Thompson (2011). "Protocol Dependence of Sequencing-Based Gene Expression Measurements." PLOS ONE **6**(5): e19287.
48. Rebelo, A. R., V. Bortolaia, J. S. Kjeldgaard, S. K. Pedersen, P. Leekitcharoenphon, I. M. Hansen, B. Guerra, B. Malorny, M. Borowiak, J. A. Hammerl, A. Battisti, A. Franco, P. Alba, A. Perrin-Guyomard, S. A. Granier, C. De Frutos Escobar, S. Malhotra-Kumar, L. Villa, A. Carattoli and R. S. Hendriksen (2018). "Multiplex PCR for detection of plasmid-mediated colistin resistance determinants, mcr-1, mcr-2, mcr-3, mcr-4 and mcr-5 for surveillance purposes." Eurosurveillance **23**(6): 17-00672.
49. Reynolds, C. M., S. R. Kalb, R. J. Cotter and C. R. Raetz (2005). "A phosphoethanolamine transferase specific for the outer 3-deoxy-D-manno-octulosonic acid residue of Escherichia coli lipopolysaccharide. Identification of the eptB gene and Ca2+ hypersensitivity of an eptB deletion mutant." J Biol Chem **280**(22): 21202-21211.
50. Rhouma, M., F. Beaudry, W. Thériault and A. Letellier (2016). "Colistin in Pig Production: Chemistry, Mechanism of Antibacterial Action, Microbial Resistance Emergence, and One Health Perspectives." Frontiers in Microbiology **7**(1789).
51. Sader, H. S., P. R. Rhomberg, R. K. Flamm and R. N. Jones (2012). "Use of a surfactant

(polysorbate 80) to improve MIC susceptibility testing results for polymyxin B and colistin." Diagnostic Microbiology and Infectious Disease **74**(4): 412-414.
52. Saidel-Odes, L. and A. Borer (2014). "Limiting and controlling carbapenem-resistant Klebsiella pneumoniae." Infection and Drug Resistance **7**: 9-14.
53. Sanger, F., S. Nicklen and A. R. Coulson (1977). "DNA sequencing with chain-terminating inhibitors." Proceedings of the National Academy of Sciences of the United States of America **74**(12): 5463-5467.
54. Sommer, M. O. A., G. Dantas and G. M. Church (2009). "Functional characterization of the antibiotic resistance reservoir in the human microflora." Science (New York, N.Y.) 325 (5944): 1128-1131.
55. Srinivasan, V. B. and G. Rajamohan (2013). "KpnEF, a new member of the Klebsiella pneumoniae cell envelope stress response regulon, is an SMR-type efflux pump involved in broad-spectrum antimicrobial resistance." Antimicrob Agents Chemother **57**(9): 4449-4462.
56. Stefani, S., D. R. Chung, J. A. Lindsay, A. W. Friedrich, A. M. Kearns, H. Westh and F. M. MacKenzie (2012). "Meticillin-resistant Staphylococcus aureus (MRSA): global epidemiology and harmonisation of typing methods." International Journal of Antimicrobial Agents **39**(4): 273-282.
57. Stefani, S. and P. E. Varaldo (2003). "Epidemiology of methicillin-resistant staphylococci in Europe." Clin Microbiol Infect **9**.
58. Turlej-Rogacka, A., B. B. Xavier, L. Janssens, C. Lammens, O. Zarkotou, S. Pournaras, H. Goossens and S. Malhotra-Kumar (2017). "Evaluation of colistin stability in agar and comparison of four methods for MIC testing of colistin." European Journal of Clinical Microbiology & Infectious Diseases.
59. Walsh, T. R. and M. A. Toleman (2012). "The emergence of pan-resistant Gram-negative pathogens merits a rapid global political response." J Antimicrob Chemother **67**(1): 1-3.
60. Wand, M. E., L. J. Bock, L. C. Bonney and J. M. Sutton (2015). "Retention of virulence following adaptation to colistin in Acinetobacter baumannii reflects the mechanism of resistance." Journal of Antimicrobial Chemotherapy.
61. Wang, G.-Q., C.-M. Wu, X.-D. Du, Z.-Q. Shen, L.-H. Song, X. Chen and J.-Z. Shen (2008). "Characterization of integrons-mediated antimicrobial resistance among Escherichia coli strains isolated from bovine mastitis." Veterinary Microbiology **127**(1): 73-78.
62. Wang, Z., M. Gerstein and M. Snyder (2009). "RNA-Seq: a revolutionary tool for transcriptomics." Nature reviews. Genetics **10**(1): 57-63.
63. WHO (2014). Antimicrobial resistance: global report on surveillance
64. Willner, D. and P. Hugenholtz (2013). "Metagenomics and Community Profiling: Culture-Independent Techniques in the Clinical Laboratory." Clinical Microbiology Newsletter **35**(1): 1-9.
65. Wright, G. D. (2010). "Q&A: Antibiotic resistance: where does it come from and what can we do about it?" BMC Biology **8**(1): 123

1.5 SCOPE OF THE THESIS

Antimicrobial resistance (AMR) threatens to reverse many of the enormous strides modern medicine has made over the last century. Some bacterial species such as *Mycobacterium tuberculosis* and carbapenemase-producing colistin-resistant Gram-negative bacteria have developed resistance against all useful human antibiotics. The golden age of new antibiotic discovery has stalled, with few new genuinely novel antibiotic classes in the developmental pipeline. The natural ability of bacteria to adapt caused the rapid emergence of resistance to multiple antibiotics, and along with their ability to acquire resistance genes through horizontal gene transfer (HGT) resulted in the spread of these genes among different bacterial species. Therefore, genetic tools are instrumental in tracing and identifying novel resistant genetic determinants rapidly. Although polymerase chain reaction (PCR) and hybridisation techniques are available to identify antibiotic resistance, the time-to-results of these techniques is still from two days to weeks, depending on the bacterial species. Moreover, the information which we obtain from these techniques is limited and too specific. However, knowledge of the molecular mechanisms of resistance and the origin of the resistance genes will help us to develop and implement better control strategies against the spread of these genes. Fortunately, substantial technological progress was made of sequencing technologies, such as Whole Genome Mapping (WGM) and Next-Generation Sequencing (NGS). These technologies offer high resolution and comprehensive information, and both techniques complement each other.

In this thesis, we first assessed, improved and developed genetic tools to analyse sequencing information rapidly. We then applied these tools to study the multi-drug-resistant bacterial genetic characteristics and novel resistant mechanisms of the antibiotic colistin. This antibiotic is one of the last-line therapeutic options available to treat multi-drug resistant Gram-negative bacterial infections. Therefore, understanding the global emergence of resistance to colistin among *K. pneumoniae* and *E. coli* clones is of utmost importance. Finally, we applied WGS and metagenomics tools to study HGT and the resistome in the gut.

The overall aims of the thesis were to develop, validate and apply bacterial whole genome and metagenomic analysis tools in clinical microbiology (**Aim 1**), and to comprehensively dissect novel colistin resistance mechanisms in *E. coli* and *K. pneumoniae* using sequencing and an ‘omics’ based strategy (**Aim 2**).

The following specific aims and studies were defined to achieve the two overall aims of the thesis

1.1 To review and identify limitations and advantages of existing antibiotic resistance gene data resources.

1.2 To develop a fully automated bioinformatics pipeline to facilitate faster interpretation of bacterial whole genome sequences.

1.3 To identify misassemblies of short read sequencing reads and improve genome assembly using whole genome mapping.

1.4 To study clonal diversity, single-clone dynamics and genomic rearrangements in human bacterial pathogens using whole genome mapping and sequencing

1.5 To detect potential *in vivo* inter-genus transfer of an ESBL-harbouring plasmid using whole genome sequencing.

1.6 To understand the impact of moxifloxacin use on the healthy human gut flora and the potential of FMT to enable a faster recovery of gut dysbiosis: a proof-of-concept metagenomics study.

The above topics are covered in **chapters 2 & 4**

2.1 To identify molecular mechanisms of colistin resistance in *E. coli* using whole genome sequencing.

2.2 To elucidate molecular mechanisms of colistin resistance in *K. pneumoniae* using omics-based strategy.

The above topics are covered in **chapter 3**

CHAPTER II

DEVELOPMENT OF ANTIBIOTIC RESISTANCE GENE DATABASES AND BACTERIAL WHOLE GENOME ANALYSIS TOOLS IN CLINICAL MICROBIOLOGY

Chapter II is further divided into the following sub-chapters:

2.1 Consolidating and exploring antibiotic resistance gene data resources. (**Xavier BB,** Das AJ, Cochrane G, De Ganck S, Kumar-Singh S, Aarestrup FM, Goossens H, Malhotra-Kumar S. *Journal of Clinical Microbiology, 54*(4), 851–859.) **IF:3.993**

2.2 BacPipe: A rapid, user-friendly whole genome sequencing pipeline for clinical diagnostic bacteriology and outbreak detection. (**Xavier BB***, Mysara M*, Bolzan M, Lammens C, Kumar-Singh S, Goossens H, Malhotra-Kumar S. (*In revision*)

2.3 Whole genome mapping as a utility tool to study genomic stability and variations in human pathogens. (Sabirova JS, **Xavier BB,** Ieven M, Goossens H, Malhotra-Kumar S. *BMC Research Notes*. 2014; 7: 704)

2.4 Employing whole genome mapping for optimal *de novo* assembly of bacterial genomes (**Xavier BB**, Sabirova JS, Pieter M, Hernalsteen, J-P, Greve de H, Goossens H, Malhotra-Kumar S. *BMC Research Notes* 2014**7**:484)

2.5 Whole-genome typing and characterization of blaVIM19-harbouring ST383 *Klebsiella pneumoniae* by PFGE, whole-genome mapping and WGS. (Sabirova JS*, **Xavier BB***, Coppens J, Zarkotou O, Lammens C, Janssens L, Burggrave R, Wagner T, Goossens, H, Malhotra-Kumar S. *Journal of Antimicrobial Chemotherapy*, Volume 71, Issue 6, 1 June 2016, Pages 1501–1509) **IF:5.313**

2.6 Comparison of typing methods for hospital-associated *Acinetobacter baumannii* clones isolated in Greece and India. (**Xavier BB,** Coppens J, Vanjari L, Sabirova JS, Lammens C, Vemu L, Wagner T, Pournaras S, Goossens H, Malhotra-Kumar S,) (*In preparation*)

2.1 CONSOLIDATING AND EXPLORING ANTIBIOTIC RESISTANCE GENE DATA RESOURCES

Xavier BB[1], Das AJ[1], Cochrane G[2], De Ganck S[1], Kumar-Singh S[1,3], Aarestrup FM[4], Goossens H[1], Malhotra-Kumar S[1]

[1]Laboratory of Medical Microbiology, Vaccine & Infectious Disease Institute, Universiteit Antwerpen, Antwerp, Belgium. [2]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Cambridge, United Kingdom. [3]Laboratory of Molecular Pathology, Cell Biology and Histology, Faculty of Medicine and Health Sciences, Universiteit Antwerpen, Antwerp, Belgium. [4]National Food Institute, Technical University of Denmark, Lyngby, Denmark.

Abstract

The unrestricted use of antibiotics has resulted in the rapid acquisition of antibiotic resistance (AR) and spread of multi-drug-resistant (MDR) bacterial pathogens. With the advent of next generation sequencing technologies and their application in understanding MDR pathogen dynamics, it has become imperative to unify AR gene data resources for easy accessibility for researchers. However, due to the absence of a centralized platform for AR gene resources, availability, consistency, and accuracy of information vary considerably across different databases. In this article, we sought to explore existing AR gene data resources in order to make them more visible to the clinical microbiology community, to identify their limitations, and to propose potential solutions.

2.1.1 Introduction

Over the years, antibiotics have vastly benefitted human and animal health in combatting bacterial infections. Apart from being widely used in clinical practice, antibiotics are also employed in agriculture, aquaculture and intensive animal farming either as prophylactic agents or for therapeutic purposes (Aarestrup 2015; Burns, et al. 2014; Cabello 2006). The unrestrained use of antibiotics, however, has resulted in higher frequency of resistant human pathogens (Davies and Davies 2010). Acquisition of antibiotic resistance can result from a variety of genomic alterations, for instance, single nucleotide mutations, large genomic changes such as insertions or deletions, chromosomal rearrangements, gene duplications, and importantly, factors that have facilitated their rampant spread i.e., carriage on plasmids and other mobile genetic elements (MGEs) including integrons and transposons (Mwangi, et al. 2007; Ploy, et al. 2000). A fitting example is the recently reported *mcr-1* gene that has been linked to colistin-resistance in humans and animals (Hasman, et al. 2015; Liu, et al. 2015a; Malhotra-Kumar, et al. 2016a; Malhotra-Kumar, et al. 2016b), and has been found to be associated with at least three different plasmids up till now (Hasman, et al. 2015; Liu, et al. 2015a; Malhotra-Kumar, et al. 2016a).

In the last decade, emergence of multi drug-resistant (MDR) bacteria that harbor multiple antibiotic resistance mechanisms/genes has severely limited therapeutic options (Davies and Davies 2010). Common examples are some of the most important Gram-negative human pathogens such as *Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa* and *Escherichia coli* that harbor MGEs carrying genes encoding enzymes like extended-spectrum beta-lactamases (ESBLs) that can hydrolyze penicillins, cephalosporins and monobactams, along with aminoglycoside-modifying enzymes, and the Qnr protection proteins that confer resistance to the fluoroquinolones (Carattoli 2013; Vervoort, et al. 2014a). Thus, a single conjugation event involving such MGEs is enough to transform an antibiotic sensitive pathogen to a MDR organism that can potentially cause infections that are non-treatable by the current antibiotic arsenal (Alekshun and Levy 2007). Extremely worrisome are the rising rates of resistance to carbapenems, one of the most important last-line antibiotics available to us (Pournaras, et al. 2010a). Carbapenemases such as bla_{VIM}, bla_{IMP}, bla_{KPC}, bla_{NDM} are also primarily MGE-encoded and coupled with the high and varying antibiotic selection pressure in hospitals has led to a rapid evolution and spread of these enzyme with varying substrate specificities (Queenan and Bush 2007). Currently, more than 40 bla_{VIM} and 10 bla_{NDM} variants are known (Canton, et al. 2012). On the other

hand, in important Gram-positive pathogens such as *Staphylococcus aureus*, resistance to beta-lactam antibiotics is mediated by the 'staphylococcal cassette chromosome' (SCC*mec*) MGEs that integrate in a site-specific manner in the *Staphylococcus* genome (Malachowa and DeLeo 2010). Interestingly, the marked differences in antibiotic resistance profiles of community- (CA) and hospital-associated (HA-) of methicillin-resistant S. aureus (MRSA) can be largely attributed to the kind of SCC*mec* element harboured by these strains. CA-MRSA harbor the smaller SCC*mec*IV, V, or VII elements, whereas HA-MRSA typically contain the larger SCC*mec*I, II, III, VI, or VIII elements that encode multiple resistance determinants in addition to the gene encoding beta-lactam resistance, *mecA* (Malachowa and DeLeo 2010).While these examples highlight the complex combinations of emerging AR mechanisms and genes, they represent only the tip of the proverbial iceberg. The application of next-generation sequencing (NGS) technology to the study of pathogen genomes as well as to soil-, marine-, and human-associated metagenomes has given us unprecedented insights into unknown reservoirs and novel AR genes (Arumugam, et al. 2011; Canica, et al. 2015; Didelot, et al. 2012a; Vervoort, et al. 2015). Currently, a wealth of information with respect to AR genes is available online in AR gene databases (Table 2.1.1, Figure 2.1.1). As costs of sequencing are steadily decreasing and response times getting shorter, its utility as a tool for tracking MDR pathogens in real time for routine hospital epidemiology or as an early warning system for outbreak detection is steadily increasing. Application of NGS to routine clinical microbiology and diagnostics will be especially useful in simplifying the technical algorithms utilized for typing and for antibiotic resistance detection. Currently, majority of the routine microbiology laboratories still screen for MDR based on phenotypic susceptibility testing, which is not only subject to guidelines and breakpoints but is also time-consuming as it depends on pathogen growth. Previous NGS-based studies have demonstrated high concordance between *in silico* predicted and phenotypic antimicrobial susceptibility (Zankari, et al. 2012a; Zankari, et al. 2013). Nonetheless, sequence-based predictions of phenotypic resistance for clinical purposes need to be made with caution. Firstly, in contrast to phenotypic testing, sequencing data yields information only on resistance to antibiotics but not on susceptibility to others. Also, absence of a resistance gene does not preclude sensitivity to that antibiotic as any new resistances that are not in the utilized AR gene database might have been missed. On the other hand, sequence-based predictions might potentially identify a gene that leads to resistance during treatment although is not expressed during sensitivity testing under specific growth conditions. Thus, while application of NGS to antimicrobial susceptibility testing can result in a more efficient workflow, the large datasets generated here will be heavily reliant on the available AR gene data resources for quality reference data and interpretation making it imperative that the latter are well curated, up-to-date and comprehensive.

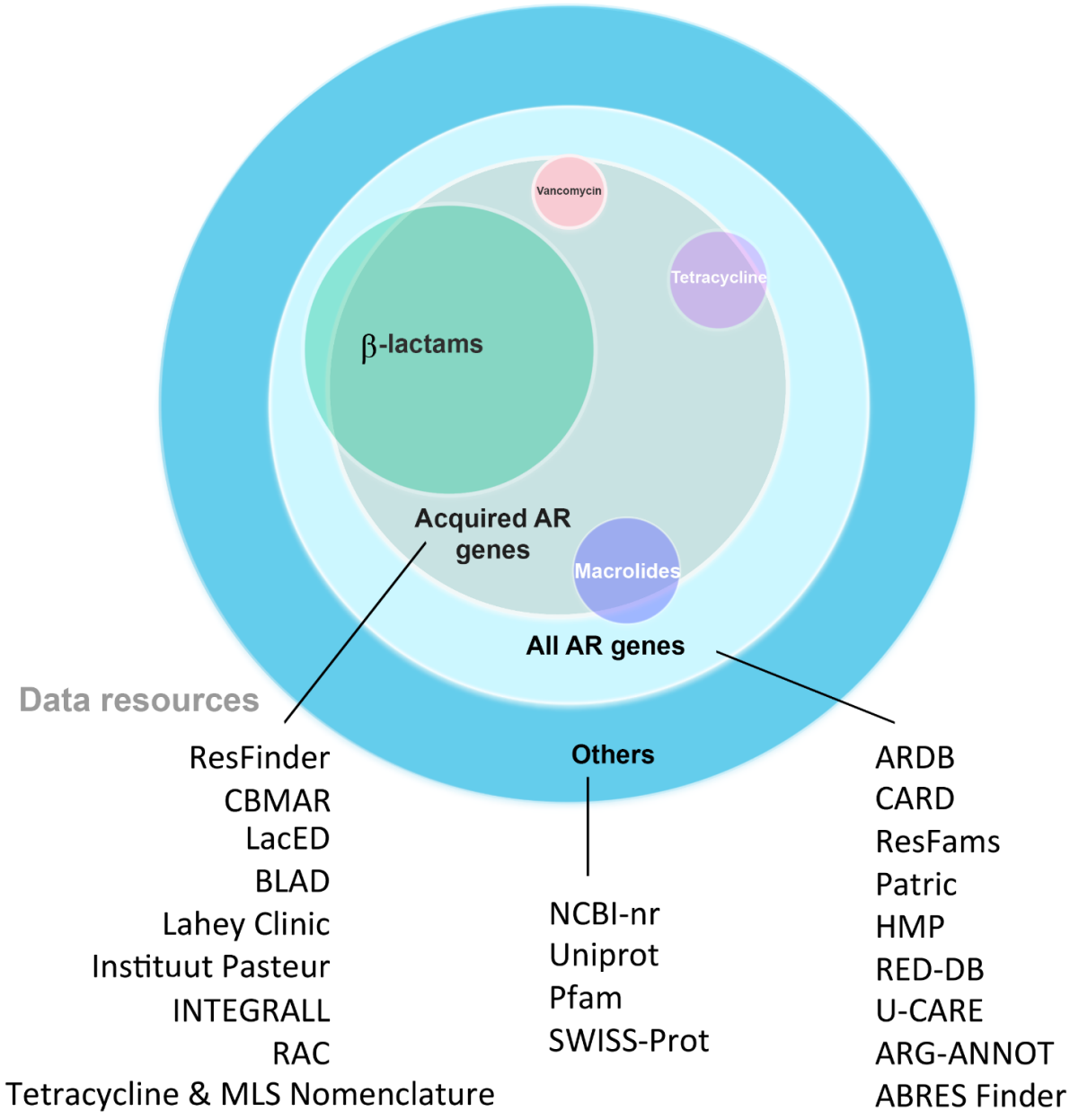


Figure 2.1.1 Non-exhaustive overview of available data resources in light of the functional classifications of resistance genes targeting different antibiotic classes. For instance, β-lactams refers to all beta-lactamase genes including ESBLs and carbapenemases. Colours indicate the subset of genes represented in the databases. Not drawn to scale.

Table 2.1.1 Characteristics of available AR gene data resources

No.	Database/Repository	AR Gene Spectrum	Functionality & Features	Last update*	References
1	ARDB	All AR genes	Webtool, BLASTp, BLASTn	39995	(Liu and Pop 2009)
2	CARD	All AR genes	Webtool, BLASTp, BLASTn, Gene ontology, Gene identifier V2, Annotation	41730	(McArthur, et al. 2013a)
3	ResFinder	All AR genes (except chromosome-specific)	Webtool, BLASTn	42156	(Zankari, et al. 2012a)
4	LacED	β-lactamases	Webtool, BLASTp, Clustalw	-	(Thai, et al. 2009)
5	ResFams	All AR genes	BLASTp, Local BLAST, HMM profile	42005	(Gibson, et al. 2014)
6	Patric	All AR genes	Webtool link to CARD and ARDB	42339	(Wattam, et al. 2014)
7	HMP	Human body site-specific study resources	Webtool, BLASTp, BLASTn	40848	(Human Microbiome Project 2012)
8	RED-DB	All AR genes	BLASTn, BLASTp	-	-
9	U-CARE	Organism-specific (*E.coli*)	BLASTp	-	(Saha, et al. 2015)
10	ARG-ANNOT	All AR genes	BLAST, BioEdit V7.25, Annotation	-	(Gupta, et al. 2014)
11	BLAD	β-lactamases	Webtool	-	(Danishuddin, et al. 2013)
12	CBMAR	β-lactamases	Webtool, BLASTn, BLASTp,Clustalw, MEME/MAST	41883	(Srivastava, et al. 2014)
13	Lahey Clinic	β-lactamases	ß-lactamase classification and assigning allelic number**	42064	-
14	Instituut Pasteur	OKP, LEN, OXY	MLST database with additional information on specific β-lactamases	42217	-
15	Tetracycline & MLS Nomenclature	Tetacycline and macrolide AR genes	Information on resistance mechanisms and nomenclature	42156	-
16	ABRES Finder	All AR genes	Links to external databases	-	-
17	INTEGRALL	Integron types and genetic context of AR genes	Webtool, BLASTn	42217	(Moura, et al. 2009)
18	RAC	Genetic context of AR genes	Webtool, Resistance gene cassette annotation	-	(Tsafnat, et al. 2011a)
19	Mvirdb	Virulence and toxin factors	BLASTn, BLASTp, link to ARGODB for AR genes	41730	(Zhou, et al. 2007)

*Based on information available on their respective websites and/or publications

**Moved to http://www.ncbi.nlm.nih.gov/pathogens/submit_beta_lactamase/

We reviewed the currently available AR gene data resources with the aim of making them more visible to the clinical microbiology community, particularly emphasizing on regular updates and easy accessibility to resources that include metadata from published literature. Additionally, we also demonstrated test runs on 4 available databases using in-house and publicly available data. This exercise revealed inconsistent search results, which we discuss in detail and propose two complementary approaches that call for a combined effort in addressing this issue.

2.1.2 Antibiotic resistance gene databases

AR gene data resources are online platforms that offer AR-related reference data in support of prediction of resistome and gene-based antibiograms along with online bioinformatics tools for sequence comparisons, alignment and annotation. These resources accept user nucleotide or protein sequences as queries and return predictions of their AR gene content, often with confidence-related statistics, annotation and onward links to external related resources. We consider first several generalist resources and then move on to AR-focused resources with an anecdotal commentary and a tabulated summary of each open access data resource's key characteristics.

The NCBI non-redundant (NCBI-nr) data set (http://www.ncbi.nlm.nih.gov/nuccore) represents one of the largest publicly available generalist data resources that include (AR) genes and associated information. In some cases, however, search results obtained with NCBI-nr might not be specific in terms of gene sub-types resulting in multiple hits with similar level of identity and query coverage. It is important to note that results may vary depending on whether the query sequence is a part of the gene or also includes regions flanking the gene. Thus, additional manual verifications may be required for accurate predictions.

Popular further generalist options that relate to protein-level similarity are the UniProt Knowledgebase (UniProtB; http://www.uniprot.org/) and the Protein Families Database (Pfam; http://pfam.xfam.org/), which together provide information on protein sequences, functional annotations and conserved protein families (Ashburner, et al. 2000; Eilbeck, et al. 2005). UniProtKB offers an exhaustive collection of protein annotation, cross-references and literature-derived annotations, while Pfam offers conserved protein families. Pfam uses the profile HMM software, HMMER3 (http://hmmer.janelia.org/), in order to identify and build Hidden Markov Models (HMM) of protein families. These generalist resources are of value not only as they are

comprehensive for publicly available data but also since they serve to feed data to more specialist AR-gene data resources (Table 1).

Antibiotic Resistance Genes Database (ARDB; http://ardb.cbcb.umd.edu/index.html), a manually curated specialist AR-gene database, appeared very promising at the time of its introduction, combining information from several existing resources, offering AR gene (sub)types and ontology information. At launch, it comprised 13293 genes, 377 types, 257 antibiotics, 632 genomes, 933 species and 124 genera to which were applied a two-step filtering of vector sequences, synthetic constructs and redundant genes and then removal of incomplete sequences, yielding 4545 antibiotic resistance gene sequences (Liu and Pop 2009). The resource features various tools for annotation and comparison of genes and genomes. Furthermore, a tool for mutation detection is also provided (Liu and Pop 2009). The site allows upload of data as a single gene or in a batch mode for multiple genes or protein sequences. Though the site is functional and user-friendly, the major concern is with the updates of the database as, according to the database statistics, the last update was in July 2009. Following the last update, the database reported 23137 genes, 380 types, 249 antibiotics, 632 genomes, 1737 species, 267 genera including information on 2881 vectors (vehicles for transmitting genetic material/ genes from one organism to another) including plasmids.

The Comprehensive Antibiotic Gene Database (CARD; http://arpcard.mcmaster.ca/) was first introduced with a β-lactamase ontology feature. Since its introduction in 2009, regular updates have been announced with the most recent one in April 2014. The database facilitates access to exhaustive knowledge resources regarding antibiotic resistance genes and their associated proteins that additionally include antibiotics and corresponding targets. CARD presents a well-developed AR Ontology (ARO) platform that has been expanded from the initial efforts of the ARDB. The ARO allows efficient investigation of molecular data by including classification of AR genes, functional ontology information, SNP details for resistance genes, extensive microarray targets, Gene Ontology (GO), Sequence Ontology (SO) and Infectious Disease Ontology (IDO) (Ashburner, et al. 2000; Eilbeck, et al. 2005; Goldfain, et al. 2010; McArthur, et al. 2013a). Additionally, CARD also features a graphical web tool called Resistance Gene Identifier (RGI), Version 2, in October 2011 for annotation of query sequences. As of the latest update, CARD provides 3008 genes tagged specifically for antibiotic resistance and 4120 genes with AR-related functions. It permits query sequence upload in both batch mode (limited to 20 Mb) and as single sequences. The graphical interface was found to be user friendly and highly descriptive with functional based classification of AR genes.

ResFinder version 2.1(https://cge.cbs.dtu.dk//services/ResFinder/), most recently updated in July 2015, is a database that provides exhaustive information on AR genes from sequenced or

partially sequenced bacterial isolates. Information of acquired resistance genes through horizontal gene transfer can be obtained here. ResFinder provides not only up-to-date information on AR genes but also offers enhanced flexibility in the user interface, which helps minimize unspecific hits. The current version of ResFinder allows a user to set the identity and length coverage thresholds as low as 30% and 20% respectively (Zankari 2014). One of the major advantages of Resfinder over other tools is that it accepts NGS raw reads including *de novo* assembled contigs without any limitations on size or sequence length. However, one of the limitations is that the information currently contained in the database is specific for acquired genes and therefore does not include AR mechanisms mediated by chromosomal mutations. Furthermore, the database accepts only nucleotide (and not protein) sequence queries for comparison.

The Lactamase Engineering Database (LacED; http://www.laced.uni-stuttgart.de/) provides systematic analysis and annotation of sequences that helps compare new entries to already existing ones. Furthermore, the database provides integrated tools for sequence comparison and multiple sequence alignments such as Basic Local Alignment Search Tool (BLAST) and ClustalW, respectively. LacED database, however, specializes in information related to mutations, sequences and structures of TEM and SHV, β –lactamases (Thai, et al. 2009).

ResFams (http://www.dantaslab.org/resfams) is a recently established resource for protein families, which are linked to their HMMs associated with AR function. It aims at providing accurate identification and annotation of AR genes. With the information provided, one can also get an overview of the ecology and evolution of the resistant pathogens. The ResFams platform is specifically targeted towards AR gene families and their HMMs, which are further associated with functional metagenomic datasets acquired from various sources such as soil and human feces as well as from 6000 sequenced microbial isolate genomes across diverse phylogenies and habitats. This data was then utilized to derive 166 HMM profiles comprising the major AR gene classes (Gibson, et al. 2014). The authors emphasize that for resistome analysis this HMM-based approach is superior to that of BLAST-based pairwise sequence alignment to AR-specific databases that are biased towards human-associated organisms and vastly underestimate the potential impact of environmental resistance reservoirs on AR in pathogens (Gibson, et al. 2014). The authors demonstrated this by comparing ResFams HMMs with the BLAST-based ARDB and CARD databases for their ability to predict AR function, and showed that 64% of AR proteins identified using ResFams in both the soil and the human gut microbiota were not detected by BLAST. This increased sensitivity over other AR data resources is expected with HMM-based analysis. HMMs are specific models that are constructed based on observed sequence variation sampled across gene

or protein families and capture nuanced positional variability for the family. Search of query sequence using these models returns resulting matches that can be distant, and not detectable using sequence-based matching such as in BLAST, but represent valid homologous genes or proteins. However, HMM approaches come at a computational cost, in particular where models are used in scans of high volume whole genome shotgun data. The implication of this cost for ResFams is that the user must provide local computational resources in order to run HMM-based searches. These HMMER tools need to be installed locally under LINUX/UNIX, and the results appear in a tabular form without a graphical interface.

Antibiotic Resistance Gene ANNOTation (ARG-ANNOT; http://en.mediterranee-infection.com/article.php?laref=283%26titre=arg-annot-) is a rapid bioinformatics tool that is used in identifying putative new AR genes in bacterial genomes. ARG-ANNOT also provides data relating to point mutations. In another study, the tool has been tested for its enhanced sensitivity and specificity for both complete and partial gene sequences (Gupta, et al. 2014).

The Pathosystems Resource Integrated System (Patric; https: //www. Patricbrc.org/portal/ portal/ patric/AntibioticResistance) provides a platform for genome assembly, protein family comparisons, genome annotations, meta-data information such as AR and pathway comparisons. Patric collects public genome data and currently provides AR data from ARDB and CARD (Wattam, et al. 2014).

The Human Microbiome Project (http://hmpdacc.org/HMGOI/), in its efforts to characterize the human microbiome, developed a reference set of 3000 microbial isolate genomes. The HMP also provides a large collection of AR genes (Human Microbiome Project 2012).

Resistance Determinants Database (RED-DB; http://www.fibim.unisi.it/REDDB/) is a non-redundant collection of resistance genes from various nucleotide sequence databases. One can easily look up the database based on the cluster or reference gene names. User-friendly Comprehensive Antibiotic Resistance Repository of *Escherichia coli* (U-CARE; http://www.e-bioinformatics.net/ucare/), is a manually-curated resource that provides *E.coli* related AR information, including information from 52 antibiotics, 107 resistance genes and associated information of transcription factors and SNPs (Saha, et al. 2015).

β-lactamase Database (BLAD; http://www.blad.co.in) includes resistance patterns of all classes of β -lactamases collected from published data, NCBI and the crystal structure of proteins from the Protein Data Bank (PDB; http://www.rcsb.org/pdb/home/home.do). BLAD allows sequence comparison using BLAST search tool. Apart from facilitating information regarding the 3D structure and physiochemical properties of bound ligands, BLAD also provides links to the popular nucleotide and protein databases (Danishuddin, et al. 2013). The resource specializes in β-lactamase related information.

Comprehensive β-lactamase molecular annotation resource. (CBMAR; http://14.139.227.92/mkumar/lactamasedb) is a recently established AR gene resource, which provides a fully interactive environment for data access to exhaustive range of β-lactamase resources (Srivastava, et al. 2014). It provides extensive metadata along with detailed molecular and biochemical information, which could reveal further insights into novel β –lactamases. CBMAR also features tools such as BLAST and searches for family-specific fingerprints employing MAST (Motif alignment and search tool; http://meme-suite.org/tools/mast). Information related to protein, nucleotide, protein structures, alignments, mutation profiles and phylogenetic trees can be downloaded from the database. According to site statistics, the most recent update was performed on 9 September 2014.

Lahey (http://www.lahey.org/studies/) is a conventional database/repository for β -lactamase classification and amino acid sequences for *bla*$_{TEM}$, *bla*$_{SHV}$ and *bla*$_{OXA}$ extended-spectrum and other inhibitor-resistant enzymes. Information on plasmid-borne quinolone resistance genes *qnr* genes and *qnr* allele designations can also be accessed from http://www.lahey.org/qnrStudies/.

At the time of this review, due to the transitioning of the database to a new location, it was not assessed. Institute Pasteur database (http://bigsdb.web.pasteur.fr/) provides MLST data for *Klebsiella pneumoniae*, including information particularly for specific β –lactamases, OKP, LEN and OXY. Latest resources regarding tetracycline and macrolide-lincosamide-streptogramin (MLS) AR genes can be obtained from the Tetracycline and MLS Nomenclature database (faculty.washington.edu/marilynr/). The latest update for the tetracycline and MLS resources was in August 2015.

Antibiotic Resistance Gene Finder (ABRESfinder; http://www.bioindians.org/ABRES/) is an AR gene resource that includes information on the gene sub-families and on the mechanisms of resistance from various sources. ABRESfinder provides links to tools such as BLAST, ClustalW and Primer3Plus and is mainly focused on AR-related information from India. Additionally, we would also like to shed light on some of the databases that provide information on integrons, AR-related gene cassettes and virulence factors. INTEGRALL, the integron database (http://integrall.bio.ua.pt) is a freely available web tool that provides information on integron sequences and their genetic arrangements with respect to AR genes (Moura, et al. 2009). Furthermore, annotation and information of gene cassettes in mobile integrons can also be accessed using the Repository of Antibiotic Resistance Cassettes (RAC; http://rac.aihi.mq.edu.au/rac/) (Tsafnat, et al. 2011a).

The MvirDB (http://mvirdb.llnl.gov/) is a database that targets genes for signature discovery, mainly in identification and characterization of both functional and genetic signatures. MvirDB has a collection of virulence factors and toxins, including AR gene related resources from several other databases. Apart from aiding medical researchers, the database also aims at centralizing information for bio-defense purposes concerning virulence factors and toxins, especially for tracking genetically engineered organisms. The web interface includes two tools – Virulence browser and Virulence BLAST Interface for sequence identification and comparison. MvirDB also compares entries to a high-throughput microbial annotation database (MannDB; http://manndb.llnl.gov/) (Zhou, et al. 2007). Although the web tools were found to be functional and easy to work with, we observed some broken links in the documentation section.

Besides the specific characteristics discussed above, one of the common issues that arises with AR gene databases is of false positive predictions due to certain housekeeping genes that are ubiquitously present in bacterial and, sometimes, in mammalian genomes. For instance, dihydrofolate reductases (*dhfr*) are important enzymes catalyzing the folic acid pathway in bacteria and are targeted by the antibiotic trimethoprim. Resistance arises either by overproduction of chromosomal DHFR due to a promoter mutation in *E. coli* or by an altered chromosomally encoded DHFR due to a single amino acid substitution in the *dhfr* gene in *S. aureus (Eliopoulos GM 2001).* However, naturally insensitive enzymes have also been reported in some organisms (Eliopoulos GM 2001; Forsberg, et al. 2014) and the mammalian *dhfr genes* are also highly similar to their bacterial counterparts. Expectedly, this gene is highly represented in metadata from various communities including fecal and soil sources (Forsberg, et al. 2014), and may be challenging for most AR gene databases, which do not include information from soil and ecological microbiome studies, to single out as false positive predictions. ResFams is one database that includes soil resistome metadata, which aids accurate predictions. Furthermore, terminologies or gene names might vary, for example, *dhfr* is often referred to as *dfrA* in certain databases. This ambiguity can be counteracted by conducting a parallel protein-domain based search using related databases such as Pfam. Thus, although a number of databases are available at our disposal, working with large amount of data still requires fine tuning of parameters, such as identity levels, e-value and bit score to predict the right AR genes and obtain better sensitivity.

2.1.3 Assessment of the performance of AR gene databases using gene sequences, whole genome sequences and (functional) metagenomics data

Next, we carried out test runs on the selected databases namely the ARDB, CARD, Resfinder and CBMAR, using our in-house data and those of others. We particularly selected these databases, as the ARDB, CARD and Resfinder are among the most popular AR related reference data sources. CBMAR is a recently established database that offers a comprehensive collection of data resources and tools related to AR genes. The query sequences used to assess the databases comprised of AR gene sequences, whole genome sequences and metagenomics data, including whole genome shotgun and functional metagenomics sequences.

To further verify the availability of latest resources and the accuracy of AR gene predictions from the 4 databases, we selected some of the known carbapenemase genes, *bla*$_{\text{VIM}}$ and *bla*$_{\text{NDM,}}$ and their variants as query sequences. The entire sequence of *bla*$_{\text{VIM-1}}$ (KT124311), *bla*$_{\text{VIM-2}}$ (KR337992.1), *bla*$_{\text{VIM-4}}$ (AJ585042.1), *bla*$_{\text{VIM-19}}$ (KT124310), *bla*$_{\text{VIM-35}}$ (JX982634.1) and *bla*$_{\text{NDM}}$ genes such as *bla*$_{\text{NDM-1}}$ (KP770030.1), *bla*$_{\text{NDM-2}}$ (JF703135.1), *bla*$_{\text{NDM-4}}$ (KP772213), *bla*$_{\text{NDM-6}}$ (KJ872581.1) and *bla*$_{\text{NDM-8}}$ (NG_036906.1) were downloaded from the NCBI database and used in our analysis. Runs were performed with the BLAST parameters set to the default for each of the databases used. Out of the 10 genes that we used for screening the 4 databases, 3 (*bla*$_{\text{VIM-1,}}$ *bla*$_{\text{VIM-2}}$, and *bla*$_{\text{VIM-4}}$) were predicted correctly by all of the 4 databases. *bla*$_{\text{VIM-19}}$ and *bla*$_{\text{VIM-35}}$ were incorrectly predicted by ARDB and CBMAR databases. ARDB returned several non-specific hits to *bla*$_{\underline{\text{VIM}}}$ gene type with an average similarity percentage of 94.12%, 96.43% for *bla*$_{\text{VIM-19}}$ and *bla*$_{\text{VIM-35,}}$ respectively. The results shortlisted *bla*$_{\underline{\text{VIM}}}$ genes but not the variants used as query. In case of CBMAR, *bla*$_{\text{VIM-19}}$ and *bla*$_{\text{VIM-35}}$ yielded non-specific hits; the top 10 hits pointed to *bla*$_{\text{VIM-4,}}$ and *bla*$_{\text{VIM-1,}}$ *bla*$_{\text{VIM-4,}}$ *bla*$_{\text{VIM-5}}$ genes respectively. The CARD and Resfinder databases produced correct results. In case of the *bla*$_{\text{NDM}}$ genes, BLAST results from the ARDB and CBMAR databases returned no hits. Whereas, the CARD and Resfinder databases were found to consistently return correct hits (Figure 2.1.2a). We did not observe any differences in results upon use of the entire or partial sequence as query.

Next, we utilized whole genome sequences of 3 MRSA strains, UAS391, H-EMRSA-15 and JKD6008 with accession numbers CP007690, CP007659 (Sabirova, et al. 2014b) and CP002120 (Howden, et al. 2010), respectively, in order to assess the databases' ability to predict AR genes and their variants from among whole genome sequence data. Results obtained using whole genome sequences showed that CARD detected the maximum number of AR genes – 6 for UAS391 and JKD6008, and 4 for the H-EMRSA-15 strains. Resfinder predicted 5 for JKD6008, and 1 each for UAS391 and H-EMRSA-15. CBMAR detected 1 each for JKD6008 and H-EMRSA-15 and no hits

for UAS391 (Figure 2.1.2b). We were unable to receive results from ARDB as our sequence files of 2.8 Mb we not accepted as query.

Additionally, we also screened the databases with a publicly available whole genome shotgun metagenomics dataset with primary accession number PRJEB3977. The data was obtained from a study considering the effects of a decolonization strategy on the gut resistome (Buelow, et al. 2014). Utilizing this data as query, CARD database predicted the maximum number of resistance genes – a total of 11, including 2 aminoglycoside resistance genes, 7 β-lactamases and 2 either undefined or other genes. While the Resfinder detected a total of 2 including 1 aminoglycoside and 1 β -lactamase, the ARDB predicted a total of 4, 1 each from aminoglycoside, β-lactamase, tetracycline and others (Figure 2.1.2c). In this case, CBMAR detected no genes.

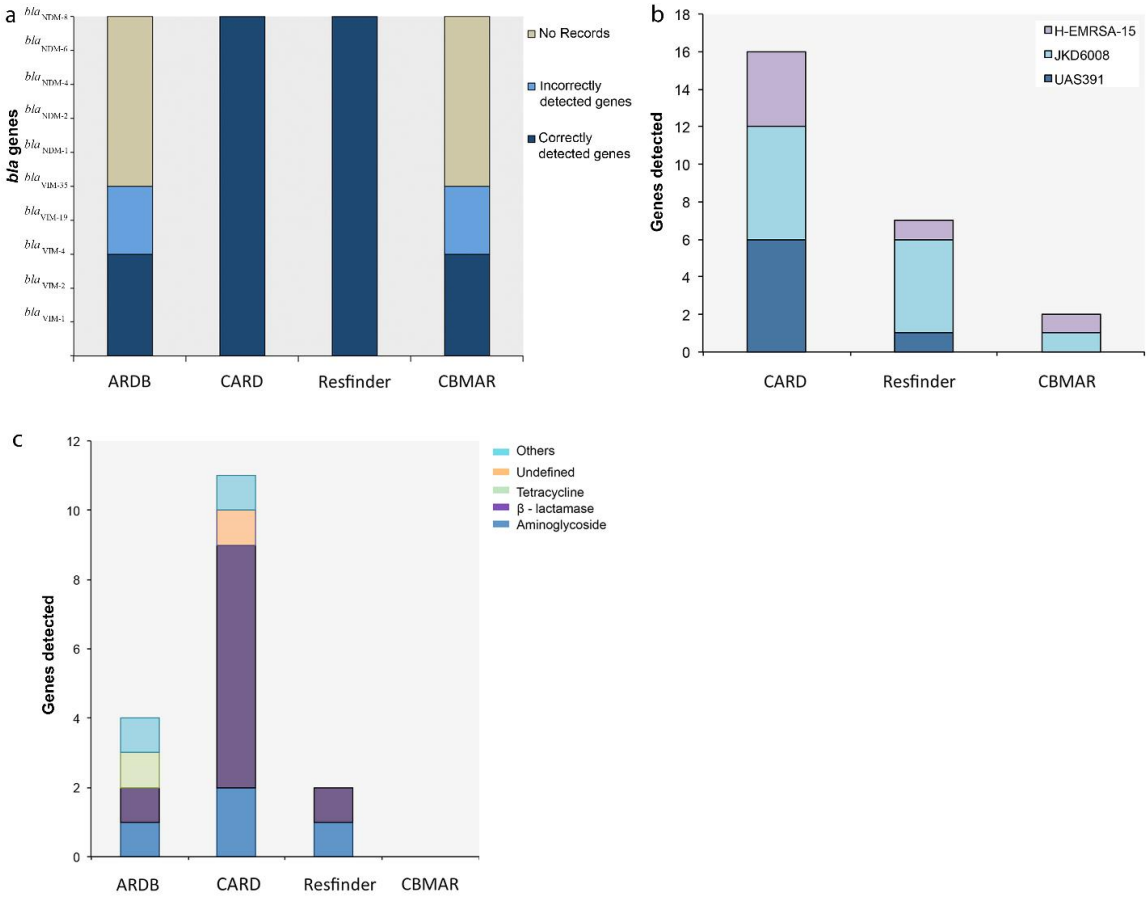


Figure 2.1.2 Comparison of AR gene data resources ARDB, CARD, Resfinder, and CBMAR using single gene sequences, whole genome sequences and metagenomics datasets as queries. Blast results obtained with *bla*$_{VIM}$ *and bla*$_{NDM}$ genes and their variants as query against the four databases (a). Results obtained using whole genome sequences (H-EMRSA-15, JKD6008, UAS391) (b) and metagenomic sequences (45) as query (c).

We also screened the databases using functional metagenomics data that came from a recently concluded study of the naso-oro-pharyngeal resistome from 150 healthy individuals across

five countries representing the Northern (Sweden), Southern (Spain), Eastern (Poland, Slovakia) and Western (Belgium) parts of Europe (Vervoort J, Xavier B.B, Joossens M, Darzi Y, Versporten A, Lammens C, Raes J, Goossens H and Malhotra-Kumar S. submitted for publication). Here, we utilized a functional metagenomic approach in order to identify differences in presence of antibiotic resistance genes harboured by healthy individuals and attempted to correlate it to antibiotics consumed in that particular country. Samples were enriched overnight in presence of different antibiotics, the DNA was isolated, sheared and cloned in *E. coli*. From the resistant clones, plasmid DNA was sequenced by Illumina (HiSeq), followed by filtering out vector-specific sequence reads and *denovo* assembly of remaining reads using Velvet v1.2.10 (Zerbino and Birney 2008). Derived contigs were used for BLAST search for AR genes against ResFinder, ARDB, CARD and CBMAR. First, we utilized Resfinder and CARD –RGI for primary screening and results were predicted by both tools. As Resfinder is restricted to acquired resistance genes, *norA,* a multi-drug efflux transporter gene identified by CARD was not identified by Resfinder. Similarly, trimethoprim resistance-conferring genes such as *dfrA/dhfr* gene were only predicted by Resfinder and not by CARD (RGI), CBMAR and ARDB databases. Apart from these predictions, we observed that HMM-based Resfams search on our resistome data gave us additional/novel *dfr* variants and also predicted two (*dfrA8* and *dfrG*) resistance genes previously identified by Resfinder. However, not all *dfr* genes identified by ResFams in our data were resistance-related, which calls for caution while interpreting output from broader databases that include soil/environmental microbiome data.

Finally, in order to check for availability of up-to-date reference information, we screened all of the selected databases for the recently reported *mcr*-1 gene, which has been linked to colistin resistance in bacteria. In our observation, as of 16 December 2015, Resfinder was the only database that correctly detected this gene.

In summary, out of the 4 popular databases that we screened for latest information and accuracy, ARDB was found to provide information limited to the gene name, but not the actual variant. Records of the bla_{NDM} genes were also missing in ARDB. Although ARDB is considered one of the most popular databases in identification of novel AR genes, lack of regular updates has limited its scope. CARD was found to accurately predict the query gene variant and provide several related hits in the BLAST results. We observed Resfinder to accurately predict all of the query genes. As for the CBMAR database, we found that predictions using nucleotide sequence of the variant gene (bla_{VIM}) provided correct hits to related bla_{VIM} gene variants in 3 out of 5 searches. Whereas BLAST search with bla_{NDM} variants returned no results, using a protein sequence query of these genes produced correct results.

Based on our results, we suggest that CARD and Resfinder are ideal while using single gene sequences as query. However, using whole genome sequences and metagenomic sequencing data, CARD performs better than the rest. The Resfinder database, which was found up to date and accurate, currently detects only acquired genes and ignores chromosomal mutations. The ARDB is limited in its scope due to lack of regular updates. As for the CBMAR database, on referring to the resources available for download, we found that the information to all the query variants were available and fully updated. This suggests that while the sequence repository of CBMAR database was found up-to-date, the search tools may need to be updated.

2.1.4 A future path

While we have noted the value of the data resources available to support AR-related work, we have also noted a number of limitations. These include gaps, inconsistent results of searches against different resources with the same query data and lack of up-to-date reference data. While it is beyond the scope in this minireview for us to formulate solutions to these issues, and it is certainly true that expertise beyond ours alone will be required for these solutions, we take the opportunity here to lay out some thinking that we hope will be useful in stimulating, and perhaps steering, community discussions as to the solutions. We trust that we, and others, will be able to take advantage of existing initiatives, such as the Horizon 2020 COMPARE (COllaborative Management Platform for detection and Analyses of (Re-) emerging and foodborne outbreaks in Europe; http://www.compare-europe.eu/) project, to facilitate and energise these community discussions.

Our proposal is to rise to the challenge with two complementary approaches, the first simpler to lay out in practical terms, the second requiring significant conceptual planning before practical work. The first approach charges the community to develop and implement appropriate best practices and standards in the gathering of reference AR data, in the description, publication and dissemination of these data and in the presentation of methodologies and algorithms offered through the services of each data resource. Establishing best practice around the open sharing of richly and systematically described reference data (such as sequences, annotations, alignments and models) is a step that will reduce redundant effort in discovering source data for analysis and curation in specialist resources and will maximise opportunities to fill gaps. Systematic descriptions of computational methods and query services offered by specialist AR data resources will aid in users' selection of appropriate tools for their analyses and minimise risk of misinterpretation. In this first approach, we do not seek to fill gaps where they exist in AR data resource services, nor to

benchmark precision and reliability, but rather seek to create a landscape of transparent and tractable elements that can contribute to many different current and future analytical infrastructures.

Our second approach calls for decisions as to how data resources, both generalist and AR specialist, should move forward to fill gaps in coverage, to provide consistency between query tools that are intended to serve the same function, to remove redundant data processing, curation and software development steps to maximise overall productivity and to guide consumers in making informed analyses using the most appropriate tools. Clearly, broad community engagement will be required to tackle these issues. While the AR data resource provider and consumer community will no doubt present very specific needs, a number of successful initiatives in other domains will be informative. RNAcentral, for example, is the product of a broad collaboration between 38 specialist data resources covering different families of non-coding RNA genes (Consortium 2015) in a model that centralises database components for non-coding sequences and comprehensive search and discovery across, so far, 22 of the collaborating data resources, while maintaining expert curation and specialist web access at the expert site. A second example, which differs in its model, comes from the Generic Model Organism Database (GMOD) project (http://gmod.org/), which provides software tools for the maintenance and presentation of model organism data across many community projects, including FlyBase (http://flybase.org/), WormBase (http://www.wormbase.org/#01-23-6) and DictyBase (http://dictybase.org/), for example.

2.1.5 Summary and conclusions

In this minireview, we have compiled information on the available data resources that relate to AR function. In the fight against the spread of MDR pathogens, a collective effort is being made in establishing these resources to share knowledge in free and accessible ways. While we find substantial value in what is already available, we note a number of limitations, including those that relate to frequency of updates, and functionality and comprehensiveness of the resources as a whole. Indeed, broader coverage, consistent gene terminologies, centralized (unifying soil, human and other microbiome/resistome data) up-to-date records will be crucial in identifying and tracking (novel) genomic alterations that are acquired by bacterial pathogens upon progression to antibiotic resistance.

At about a time when NGS has become affordable and relatively rapid, and MDR poses an ever-greater challenge to public and animal health, a greater need for comprehensive, up-to-date and interoperable AR gene data resources is created. We hope that this evaluation will initiate a

strategic response among data resource managers to come together to work out mutual solutions, which will make it easier for their operations to be sustained and kept more up-to-date. As coordination at the international level of pathogen genomics efforts grows, we urge that attention be paid to sustaining and extending AR gene data resources as a critical component of our response to MDR.

Authors contribution: BBX was involved in data analysis, interpretation and drafting manuscript.

2.1.6 References

1. Cabello FC. 2006. Heavy use of prophylactic antibiotics in aquaculture: a growing problem for human and animal health and for the environment. Environ Microbiol 8:1137-1144.
2. Burns A, Shore AC, Brennan GI, Coleman DC, Egan J, Fanning S, Galligan MC, Gibbons JF, Gutierrez M, Malhotra-Kumar S, Markey BK, Sabirova JS, Wang J, Leonard FC. 2014. A longitudinal study of Staphylococcus aureus colonization in pigs in Ireland. Vet Microbiol 174:504-513.
3. Aarestrup FM. 2015. The livestock reservoir for antimicrobial resistance: a personal view on changing patterns of risks, effects of interventions and the way forward. Philos Trans R Soc Lond B Biol Sci 370:20140085.
4. Davies J, Davies D. 2010. Origins and evolution of antibiotic resistance. Microbiol Mol Biol Rev 74:417-433.
5. Mwangi MM, Wu SW, Zhou Y, Sieradzki K, de Lencastre H, Richardson P, Bruce D, Rubin E, Myers E, Siggia ED, Tomasz A. 2007. Tracking the in vivo evolution of multidrug resistance in Staphylococcus aureus by whole-genome sequencing. Proc Natl Acad Sci U S A 104:9451-9456.
6. Ploy MC, Lambert T, Couty JP, Denis F. 2000. Integrons: an antibiotic resistance gene capture and expression system. Clin Chem Lab Med 38:483-487.
7. Liu YY, Wang Y, Walsh TR, Yi LX, Zhang R, Spencer J, Doi Y, Tian G, Dong B, Huang X, Yu LF, Gu D, Ren H, Chen X, Lv L, He D, Zhou H, Liang Z, Liu JH, Shen J. 2015. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. Lancet Infect Dis doi:10.1016/S1473-3099(15)00424-7.
8. Hasman H, Hammerum AM, Hansen F, Hendriksen RS, Olesen B, Agerso Y, Zankari E, Leekitcharoenphon P, Stegger M, Kaas RS, Cavaco LM, Hansen DS, Aarestrup FM, Skov RL. 2015. Detection of mcr-1 encoding plasmid-mediated colistin-resistant Escherichia coli isolates from human bloodstream infection and imported chicken meat, Denmark 2015. Euro Surveill 20.
9. Malhotra-Kumar S, Xavier BB, Das AJ, Lammens C, Hoang HTT, Pham NT, Goossens H. 2016. Colistin-resistant Escherichia coli harbouring mcr-1 isolated from food animals in Hanoi, Vietnam. Lancet Infect Dis, in press.
10. Malhotra-Kumar S, Xavier BB, Das AJ, Lammens C, Butaye P, Goossens H. 2016. Colistin resistance conferring mcr-1 isolated from Belgian bovine and pig farms is harbored on a novel multi-drug resistant plasmid. Lancet Infect Dis, in press.
11. Carattoli A. 2013. Plasmids and the spread of resistance. Int J Med Microbiol 303:298-304.
12. Vervoort J, Gazin M, Kazma M, Kotlovsky T, Lammens C, Carmeli Y, Goossens H, Malhotra-Kumar S, Saturn WP, groups MWs. 2014. High rates of intestinal colonisation with fluoroquinolone-resistant ESBL-harbouring Enterobacteriaceae in hospitalised patients with antibiotic-associated diarrhoea. Eur J Clin Microbiol Infect Dis 33:2215-2221.
13. Alekshun MN, Levy SB. 2007. Molecular mechanisms of antibacterial multidrug resistance. Cell 128:1037-1050.

14. Pournaras S, Poulou A, Voulgari E, Vrioni G, Kristo I, Tsakris A. 2010. Detection of the new metallo-beta-lactamase VIM-19 along with KPC-2, CMY-2 and CTX-M-15 in Klebsiella pneumoniae. J Antimicrob Chemother 65:1604-1607.
15. Queenan AM, Bush K. 2007. Carbapenemases: the versatile beta-lactamases. Clin Microbiol Rev 20:440-458, table of contents.
16. Canton R, Akova M, Carmeli Y, Giske CG, Glupczynski Y, Gniadkowski M, Livermore DM, Miriagou V, Naas T, Rossolini GM, Samuelsen O, Seifert H, Woodford N, Nordmann P, European Network on C. 2012. Rapid evolution and spread of carbapenemases among Enterobacteriaceae in Europe. Clin Microbiol Infect 18:413-431.
17. Malachowa N, DeLeo FR. 2010. Mobile genetic elements of Staphylococcus aureus. Cell Mol Life Sci 67:3057-3071.
18. Vervoort J, Xavier BB, Stewardson A, Coenen S, Godycki-Cwirko M, Adriaenssens N, Kowalczyk A, Lammens C, Harbarth S, Goossens H, Malhotra-Kumar S. 2015. Metagenomic analysis of the impact of nitrofurantoin treatment on the human faecal microbiota. J Antimicrob Chemother 70:1989-1992.
19. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Dore J, Meta HITC, Antolin M, Artiguenave F, Blottiere HM, Almeida M, Brechot C, Cara C, Chervaux C, Cultrone A, Delorme C, Denariaz G, Dervyn R, Foerstner KU, Friss C, van de Guchte M, Guedon E, Haimet F, Huber W, van Hylckama-Vlieg J, Jamet A, Juste C, Kaci G, Knol J, Lakhdari O, Layec S, Le Roux K, Maguin E, Merieux A, Melo Minardi R, M'Rini C, Muller J, Oozeer R, Parkhill J, Renault P, Rescigno M, Sanchez N, Sunagawa S, Torrejon A, Turner K, Vandemeulebrouck G, Varela E, Winogradsky Y, Zeller G, Weissenbach J, Ehrlich SD, Bork P. 2011. Enterotypes of the human gut microbiome. Nature 473:174-180.
20. Canica M, Manageiro V, Jones-Dias D, Clemente L, Gomes-Neves E, Poeta P, Dias E, Ferreira E. 2015. Current perspectives on the dynamics of antibiotic resistance in different reservoirs. Res Microbiol 166:594-600.
21. Didelot X, Bowden R, Wilson DJ, Peto TE, Crook DW. 2012. Transforming clinical microbiology with bacterial genome sequencing. Nat Rev Genet 13:601-612.
22. Zankari E, Hasman H, Kaas RS, Seyfarth AM, Agerso Y, Lund O, Larsen MV, Aarestrup FM. 2013. Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. J Antimicrob Chemother 68:771-777.
23. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 67:2640-2644.
24. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25-29.
25. Eilbeck K, Lewis SE, Mungall CJ, Yandell M, Stein L, Durbin R, Ashburner M. 2005. The

Sequence Ontology: a tool for the unification of genome annotations. Genome Biol 6:R44.
26. Liu B, Pop M. 2009. ARDB--Antibiotic Resistance Genes Database. Nucleic Acids Res 37:D443-447.
27. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, Bhullar K, Canova MJ, De Pascale G, Ejim L, Kalan L, King AM, Koteva K, Morar M, Mulvey MR, O'Brien JS, Pawlowski AC, Piddock LJ, Spanogiannopoulos P, Sutherland AD, Tang I, Taylor PL, Thaker M, Wang W, Yan M, Yu T, Wright GD. 2013. The comprehensive antibiotic resistance database. Antimicrob Agents Chemother 57:3348-3357.
28. Goldfain A, Smith B, Cowell LG. 2010. Dispositions and the Infectious Disease Ontology. Formal Ontology in Information Systems (Fois 2010) 209:400-413.
29. Zankari E. 2014. Comparison of the web tools ARG-ANNOT and ResFinder for detection of resistance genes in bacteria. Antimicrob Agents Chemother 58:4986.
30. Thai QK, Bos F, Pleiss J. 2009. The Lactamase Engineering Database: a critical survey of TEM sequences in public databases. BMC Genomics 10:390.
31. Gibson MK, Forsberg KJ, Dantas G. 2014. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. The ISME Journal 9:207-216.
32. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, Rolain JM. 2014. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 58:212-220.
33. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R, Machi D, Mao C, Nordberg EK, Olson R, Overbeek R, Pusch GD, Shukla M, Schulman J, Stevens RL, Sullivan DE, Vonstein V, Warren A, Will R, Wilson MJ, Yoo HS, Zhang C, Zhang Y, Sobral BW. 2014. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res 42:D581-591.
34. Human Microbiome Project C. 2012. A framework for human microbiome research. Nature 486:215-221.
35. Saha SB, Uttam V, Verma V. 2015. u-CARE: user-friendly Comprehensive Antibiotic resistance Repository ofEscherichia coli. Journal of Clinical Pathology doi:10.1136/jclinpath-2015-202927:jclinpath-2015-202927.
36. Danishuddin M, Hassan Baig M, Kaushal L, Khan AU. 2013. BLAD: a comprehensive database of widely circulated beta-lactamases. Bioinformatics 29:2515-2516.
37. Srivastava A, Singhal N, Goel M, Virdi JS, Kumar M. 2014. CBMAR: a comprehensive beta-lactamase molecular annotation resource. Database (Oxford) 2014:bau111.
38. Moura A, Soares M, Pereira C, Leitao N, Henriques I, Correia A. 2009. INTEGRALL: a database and search engine for integrons, integrases and gene cassettes. Bioinformatics 25:1096-1098.
39. Tsafnat G, Copty J, Partridge SR. 2011. RAC: Repository of Antibiotic resistance Cassettes. Database (Oxford) 2011:bar054.
40. Zhou CE, Smith J, Lam M, Zemla A, Dyer MD, Slezak T. 2007. MvirDB--a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. Nucleic Acids Res 35:D391-394.
41. Eliopoulos GM HP. 2001. Resistance to Trimethoprim-Sulfamethoxazole. Clinical Infectious Disease 32:6.

42. Forsberg KJ, Patel S, Gibson MK, Lauber CL, Knight R, Fierer N, Dantas G. 2014. Bacterial phylogeny structures soil resistomes across habitats. Nature 509:612-616.
43. Sabirova JS, Xavier BB, Hernalsteens JP, De Greve H, Ieven M, Goossens H, Malhotra-Kumar S. 2014. Complete Genome Sequences of Two Prolific Biofilm-Forming Staphylococcus aureus Isolates Belonging to USA300 and EMRSA-15 Clonal Lineages. Genome Announc 2.
44. Howden BP, Seemann T, Harrison PF, McEvoy CR, Stanton JA, Rand CJ, Mason CW, Jensen SO, Firth N, Davies JK, Johnson PD, Stinear TP. 2010. Complete genome sequence of Staphylococcus aureus strain JKD6008, an ST239 clone of methicillin-resistant Staphylococcus aureus with intermediate-level vancomycin resistance. J Bacteriol 192:5848-5849.
45. Buelow E, Gonzalez TB, Versluis D, Oostdijk EA, Ogilvie LA, van Mourik MS, Oosterink E, van Passel MW, Smidt H, D'Andrea MM, de Been M, Jones BV, Willems RJ, Bonten MJ, van Schaik W. 2014. Effects of selective digestive decontamination (SDD) on the gut resistome. J Antimicrob Chemother 69:2215-2223.
46. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821-829.
47. Consortium RN. 2015. RNAcentral: an international database of ncRNA sequences. Nucleic Acids Res 43:D123-129.

2.2 BACPIPE: A RAPID, USER-FRIENDLY WHOLE GENOME SEQUENCING PIPELINE FOR CLINICAL DIAGNOSTIC BACTERIOLOGY AND OUTBREAK DETECTION

Basil Britto Xavier[1]*, Mohamed Mysara[1*], Mattia Bolzan[1], Christine Lammens[1], Samir Kumar-Singh[1,2], Herman Goossens[1], Surbhi Malhotra-Kumar[1]

[1]Laboratory of Medical Microbiology, Vaccine & Infectious Disease Institute, University of Antwerp, Wilrijk, Belgium, [2]Molecular Pathology Group, Cell Biology and Histology, University of Antwerp, Wilrijk, Belgium

Abstract

Despite rapid advances in whole genome sequencing (WGS) technologies, their integration into routine microbiological diagnostics and infection control has been hampered by the need for downstream bioinformatics analyses that require considerable expertise. We have developed a comprehensive, rapid, and computationally low-resource bioinformatics pipeline (BacPipe) that enables direct analyses of bacterial whole-genome sequences (raw reads, contigs or scaffolds) obtained from second and third-generation sequencing technologies. BacPipe is an ensemble of state-of-the-art, open-access tools for quality verification, genome assembly, annotation, and identification of the bacterial genotype (MLST, *emm* typing), resistance genes, plasmids, virulence genes, and single nucleotide polymorphisms (SNPs). The outbreak module (SNPs and patient metadata) can simultaneously analyse many strains to identify evolutionary relationships and transmission routes. Importantly, parallelization of tools in BacPipe considerably reduces the time-to-result. Validation of BacPipe using prior published WGS datasets from hospital, community and food-borne outbreaks and from transmission studies of important pathogens demonstrated the speed and simplicity of the pipeline that reconstructed the same analyses and conclusions within a few hours. We believe this fully automated pipeline will contribute to overcoming one of the primary hurdles to WGS data analysis and interpretation, facilitating its application for routine patient-care in hospitals and public-health and infection-control monitoring.

2.2.1 Introduction

Next-generation sequencing (NGS) technologies hold the promise to revolutionize the public health sector especially clinical diagnostic microbiology, infection control, outbreak detection, and antibiotic stewardship in hospitals (Arnold 2015; Kwong, et al. 2015; Moran-Gilad 2017). As costs of sequencing are steadily decreasing and response times getting shorter, their utility as tools for tracking multi-drug resistant (MDR) pathogens in real time for routine hospital epidemiology or as an early warning system for outbreak detection is steadily increasing (Punina, et al. 2015). Currently, depending on the pathogen, the identification and characterization process may take 1 to 7 days for culture, an additional 1 to 2 days for species identification and susceptibility testing, and one to several weeks for molecular typing. Whole genome sequencing (WGS) of bacterial isolates combines identification, molecular typing, and prediction of antimicrobial susceptibility and virulence, theoretically reducing the time-to-result for these procedures to a few days (Didelot, et al. 2012b; Joensen, et al. 2014; Koser, et al. 2012). However, despite rapid advances in WGS workflows and in NGS technologies, their integration into routine microbiological diagnostics and infection control has been hampered by the need for downstream bioinformatics analyses that is challenging and requires considerable expertise (Deurenberg, et al. 2017; Muir, et al. 2016). WGS analysis comprises different stages, and each stage is crucial for correct data interpretation. While there are commercial software such as CLC Genomics Workbench (qiagen), DNA Star (DNASTAR Inc., USA), BioNumerics (Applied Maths), SeqSphere^{+} (Ridom GmbH, Münster, Germany) etc., available, apart from being expensive, these might also be restrictive in the analyses offered. For instance, genome assembly, reference mapping, SNP analysis can be performed but not all commercial softwares offer the option for genome annotation, or resistance and virulence gene predictions and some offer only organism-specific analysis. Thus, these do not offer a comprehensive, 'raw data in ─ results out' solution for users with limited or no bioinformatics expertise (Deurenberg, et al. 2017). Mykrobe, a non-commercial software was also recently introduced (http://www.mykrobe.com), however, analyses of only specific organisms, *Mycobacterium tuberculosis* and *Staphylococcus aureus*, is as yet possible (Bradley, et al. 2016).

There also exist many open-access tools for specific analyses, but most require Linux- or Unix-based operating systems, and therefore, a basic command-line knowledge to run the analysis (Bodi 2011). Few web-based open access pipelines such as Orione (http://orione.crs4.it) (Cuccuru, et al. 2014) and the Bacterial analysis pipeline (https://cge.cbs.dtu.dk/services/cge/) (Thomsen, et al. 2016), and the microbial genomics virtual laboratory

https://www.melbournebioinformatics.org.au/project/micro-gvl/) are also available. Orione is available in the Galaxy portal (https://usegalaxy.org/) and can also be locally installed. It offers WGS quality control, assembly and annotation, and variant calling (Cuccuru, et al. 2014). The Bacterial analysis pipeline (https://cge.cbs.dtu.dk/services/cge/) is a widely used comprehensive pipeline that offers molecular typing tools as well as resistance and virulence gene predictions, and SNP based phylogeny However, performance of web-server based analysis depends on the server load and requires a fast and consistent internet connection to upload large raw data files, which for multiple strains might take hours. Other issues encountered are mandatory user registration to save the data and run the analysis, and importantly for hospitals, barriers to uploading data into the public domain which remains a sensitive matter with policies varying between countries (Akgün, et al. 2015; Muir, et al. 2016).

Other locally installable tools developed specifically for running and managing microbial genomics pipelines are IRIDA (irida.ca), Innuendo (http://www.innuendoweb.org/project-definition), and nullarbor (https://github.com/tseemann/nullarbor). IRIDA provides a workflow for assembly (SPAdes), annotation (Prokka), SNP phylogeny (SNVPhyl), resistance (CARD) and virulence (Islandviewer) but not for plasmids and MLST typing. Similarly, INNUENDO with its INNUca workflow provides quality control of reads, *de novo* assembly and contigs quality assessment. Nullarbor supports Illumina paired-end sequencing data but not single-end reads from either Illumina or Ion Torrent. All jobs are run on a parallelization mode by distributing the work across a high performance cluster.

To overcome the various issues discussed above, we have developed a rapid, ‘one-stop’ bacterial WGS analysis pipeline, BacPipe, which also offers a unique graphical interface. Following selection of the raw data files, the analysis spans a quality check, genome assembly, and annotation, resulting in bacterial typing, resistance, and virulence gene predictions, as well as single nucleotide polymorphisms (SNP)-based phylogeny.

2.2.2 Materials and Methods

We first reviewed the open-access tools available online that were potential candidates for inclusion within BacPipe for each stage of analyses. These included tools required for quality control check, assembly, as well as specialized tools for bacterial and plasmid typing, and for resistance and virulence gene predictions. As some of these specialized tools required annotation, these were divided within BacPipe into those that required assembly (post-assembly tools) or annotation (post-

annotation analysis) (Figure 2.2.1). The optimal tools were selected for inclusion in BacPipe as described below. Furthermore, to increase the user-friendliness of the pipeline, we also integrated a graphical user interface (GUI) (Figure 2.2.2). The GUI was developed using AppJar package (http://appjar.info/), split into four tabs: Settings (inputs and tools parameters), Progress (percentage finished and log file), Results (overall summary files are shown), Help (information regarding the input/outputs), Citation (for all tools included). BacPipe is also modular i.e., depending on the analyses required, the user can select for a particular tool or a set of tools that will also further speed up the time-to-result (Figure 2). Finally, users can choose to directly analyse raw sequencing reads or contigs or scaffolds (for instance, published sequences) in BacPipe.

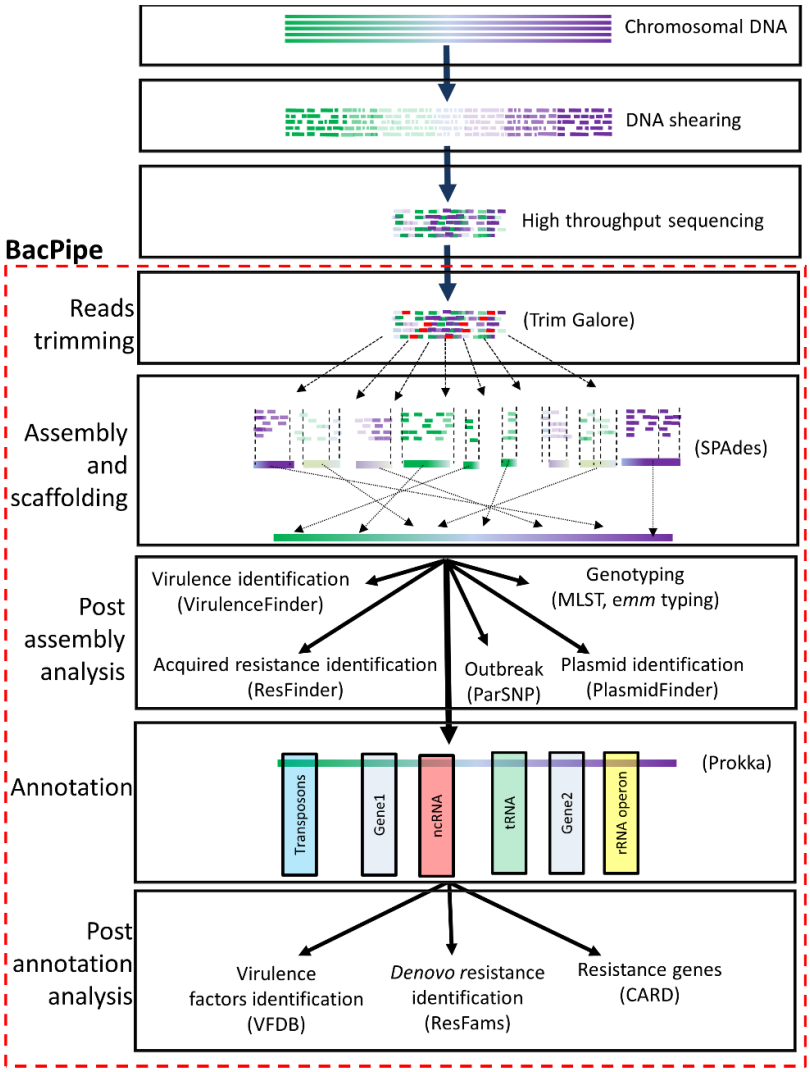


Figure 2.2.1 Complete overview of NGS workflow and analysis performed within BacPipe.

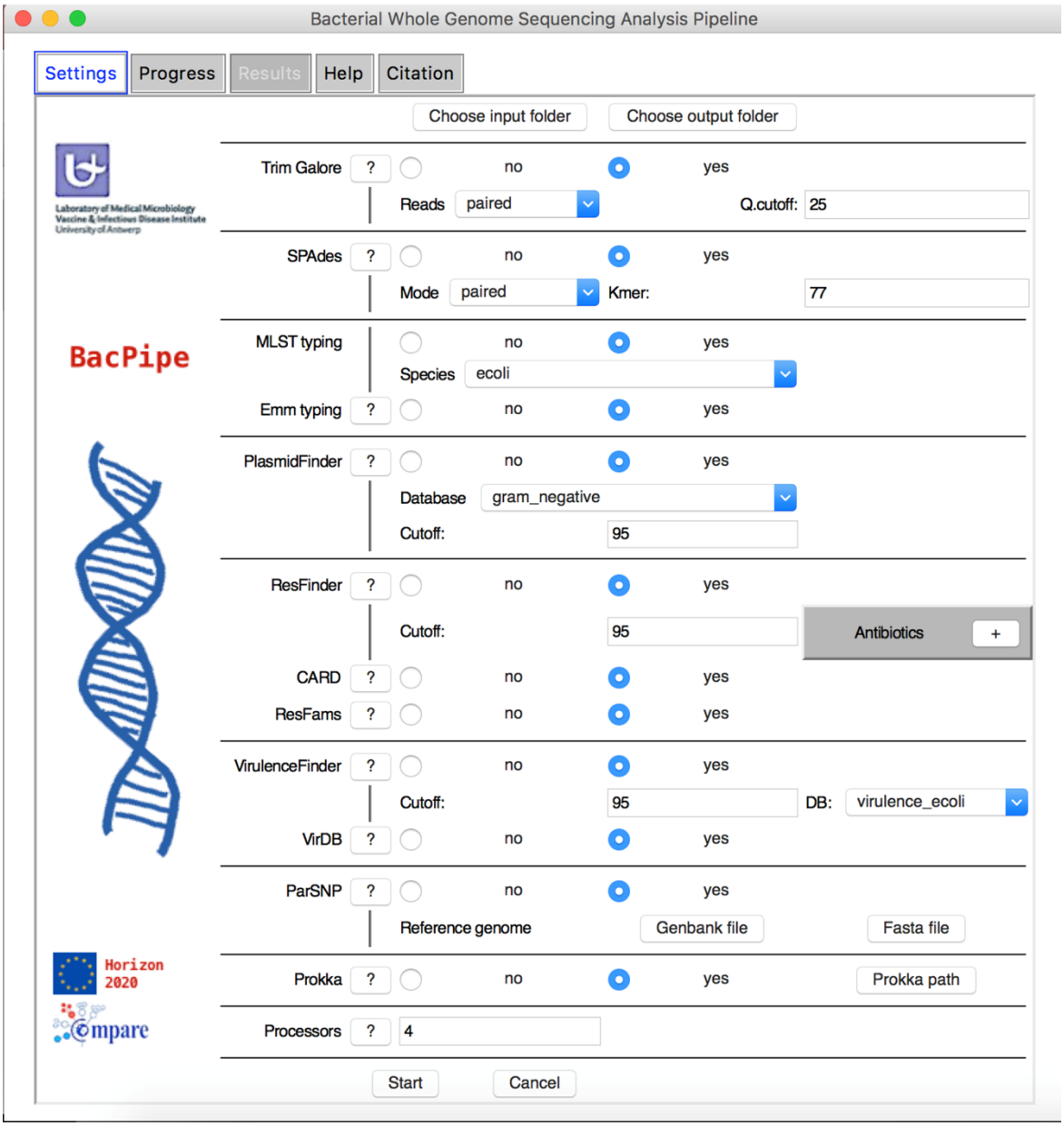


Figure 2.2.2 Screenshot of the BacPipe graphical user interface (GUI).

2.2.3 Development of BacPipe

2.2.3.1 Reads quality filtering and adapter trimming

Quality control and processing of raw reads are the initial steps and are extremely crucial for robust downstream analysis. For quality control, we opted for Trim Galore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) that removes sequencing technology-specific sequences and trims low-quality reads. This tool consists of FastQC and

Cutadapt, the former to check the quality of the reads and the latter to trim the sequencing-specific (adapter and index) sequences.

2.2.3.2 Generating contigs and scaffolds from reads: Genome assembly

To obtain any reliable information on the genetic context of genes, an error-free assembly of the genome sequences is a must. Here, we utilized SPAdes that can automatically optimize the k-mer based on read length in combination with a scaffolding step (Bankevich, et al. 2012). The tool produces reasonably large scaffolds and with higher N50 values compared to other existing tools. Also, there are multiple options in SPAdes for different assembly purposes, such as plasmid and hybrid assembly, and output data from diverse sequencing technologies such as Ion Torrent, Pacbio, and Oxford Nanopore can be processed. Additionally, our previous work validating various assemblers using whole genome mapping (Xavier, et al. 2014a), showed that SPAdes produces larger scaffolds that are free from misassemblies and, in our hands, was the best performing assembler among the studied tools (i.e., Velvet and IDBA).

2.2.3.3 Pathogen typing, detection of plasmids, virulence and resistance genes, and outbreak prediction: Post assembly analysis

The post-assembly analysis tools include those for multi-locus sequence typing (MLST) (Larsen, et al. 2012), plasmid incompatibility typing (PlasmidFinder) (Carattoli, et al. 2014), antibiotic resistance gene predictions (Resfinder) (Zankari, et al. 2012b), and for virulence gene predictions (VirulenceFinder) (Joensen, et al. 2014). PlasmidFinder can classify plasmids into various incompatibility types based on some plasmid-specific genes (Carattoli, et al. 2014). VirulenceFinder can predict putative virulence genes in the scaffolds, however, currently, gene predictions are only possible for select bacterial species (*E. coli*, *S. aureus*, *Enterococcus spp.* and *Listeria*) (Joensen, et al. 2014). All post-assembly detection tools in BacPipe described above work with a local BLAST search utilizing the databases downloaded from the Center for Genomic Epidemiology (CGE) (https://bitbucket.org/account/user/genomicepidemiology/projects/DB).

Additionally, as a first step to adding pathogen-specific tools, we have added the option of utilizing the *S. pyogenes*-specific *emm* typing tool in BacPipe (Facklam, et al. 1999). This tool also uses a local BLAST search against the curated database (identity 100%; minimum length 90%) downloaded from https://www.cdc.gov/streplab/m-proteingene-typing.html (Center for Disease Control, CDC, United States). Very interestingly, BacPipe also includes a module that enables

prediction of a suspected outbreak. This module utilizes ParSNP, a tool using MUMmer (*mummer.sourceforge.net/)* for comparison of scaffolds/genomes and generates a SNP-based core genome phylogeny. This module allows the user to choose the index or reference strain (GenBank or FASTA file) to compare. The analysis produces the output variant calling file (VCF), tree file and multiple alignment files (Treangen, et al. 2014). The Newick file or tree file can then be visualized using Gingr, Figtree (http://tree.bio.ed.ac.uk/software/figtree/), or other online tools (www.phylogeny.fr).

2.2.4 Genome annotation

Correct annotation of a gene/genome is a requisite for correct biological interpretation of sequencing data, and to study the genomic features, structural variations and evolutionary relationships. It is also used to assess the quality of the assembly and required sequencing depth by identified complete or core genes of the organism sequenced. Prokka (Prokaryotic annotation tool) was used in our pipeline as an annotation tool to predict protein coding genes, and Barranp and ARAGORN tools were used for ribosomal RNAs (5S, 23S, 16S) predictions and for transfer RNA (tRNA) and transfer messenger RNA (tmRNA) predictions, respectively (Seemann 2014). As an output, generic output file formats gbk, gff3 and sqn (sequin) are generated as these file formats are required for the downstream analysis and for uploading to public databases such as NCBI Genbank for submission.

2.2.5 Post annotation analysis

Post-annotation tools in BacPipe include those that require protein sequences as queries such as the Comprehensive Antibiotic Resistance Database (CARD) that predicts resistance genes using local blastp (minimum identiy 80%; minimum length of 80%) (Jia, et al. 2017; McArthur, et al. 2013b). CARD adds to the ResFinder v2.1 database as the former also identifies resistance genes with a chromosomal origin. The current version of ResFinder 3.0 is also predict chromosomal specific resistance genes in addition to acquired ones but restricted to only few organisms. Another resistance gene prediction tool integrated into BacPipe is Resfams, a new curated database of protein families that uses hidden Markov model profiles of resistance genes to screen for distantly-related or novel resistance genes (Gibson, et al. 2015).

For virulence gene predictions, we have also integrated Virulence Factor Database VFDB (with 80% identity and 80% minimum query length for blastp), an extensive database of virulence

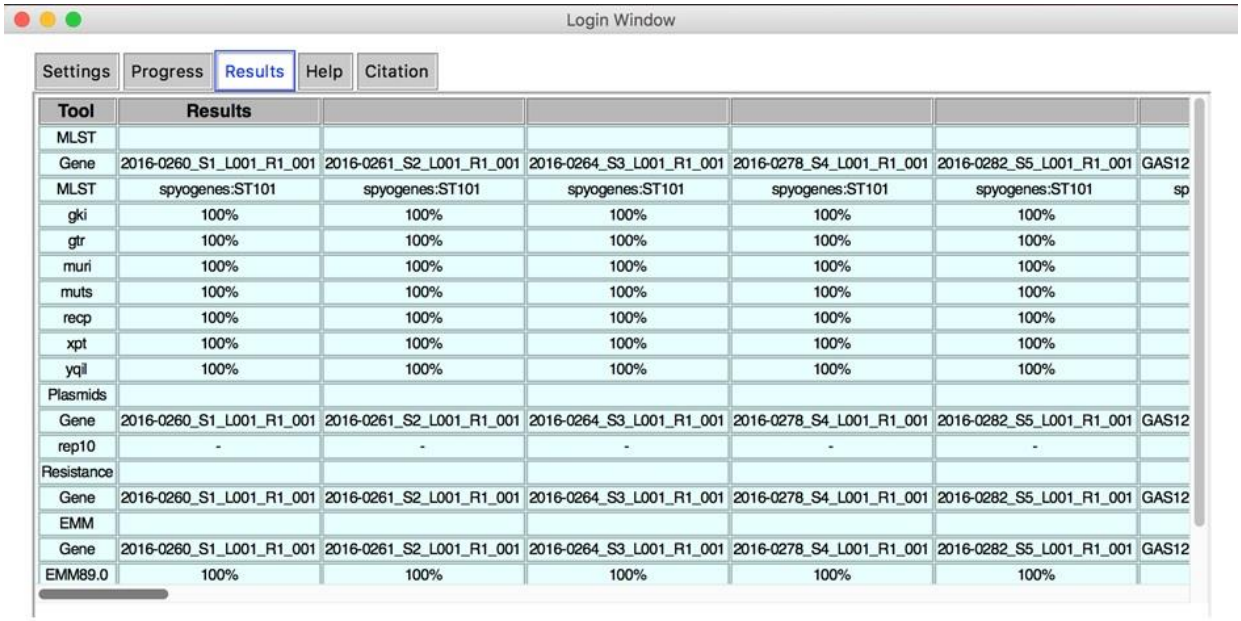
genes in pathogenic bacteria, as VirulenceFinder is restricted to a few organisms (Chen, et al. 2016). Finally, all these tools were linked through python script.

2.2.6 Tools parallelization

BacPipe is designed to run multiple tools simultaneously which considerably reduces the time-to-result. To demonstrate the impact of bacterial genome size on the computational time required, we processed three bacterial genomes that vary considerably in size, i.e., *Streptococcus pyogenes* (~1.8 Mb), *Escherichia coli* (~5.2 Mb), and *Pseudomonas aeruginosa* (~6.8 Mb). Whole genome sequences of all three pathogens were normalized to the same 70-fold-coverage before running them through BacPipe. Additionally, we also studied the impact of the sequencing depth on computational time by subsampling and analysing the *E. coli* sequences at 50, 70, 100 and 200-fold coverage.

2.2.7 Results summary

Raw results from each tool output for each strain analysed are saved as separate folders. In addition, a summary of the results for each strain are also saved as an Excel file in the overall 'summary' folder. Furthermore, if multiple strains are being run in parallel, then results of all tested strains are available in a comprehensive table after the run is finished (Supplementary Figure 2.2.1).



Tool	Results					
MLST						
Gene	2016-0260_S1_L001_R1_001	2016-0261_S2_L001_R1_001	2016-0264_S3_L001_R1_001	2016-0278_S4_L001_R1_001	2016-0282_S5_L001_R1_001	GAS12
MLST	spyogenes:ST101	spyogenes:ST101	spyogenes:ST101	spyogenes:ST101	spyogenes:ST101	sp
gki	100%	100%	100%	100%	100%	
gtr	100%	100%	100%	100%	100%	
muri	100%	100%	100%	100%	100%	
muts	100%	100%	100%	100%	100%	
recp	100%	100%	100%	100%	100%	
xpt	100%	100%	100%	100%	100%	
yqil	100%	100%	100%	100%	100%	
Plasmids						
Gene	2016-0260_S1_L001_R1_001	2016-0261_S2_L001_R1_001	2016-0264_S3_L001_R1_001	2016-0278_S4_L001_R1_001	2016-0282_S5_L001_R1_001	GAS12
rep10	-	-	-	-	-	
Resistance						
Gene	2016-0260_S1_L001_R1_001	2016-0261_S2_L001_R1_001	2016-0264_S3_L001_R1_001	2016-0278_S4_L001_R1_001	2016-0282_S5_L001_R1_001	GAS12
EMM						
Gene	2016-0260_S1_L001_R1_001	2016-0261_S2_L001_R1_001	2016-0264_S3_L001_R1_001	2016-0278_S4_L001_R1_001	2016-0282_S5_L001_R1_001	GAS12
EMM89.0	100%	100%	100%	100%	100%	

Supplementary Figure 2.2.1 Screenshot of summary of results in BacPipe.

2.2.7.1 Validation of BacPipe's functionality using prior published data

In order to validate the different analyses integrated in BacPipe, we utilized five prior published whole genome sequenced datasets that spanned the most important multi-drug resistant and virulent pathogens causing outbreaks or infections in hospitals and the community. These included datasets of methicillin-resistant *S. aureus* (MRSA) (Sabat, et al. 2017), carbapenem-resistant *K. pneumoniae* (Snitkin, et al. 2012), *C. difficile* (Jia, et al. 2016), *M. tuberculosis* (Kohl, et al. 2014), and *S. enterica* (Taylor, et al. 2015).

2.2.8 Results and Discussion

2.2.8.1 BacPipe components and computational time as a function of bacterial genome size and sequence fold-coverage

To demonstrate the impact of bacterial genome sizes on the computational time required to obtain results with BacPipe, we used three pathogen genomes that vary considerably in size, *Streptococcus pyogenes* (~1.8 Mb), *Escherichia coli* (~5.2 Mb), and *Pseudomonas aeruginosa* (~6.8 Mb). Whole genome sequences of each pathogen were normalized to the same fold-coverage to demonstrate an increase in computational time as a function of genome size. These isolates were sequenced from our in-house collection on the MiSeq platform, and as the *P. aeruginosa* PAO1 had 70-fold coverage, we randomly selected reads from the other two strains resulting in the same coverage. Normalized sequence data of each of the three pathogens is available at https://www.uantwerpen.be/en/rg/lab-of-medical-microbiology/projects-and-publications/publications/key-publications/tools/

Expectedly, computational time increased with increasing genome size totaling 9, 25, and 41 minutes for *S. pyogenes*, *E. coli*, and *P. aeruginosa*, respectively (Figure 2.2.3A). Among all the tools employed in the pipeline, genome assembly (SPAdes) was found to be the most computationally intensive, taking on average 36% of the total running time. Also, we assessed the added value of parallelizing the post-assembly tools (PlasmidFinder, ResFinder, VirulenceFinder, MLST & *emm* typing) and post-annotation tools (ResFams, VirDB, and CARD search). Parallelizing these tools resulted in a reduction of time-to-result (computational time) by 56%, 29% and 25% for the three pathogen sequences, respectively (data not shown). Additionally, to emphasize the increase in the computational time due to higher coverage, we subsampled the *E. coli*

sequences at 50, 70, 100 and 120-fold-coverage. The required computational time for 50, 70, 100, 120-fold-coverages were 21, 25, 28 and 30 minutes, respectively (Figure 2.2.3B). For this we used MacBook Pro, 2.5 GHz, quad-core i7 with 16 GB RAMS (DDR3) and SSD hard drive.

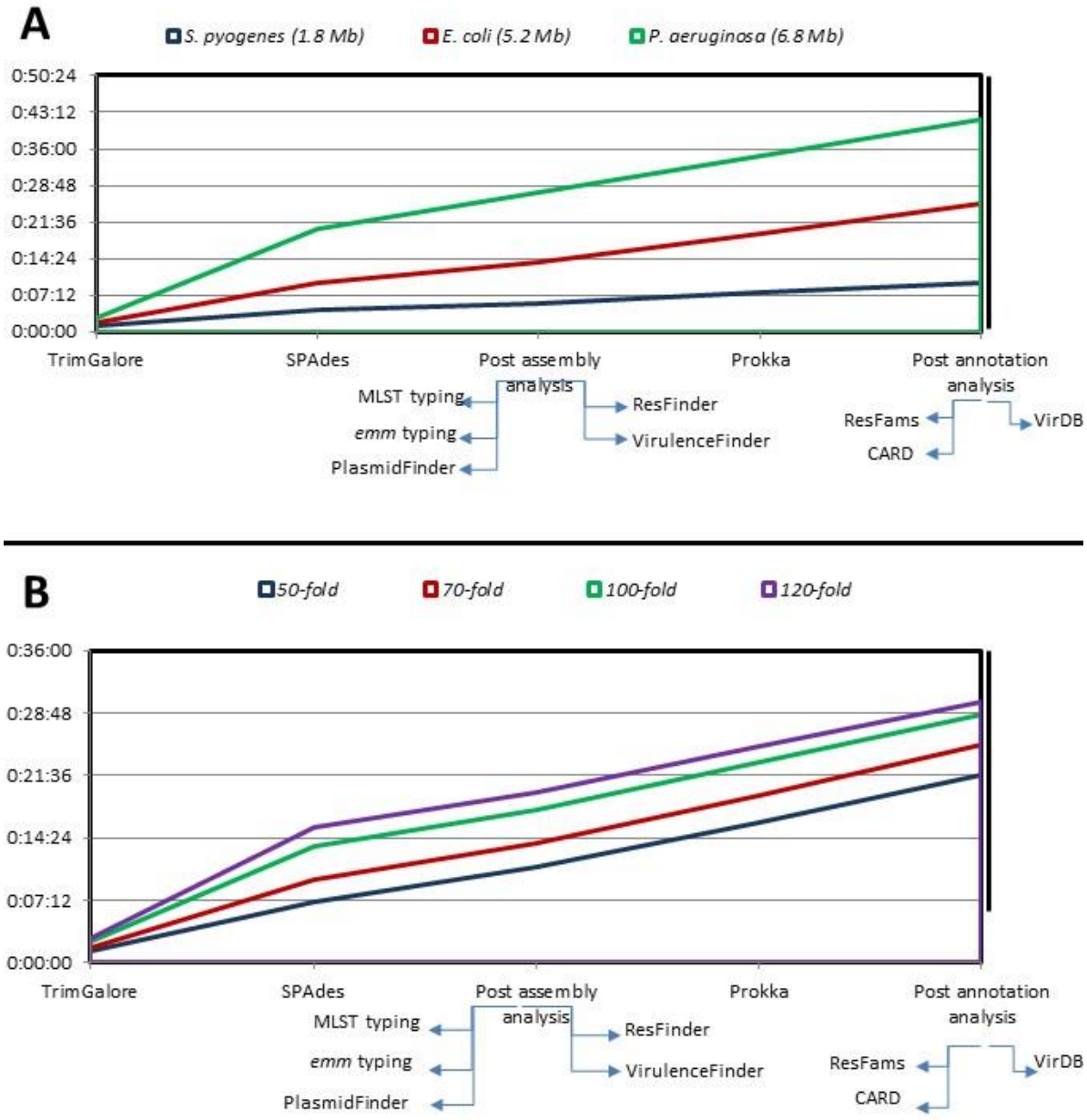


Figure 2.2.3 Impact of different genome sizes at equal sequencing coverage (70-fold) on the computational time taken for each analysis step in BacPipe (A). Impact of varying sequencing coverage of an *E. coli* genome on the computational time taken for each analysis step in BacPipe.

2.2.8.2 Validation of BacPipe's functionality using prior published data

We challenged BacPipe with various bacterial genomes including those with higher GC content and mutliple repeat regions (*M. tuberculosis*). Essentially, five previously published and analysed WGS datasets (raw reads or assembled contigs) from hospital outbreaks caused by MRSA and carbapenem-resistant *K. pneumoniae* (Snitkin, et al. 2012), a 3-year long in-hospital transmission study of *C. difficile* (Jia, et al. 2016), a community-based surveillance and transmission study of *M. tuberculosis* (Kohl, et al. 2014), and finally a food-borne outbreak caused by *S. enterica* (Taylor, et al. 2015) were utilized. We attempted to recreate the same analyses as reported in the respective publications using the 'one-stop'analysis in BacPipe.

2.2.8.3 Outbreak dynamics of MRSA in an academic hospital of Paramaribo, Republic of Suriname

The recent work of Sabat *et al* reported an investigation of an MRSA outbreak at the Academic Hospital Paramaribo (AZP), Suriname from April to May 2013. The outbreak included 12 patients and one healthcare worker/nurse at the AZP totaling 24 isolates that were used to investigate phylogenetic relatedness and transmission (Sabat, et al. 2017). In this study, isolates were sequenced on the MiSeq (V3 kit), and downstream analysis were done using commercial software SeqMan NGen, and SeqMan Pro (DNASTAR Inc., USA). Annotation was done using NCBI prokaryotic genome annotation pipeline (PGAP) (Tatusova, et al. 2016), and MLST, acquired resistance genes, and SNP analyses were performed using the CGE (http://genomicepidemiology.org/) tools. The data is available under Bioproject accession number PRJNA312385.

We analysed all raw reads belonging to 24 isolates and 63 plasmids from this study using BacPipe and produced same results. Firstly, we constructed a SNP-based phylogenetic tree similar to that of Sabat *et al* (Sabat, et al. 2017), consisting of six distinct clusters (A─F) and one singleton (SUR7) (Figure 2.2.4A & B). Secondly, the pipeline assessed the MLST of all isolates as ST8, as reported, and confirmed the loss of *splD* and *splE* genes (representing important virulence factors) from Cluster F (Supplementary File). Similar to what was reported, antibiotic resistance patterns of all isolates showed the presence of *dfrG* trimethoprim resistance- with exception to Cluster E while the *ermC* gene -conferring resistance to clindamycin- was identified

in all isolates of ClusterF and two of ClusterA isolates. For the remainder, it was possible to confirm identical resistance profiles found in all isolates to the previously reported ones including *blaZ, mecA, ermC, aphA3, str, msrA,* and *mphC* genes. Thus, utilizing our pipeline, it was possible to reveal a heterogeneous population structure during this outbreak driven by the different body sites of the same patient or existence of a direct transmission between patients. We also identified a few additional virulence factors *ssp*, *atl*, *efb*, *esa*, in all strains possibly due to the additional database VFDB in BacPipe (Chen, et al. 2016) (Supplementary File).

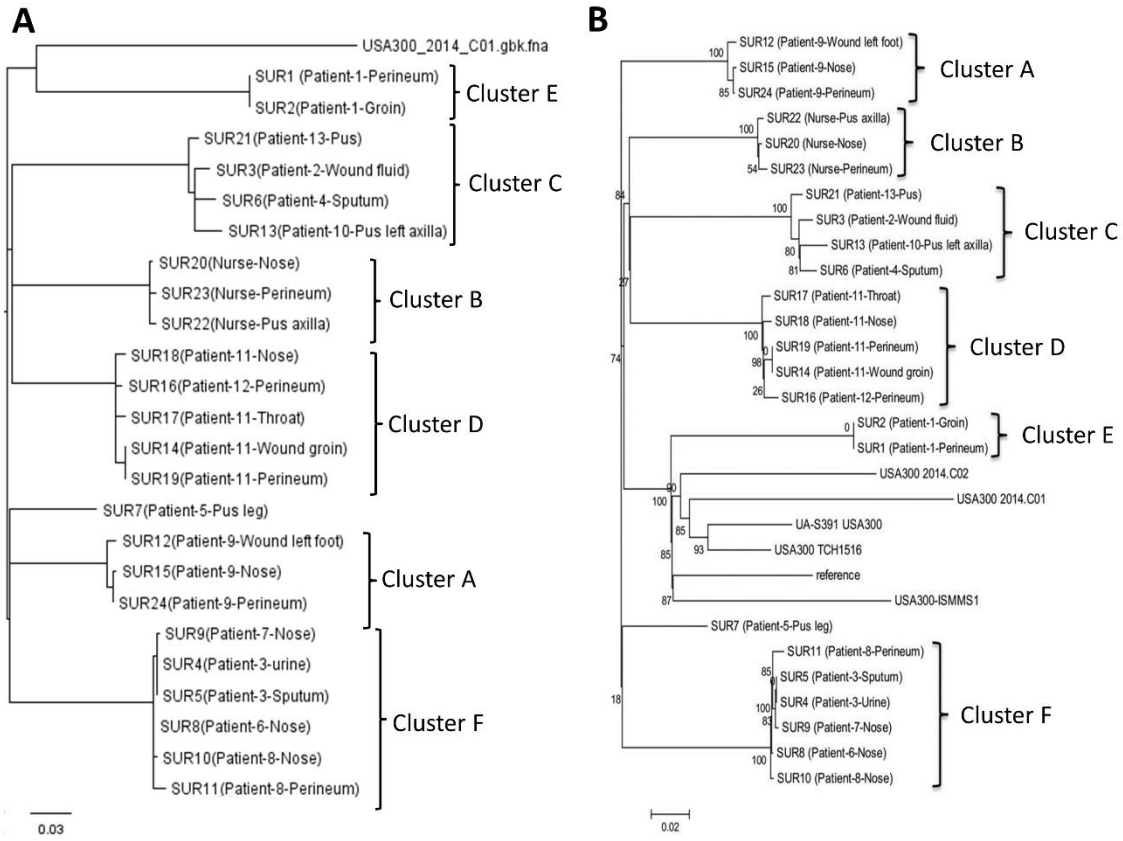


Figure 2.2.4 Phylogenetic maximum likelihood tree generated from core-genome SNPs generated through BacPipe and visualized by TreeView tool (A), and from Sabat et al (Sabat, et al. 2017) (B). The scale bar indicates the evolutionary distance between the sequences determined by 0.1 substitutions per nucleotide at the variable positions.

2.2.8.4 Tracking a Hospital Outbreak of Carbapenem-Resistant *Klebsiella pneumoniae*

Snitkin *et al* described a carbapenem-resistant *K. pneumoniae (CRE) outbreak* in 2011 at the US National Institutes of Health Clinical Center that affected 18 patients of whom 11 died (Snitkin, et

al. 2012). The first patient colonized with CRE was placed under contact isolation and treated, yet after 3 weeks of her discharge, one new case of colonization or active infection was detected every week at the center totalling up to 17 patients. To answer the central question whether patient 1 had initiated the outbreak and if so, how was she linked to the other affected patients, CRE isolated from the 18 patients' samples were analysed by WGS Roche/454 XLR instrument, (Roche Life Sciences). Assembly and annotation was done using gsAssembler and NCBI PGAP, respectively (Snitkin, et al. 2012). The data is available under Bioproject accession number PRJNA73841.

The eighteen strain sequences were processed through BacPipe. As reported in the study, all 18 CRE belonged to the epidemic ST258 clone and harboured *bla*$_{\text{KPC-3}}$. SNP-based phylogenetic construction showed two large clusters and a third cluster consisting only of patient 8 and demonstrated that patient 1 was not only linked to the outbreak but also that three independent transmissions of genetically distinct isolates occurred from patient 1 to other patients (Figure 2.2.5A & B).

Additional antibiotic resistance genes such as *bla*$_{\text{SHV}}$, *bla*$_{TEM}$, *bla*$_{OXA}$, *fosA*, *mphA*, *catA*, *oqxA*, *oqxB*, *sul1*, *dfrA12*, *aadA2* were also identified in the isolates as were plasmid types *IncFII(K)*, *IncFIB (pQil)*, *IncFIB (K)*, and *ColRNAI*, and a virulence gene, *cii* (Supplementary File).

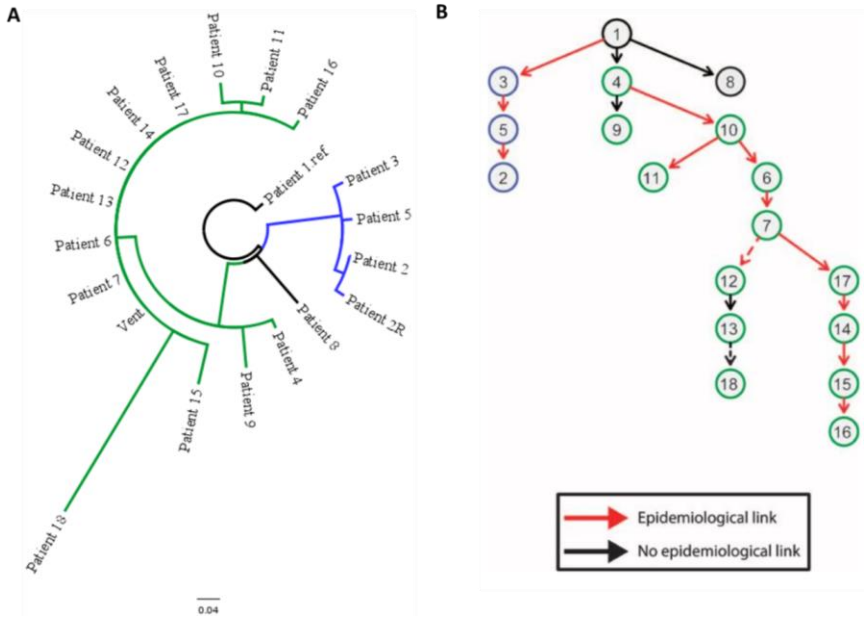


Figure 2.2.5 Phylogenetic maximum likelihood tree generated through BacPipe and visualized by TreeView tool (A). Putative map of *K. pneumoniae* transmission during outbreak reproduced from Snitkin *et al* (Snitkin, et al. 2012). Nodes represent patients, and arrows indicate a transmission event directly or indirectly from one patient to another (B).

2.2.8.5 Tracing nosocomial transmission of *Clostridium difficile* ribotype 027 in a Chinese hospital, 2012–2014

In the study by Jia *et al.* 2016 (Jia, et al. 2016), a rare case of *C. difficile* bloodstream infection (CDBI) was identified. Consequently, all cases or strains that had emerged from the same ward during the past three years were retrospectively analyzed by WGS. Of the 75 patients presenting with diarrhea, *C. difficile* was isolated from 20 patients including the case with CDBI. Isolates were sequenced on the HiSeq platform, reads were mapped to R20291 (NAP1/BI/027, ST1) reference strain using the REALPHY tool (Bertels, et al. 2014), and the phylogenetic tree was reconstructed by the BEAST tool while the genomic SNP differences between strains were detected using SOAP2 (Li, et al. 2009). The data is available under Bioproject accession number PRJNA271048.

BacPipe analysis was able to reproduce the MLST results, where the isolates were characterised into five STs: ST1 (11 patients), ST2 (2 patients), ST8 (2 patients), ST37 (2 patients), and), and ST81 (3 patients) (Supplementary File). From the SNP-based phylogenetic analysis, we confirmed the finding of Jia *et al* of a clear separation between isolates of different STs and that all ST1 isolates were monoclonal (Figure 2.2.6A & B).

Additional data not reported in this study but generated through BacPipe was as follows: *aac(6')-aph(2'')* gene conferring aminoglycoside resistance and *erm(B)* conferring macrolide resistance were identified in all isolates belonging to ST1, ST38, and ST81, while *tet(M)* conferring tetracycline resistance was identified in isolates belonging to ST37 and ST81. Additionally, for all isolates belonging to ST1, we were able to identify *rep1* plasmid, that was not detected in the other non-ST1 isolates (Supplementary File).

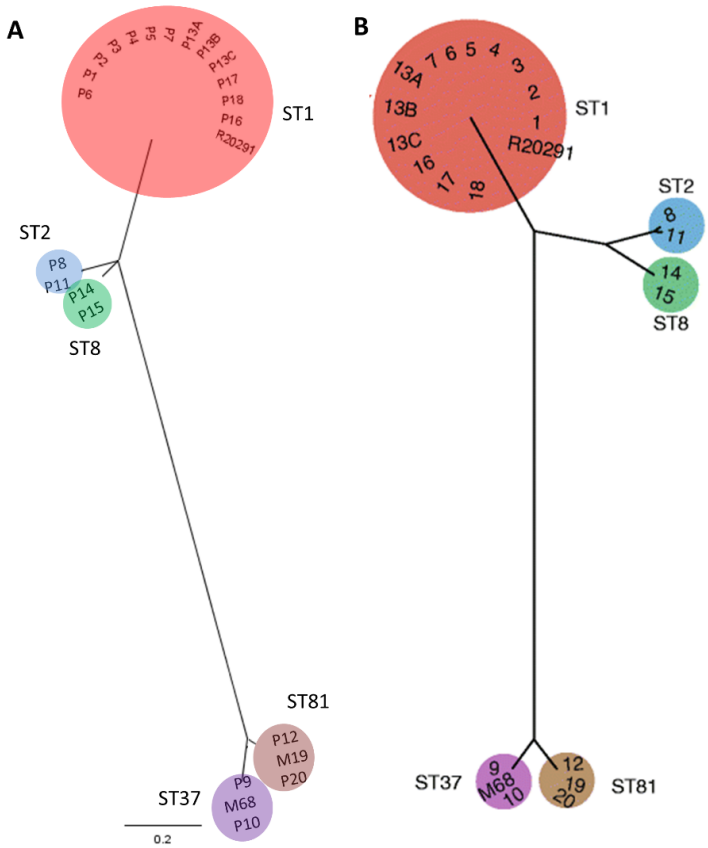


Figure 2.2.6 Phylogenetic maximum likelihood tree of *C. difficile* generated through BacPipe and visualized by TreeView tool (A), and tree reconstructed from multi mapping files via Bayesian evolutionary analysis by BEAST from Jia *et al* (Jia, et al. 2016) (B).

2.2.8.6 Whole genome-based surveillance of *Mycobacterium tuberculosis*

Kohl *et al* trace a *M. tuberculosis* complex *(*MTBC) longitudinal outbreak compromising 26 isolates (between 2001 and 2010) showing identical *IS 6110* DNA fingerprint and spoligotype patterns. These underwent WGS using MiSeq (Illumina), reads were mapped to the H37Rv reference genome using the exact alignment program SARUMAN and SNPs were extracted from the mapped reads by customized Perl scripts (Kohl, et al. 2014). Raw reads are available on Bioproject accession number PRJEB6276.

Using BacPipe, we confirmed that 22 isolates were grouped into one major cluster while four were outliers. Within the primary cluster, we also confirmed two sub-groups compromising of three and six strains, SNP-N1 and SNP-N2, respectively (Figure 2.2.7 A, B). Additional data not reported in this study but generated through BacPipe was an assessment of antibiotic resistance where all isolates harbored *aac(2')-Ic* and subclass B1 beta-lactamase genes conferring aminoglycoside and beta-lactam resistance, respectively (Supplementary File).

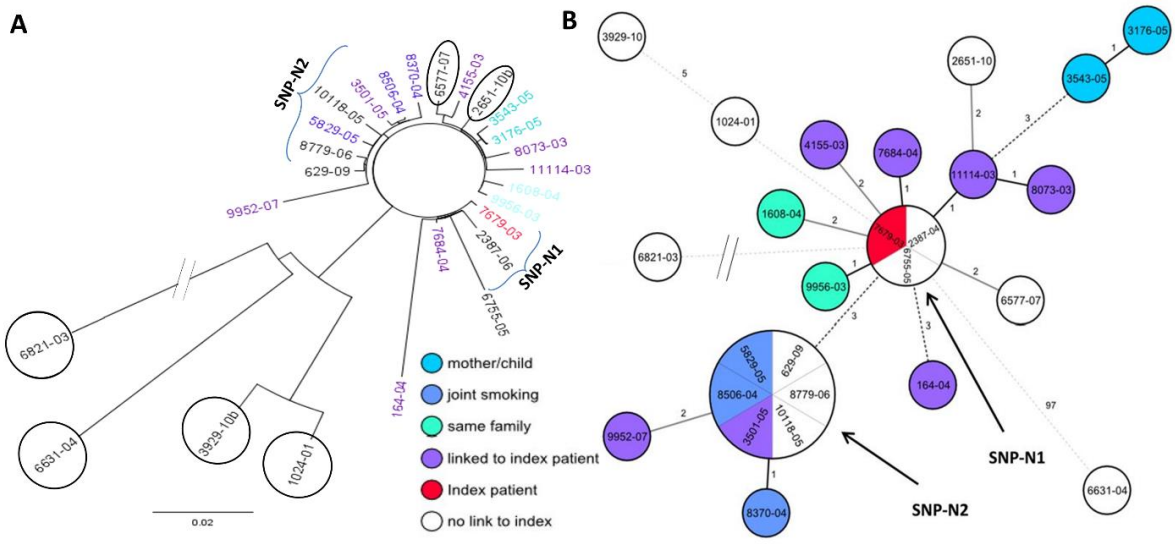


Figure 2.2.7 Phylogenetic maximum likelihood tree of *M. tuberculosis* core-genome SNPs generated through BacPipe and visualized by TreeView tool (A), and a minimum spanning tree of concatenated sequences of the 322 SNPs of the same data from Kohl *et al* (Kohl, et al. 2014) (B).

2.2.8.7 Characterization of foodborne outbreaks of *Salmonella enterica* serovar Enteritidis with whole genome sequencing for surveillance and outbreak detection

Taylor *et al.* describe the application of whole genome sequencing for the detection of *S. enterica* serovar Enteritidis outbreaks from isolates previously characterized by PFGE in Minnesota and Ohio between 2001 to 2014 (Taylor, et al. 2015). The cohort contained 28 isolates from seven epidemiologically confirmed foodborne outbreaks and 27 non-epidemiologically linked sporadic isolates that were assessed by WGS (MiSeq v2). Reads were mapped to the reference genome using BWA-MEM tools, which were later sorted and de-duplicated by the Picard tool. The variant call file (VCF) was produced with BCF tool, and the maximum-likelihood phylogenetic tree was calculated with PhyML. Raw reads are available on accession number: PRJNA237212.

BacPipe was able to retrieve the same phylogenetic tree, confirming that all isolates within the same outbreak were closely related (ranging from 2-7 isolates per outbreak) (Figure 8A). We derived the same conclusion as the original study, that the serovar Enteritidis shows little genetic diversity in the host over time, from investigating isolates MDH-2014-00222, MDH-2014-00223, MDH-2014-00225, and MDH-2014-00228 that were isolated from an individual over a 5-week period (see outbreak 2 in Figure 2.2.8B).

Additional data not reported in this study but generated through BacPipe was the assignment of all strains as ST11, and identification of *IncFIB(S)* and *IncFII(S)* plasmids within all isolates, with exception of MDH-2014-00232, MDH-2014-00245 and outbreak 6/7 isolates –where

no plasmids were detected- and MDH-2014-00215 and MDH-2014-00247 isolates –where *IncHI1B*, *Incl1*, *IncHI1A* and *IncFIA(HI1)* plasmids were identified. Similarly, from MDH-2014-00215 and MDH-2014-00247, antibiotic resistance genes such as bla_{TEM-1B}, *catA1*, *sul1*, *tet(B)*, and *dfrA7* were identified whereas none were found in the remaining isolates (Supplementary File).

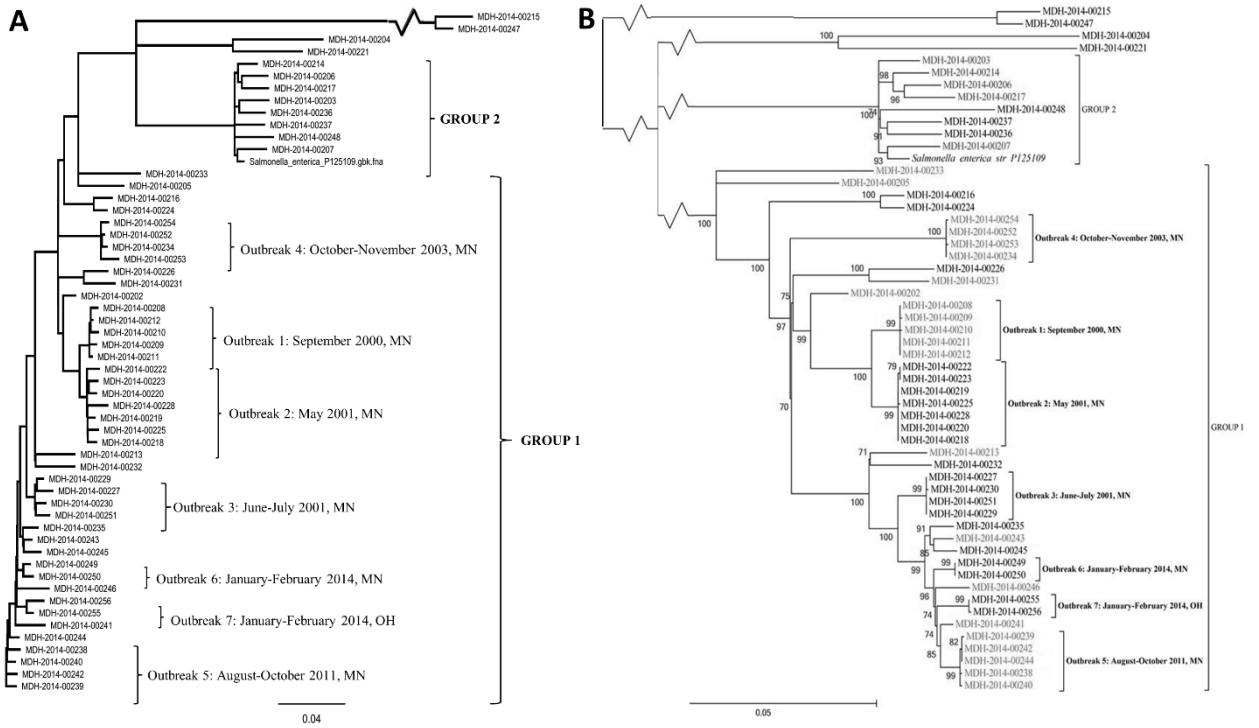


Figure 2.2.8 Maximum-likelihood tree of *S. enteritidis* produced by SNP analysis showing outbreak clusters and time frame (month[s] and year) and the State from where each isolate originated. Phylogenetic analysis generated through BacPipe and visualized by TreeView tool (A), and tree reproduced from Taylor *et al* (Taylor, et al. 2015)(B).

2.2.9 Conclusion

Here we have introduced BacPipe, a bacterial whole genome sequencing analysis pipeline and demonstrated its robustness in handling diverse genomes of clinically important pathogens characterized by different sizes, GC content, and presence of repeat regions that are challenging for downstream data analysis. Along with being comprehensive and modular, BacPipe has the advantage of being computationally low-resource as pipeline functionality does not require an internet connection or high-end computers. A simple graphical interface makes it very user-friendly; a user can direct the software to input and output folder locations, specify the tools to be included in the analysis and adjust the database/parameters from a drop-down list or buttons. BacPipe can run with raw reads from various sequencing platforms and start or pick-up the analysis from any step throughout the workflow giving tremendous flexibility. The end-point of the analysis provides

various level of details, from an over-view nutshell comparing the results across all analysed samples, to an Excel sheet with all of the compiled tool results, to very detailed folders dedicated for each tool output and log files. Additionally, the output of BacPipe can easily be used to study the pan-genome and perform comparative genome analysis and to define acquisition and loss of genes through horizontal gene transfer using additional tools such as Roary (Page, et al. 2015) and BPGA (Chaudhari, et al. 2016), the latter also producing the KEGG and COG function and pathway based comparisons. Finally, using raw data or contigs from prior publications that had utilized many different tools to delineate hospital or community-based outbreaks or pathogen transmission, we demonstrated that the collection of tools in BacPipe could reproduce the entire analyses as a 'one-stop' platform within a few hours. We believe this fully automated pipeline will overcome one of the primary barriers to analysing and interpreting WGS data, facilitating applications for routine patient care in hospitals and public health and infection control monitoring. Future development of BacPipe entails expansion of tools to enable identification of prophages, IS (insertion sequence) elements, CRISPR-Cas elements and, depending upon their open-access status, whole-/core-/pan-genome MLST schemes.

Author contribution: BBX was involved in validation, data analysis, interpretation, part of designing study and drafting manuscript

2.2.10 References

1. Kwong JC, McCallum N, Sintchenko V, Howden BP. Whole genome sequencing in clinical and public health microbiology. Pathology. 2015;47(3):199-210.
2. Arnold C. Outbreak Breakthrough: Using Whole-Genome Sequencing to Control Hospital Infection. Environ Health Perspect. 2015;123(11): A281-6.
3. Moran-Gilad J. Whole genome sequencing (WGS) for food-borne pathogen surveillance and control - taking the pulse. Euro Surveill. 2017;22(23).
4. Punina NV, Makridakis NM, Remnev MA, Topunov AF. Whole-genome sequencing targets drug-resistant bacterial infections. Human Genomics. 2015; 9:19.
5. Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW. Transforming clinical microbiology with bacterial genome sequencing. 2012; 13:601.
6. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, et al. Real-Time Whole-Genome Sequencing for Routine Typing, Surveillance, and Outbreak Detection of Verotoxigenic Escherichia coli. Journal of Clinical Microbiology. 2014;52(5):1501-10.
7. Koser CU, Ellington MJ, Cartwright EJ, Gillespie SH, Brown NM, Farrington M, et al. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. PLoS Pathog. 2012;8.
8. Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, García-Cobos S, et al. Application of next generation sequencing in clinical microbiology and infection prevention. Journal of Biotechnology. 2017; 243:16-24.
9. Muir P, Li S, Lou S, Wang D, Spakowicz DJ, Salichos L, et al. The real cost of sequencing: scaling computation to keep pace with data generation. Genome Biology. 2016;17(1):53.
10. Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, et al. Corrigendum: Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis. Nature Communications. 2016; 7:11465.
11. Bodi K. Tools for Next Generation Sequencing Data Analysis. Journal of Biomolecular Techniques: JBT. 2011;22(Suppl): S18-S.
12. Cuccuru G, Orsini M, Pinna A, Sbardellati A, Soranzo N, Travaglione A, et al. Orione, a web-based framework for NGS analysis in microbiology. Bioinformatics. 2014;30(13):1928-9.
13. Thomsen MCF, Ahrenfeldt J, Cisneros JLB, Jurtz V, Larsen MV, Hasman H, et al. A Bacterial Analysis Platform: An Integrated System for Analysing Bacterial Whole Genome Sequencing Data for Clinical Diagnostics and Surveillance. PLOS ONE. 2016;11(6):e0157718.
14. Akgün M, Bayrak AO, Ozer B, Sağıroğlu MŞ. Privacy preserving processing of genomic data: A survey. Journal of Biomedical Informatics. 2015; 56:103-11.
15. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455-77.
16. Xavier BB, Sabirova J, Pieter M, Hernalsteens J-P, de Greve H, Goossens H, et al. Employing whole genome mapping for optimal de novo assembly of bacterial genomes. BMC Research Notes. 2014; 7:484.
17. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, et al. Multilocus

Sequence Typing of Total-Genome-Sequenced Bacteria. Journal of Clinical Microbiology. 2012;50(4):1355-61.
18. 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. Antimicrobial Agents and Chemotherapy. 2014;58(7):3895-903.
19. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. Journal of Antimicrobial Chemotherapy. 2012;67(11):2640-4.
20. Facklam R, Beall B, Efstratiou A, Fischetti V, Johnson D, Kaplan E, et al. emm typing and validation of provisional M types for group A streptococci. Emerging Infectious Diseases. 1999;5(2):247-53.
21. 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 Biology. 2014;15(11):524.
22. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30(14):2068-9.
23. McArthur AG, Waglechner N, Nizam F, Yan A, Azad MA, Baylay AJ, et al. The Comprehensive Antibiotic Resistance Database. Antimicrobial Agents and Chemotherapy. 2013;57(7):3348-57.
24. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Research. 2017;45(D1): D566-D73.
25. Gibson MK, Forsberg KJ, Dantas G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. The ISME Journal. 2015;9(1):207-16.
26. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. Nucleic Acids Research. 2016;44(D1): D694-D7.
27. Sabat AJ, Hermelijn SM, Akkerboom V, Juliana A, Degener JE, Grundmann H, et al. Complete-genome sequencing elucidates outbreak dynamics of CA-MRSA USA300 (ST8-spa t008) in an academic hospital of Paramaribo, Republic of Suriname. Scientific Reports. 2017; 7:41050.
28. Snitkin ES, Zelazny AM, Thomas PJ, Stock F, Henderson DK, Palmore TN, et al. Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. Sci Transl Med. 2012;4.
29. Jia H, Du P, Yang H, Zhang Y, Wang J, Zhang W, et al. Nosocomial transmission of Clostridium difficile ribotype 027 in a Chinese hospital, 2012–2014, traced by whole genome sequencing. BMC Genomics. 2016;17(1):405.
30. Kohl TA, Diel R, Harmsen D, Rothgänger J, Walter KM, Merker M, et al. Whole-Genome-Based Mycobacterium tuberculosis Surveillance: a Standardized, Portable, and Expandable Approach. Journal of Clinical Microbiology. 2014;52(7):2479-86.
31. Taylor AJ, Lappi V, Wolfgang WJ, Lapierre P, Palumbo MJ, Medus C, et al. Characterization of Foodborne Outbreaks of Salmonella enterica Serovar Enteritidis with Whole-Genome Sequencing Single Nucleotide Polymorphism-Based Analysis for Surveillance and Outbreak Detection. Journal of Clinical Microbiology. 2015;53(10):3334-40.

32. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 2016;44(14):6614-24.
33. Bertels F, Silander OK, Pachkov M, Rainey PB, van Nimwegen E. Automated Reconstruction of Whole-Genome Phylogenies from Short-Sequence Reads. Molecular Biology and Evolution. 2014;31(5):1077-88.
34. Li R, Yu C, Li Y. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics. 2009;25.
35. 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(22):3691-3.
36. Chaudhari NM, Gupta VK, Dutta C. BPGA- an ultra-fast pan-genome analysis pipeline. 2016; 6:24373.

2.3 WHOLE GENOME MAPPING AS UTILITY TOOL TO STUDY GENOMIC STABILITY AND VARIATIONS IN HUMAN PATHOGENS

Julia S Sabirova, **Basil Britto Xavier,** Margareta Ieven, Herman Goossens, and Surbhi Malhotra-Kumar

Laboratory of Medical Microbiology, Vaccine & Infectious Disease Institute, Universiteit Antwerpen, Antwerp, Belgium

Abstract

Whole genome (optical) mapping (WGM), a state-of-the-art mapping technology based on the generation of high resolution restriction maps, has so far been used for typing clinical outbreak strains and for mapping *de novo* sequence contigs in genome sequencing projects. We employed WGM to assess the genomic stability of previously sequenced *Staphylococcus aureus* strains that are commonly used in laboratories as reference standards. *S. aureus* strains (n = 12) were mapped on the Argus™ Optical Mapping System (Opgen Inc, Gaithersburg, USA). Assembly of *NcoI*-restricted DNA molecules, visualization, and editing of whole genome maps was performed employing MapManager and MapSolver softwares (Opgen Inc). *In silico* whole genome *NcoI*-restricted maps were also generated from available sequence data and compared to the laboratory-generated maps. Strains showing differences between the two maps were resequenced using Nextera XT DNA Sample Preparation Kit and Miseq Reagent Kit V2 (MiSeq, Illumina) and *de novo* assembled into sequence contigs using the Velvet assembly tool. Sequence data were correlated with corresponding whole genome maps to perform contig mapping and genome assembly using MapSolver. Of the twelve strains tested, one (USA300_FPR3757) showed a 19-kbp deletion on WGM compared to its *in silico* generated map and reference sequence data. Resequencing of the USA300_FPR3757 identified the deleted fragment to be a 13kbp-long integrative conjugative element *ICE6013*. Frequent sub culturing and inter-laboratory transfers can induce genomic and therefore, phenotypic changes that could compromise the utility of standard reference strains. WGM can thus be used as a rapid genome screening method to identify genomic rearrangements whose size and type can be confirmed by sequencing.

2.3.1 Background

The plasticity of the bacterial genome is well-known and can be attributed to horizontal gene transfer, genome rearrangements, and the activities of mobile genetic elements (MGEs) [1]. In the case of *Staphylococcus aureus*, and also for several other pathogens, MGE DNA accounts for a substantial part of the identified inter-strain genetic variability in terms of acquisition of virulence, colonization or antibiotic resistance factors. Such MGE-associated changes have also been reported in frozen, archived *S. aureus* strains [2], and are especially problematic for maintenance of reference culture collections utilized for inter- and intra-laboratory experimental validations.

Whole genome (optical) mapping (WGM) is an advanced molecular technology by which endonuclease-digested single DNA molecules are assembled into a high-resolution restriction map constituting a detailed genomic fingerprint. Originally developed in 1993 by Schwartz and colleagues to type *Saccharomyces cerevisiae* strains [3], WGM has so far mainly been applied in clinical microbiology for typing outbreak strains [4, 5, 6, 7], and for genome sequencing to guide *de novo* assembly of sequence contigs [8]. Here, we applied WGM to rapidly assess genomic stability of key reference strains of *Staphylococcus aureus* utilizing their previously published whole genome sequences as comparators.

2.3.2 Methods

2.3.2.1 Bacterial strains and growth conditions

Twelve *S. aureus* strains that are commonly used as reference strains and whose whole genome sequences are available in public databases were requested from strain collections (Network on Antimicrobial Resistance in *S. aureus* (NARSA, http://www.beiresources.org/), or from research groups: COL (Acc.no. NC_002951), 71193 (Acc.no. NC_017673), FPR3757 (Acc.no. NC_007793), ED133 (Acc.no. NC_017337), JKD6008 (Acc.no. NC_017341), JKD6159 (Acc.no. NC_017338), LGA251 (Acc.no. NC_017349), MRSA252 (Acc.no. NC_002952), MRSA476 (Acc.no. NC_002953), N315 (Acc.no. NC_002745), TCH1516 (Acc.no. NC_010079), and HO 5096 0412 (Acc.no. NC_017763). Upon arrival, bacterial stocks were inoculated on blood agar plates and incubated overnight at 37°C. 16 hour-old colonies were used for species re-confirmation by Maldi-TOF and to prepare glycerol stocks for long-term storage at -80°C. In parallel, one of the 16 hour-old colonies were inoculated on brain heart infusion agar, incubated overnight at 37°C and used for WGM.

2.3.2.2 Genome sequencing

S. aureus strains were grown on brain heart infusion plates for 24 h at 37°C and used to extract genomic DNA using MasterPure™ Gram Positive DNA Purification Kit (Epicentre Technologies Corp). Samples were prepared and sequenced employing Nextera XT DNA Sample Preparation Kit and Miseq Reagent Kit V2 (Illumina Inc). The 2 X 150 reads were *de novo* assembled into contigs using CLC Genomics Workbench with default parameters and Velvet v1.2.10 assembler [14] with an optimised N50 *k-mer* size of 93 [10]. Sequencing data were correlated with corresponding whole genome maps to perform contig mapping in MapSolver software (Opgen Inc). Gene annotation of ICE elements was performed using RAST annotation server and blast2go for functional annotation.

2.3.2.3 Whole genome mapping

Maps were generated on an Argus™ Optical Mapping System (Opgen Inc, Gaithersburg, USA). DNA extraction, quality control, restriction using *NcoI*, and loading on a mapcard were done according to manufacturer's protocols. Briefly, *S. aureus* colonies grown overnight at 37°C on brain heart infusion plates were used for isolation of high molecular weight (HMW) DNA using Argus® HMW DNA isolation kit (Opgen, Inc). Extracted DNA preps were checked for the presence of HMW DNA molecules by using Argus® QCard kit (Opgen, Inc) and subsequently used for mapping employing Argus® MapCard II kit (Opgen, Inc). The assembly of restricted DNA molecules and identification of novel *NcoI* restriction sites was performed employing MapManager software (Opgen Inc.). Visualization and editing of maps was performed using MapSolver software (Opgen Inc.). For map editing, whole genome maps were adjusted in size, orientation and in starting point employing *in silico* maps generated from corresponding genome sequence data using MapSolver's function "Comparative genomics" and an advanced parameter "Allow alignments to wrap circular maps" with other default parameters pre-set in MapSolver. Map clustering was also performed using MapSolver's default parameters employing MapSolver's function "Create cluster".

2.3.3 Results and discussion

We utilized WGM to assess the genomic stability of previously sequenced *S. aureus* that are also commonly used as laboratory reference strains (Additional file 1: Table S1). Comparison of experimental whole genome maps to *in silico* restriction maps (*NcoI*) generated from published sequences showed that for majority of the strains (COL, MRSA476, 71193, ED133, JKD6008,

JKD6159, LGA251, MRSA252, N315, TCH1516, and HO 5096 0412), both maps were identical (100–99.9% map similarity; Figure 1). The 0.1% difference in *in silico* and experimental maps of some strains (COL, MRSA476, 71193, ED133, JKD6008, LGA251, MRSA252, N315, and TCH1516) could be attributed to the loss of small DNA fragments in the experimental maps. The rate of loss of fragments ≤ 2 kb could be as high as 75% during mapcard processing comprising *in situ* restriction, staining and washing of linearized and immobilized DNA molecules, as described previously [9]. However, for *S. aureus* FPR3757, a clear 1.5% map difference was observed between the experimental and corresponding *in silico* WGMs. The experimental WGM of FPR3757 displayed a 19-kbp genomic deletion compared to its *in silico* map (Figure 2). According to published sequence data, the deleted fragment contained genes SAUSA300_1456 to SAUSA300_1474 that included a maltose degradation operon (Figure 3A).

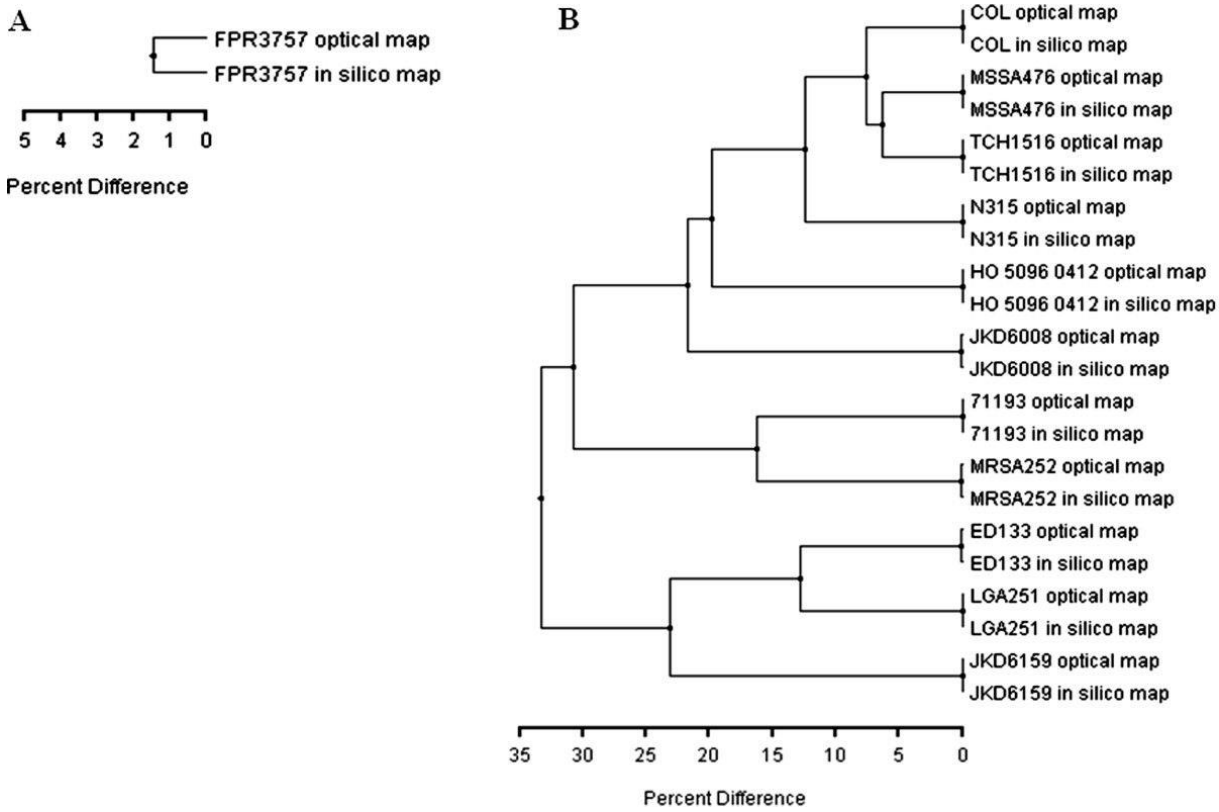


Figure 2.3.1 Map similarity cluster of experimental whole genome maps and corresponding *in silico* maps for *S. aureus* strains constructed using unweighted-pair group method with arithmetic averages (UPGMA). Whole genome map of FPR3757 features 1.5% map distance **(A)**, whereas whole genome maps of other strains feature 0–0.1% map distance compared to their corresponding *in silico* maps **(B)**.

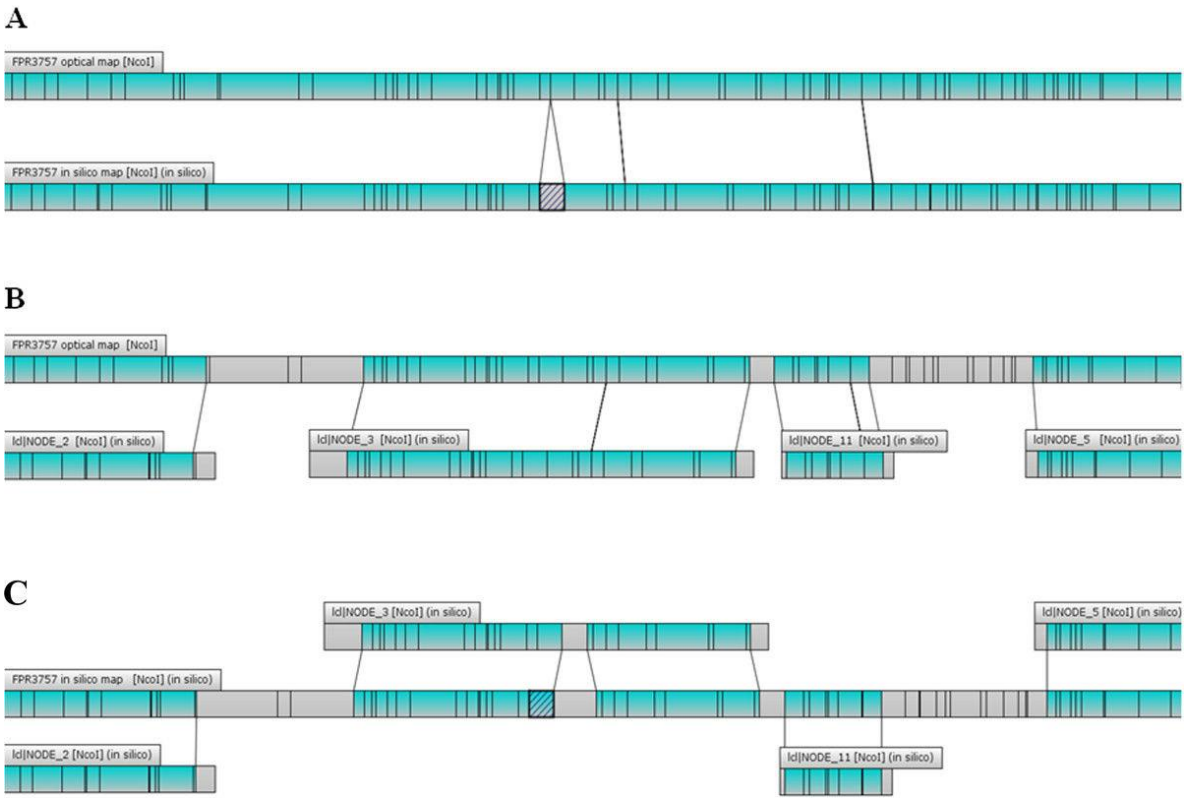


Figure 2.3.2 Aligned experimental and *in silico* whole genome maps of FPR3757 (A); alignment of experimental map of FPR3757 to sequence contigs (B); alignment of *in silico* map of FPR3757 to sequence contigs (C). Regions highlighted in green indicate perfect alignment. The region corresponding to the deletion is colored in grey and hashed with dark grey diagonal bars in the *in silico* map.

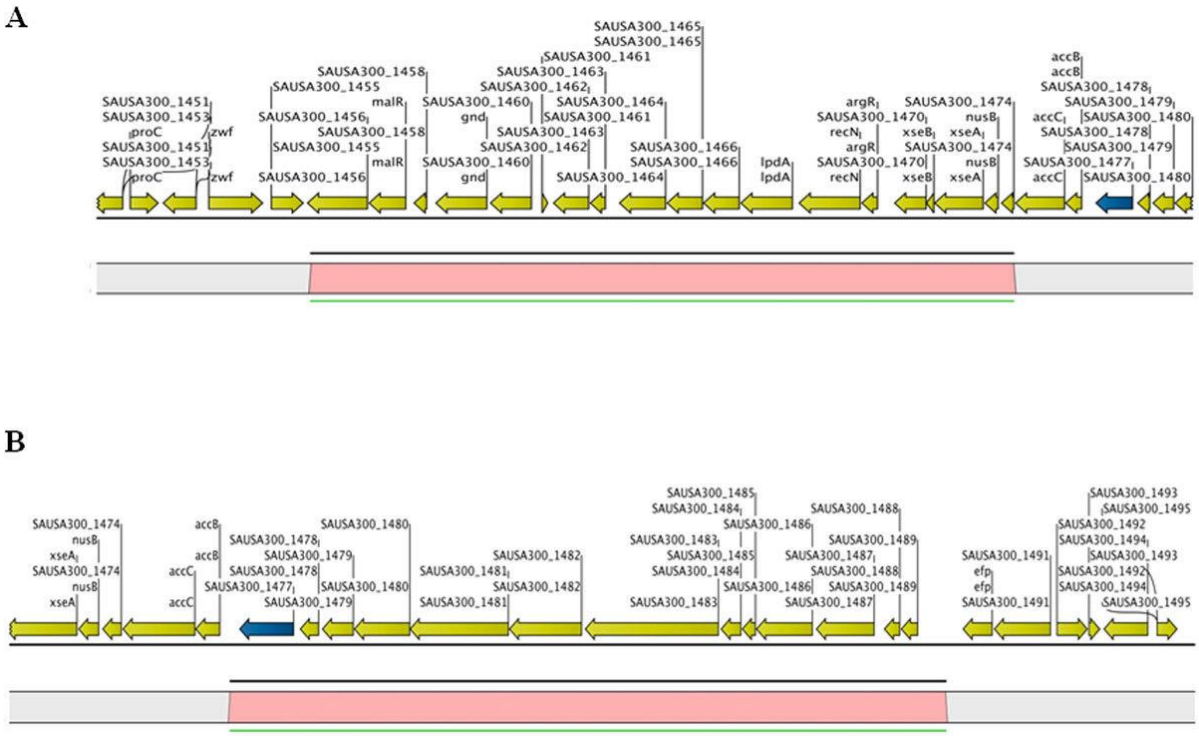


Figure 2.3.3 Genomic deletion in FPR3757 according to WGM (A), and whole genome (re) sequencing (B). All open reading frames are highlighted in yellow; transposase gene is highlighted in blue. Region corresponding to the genomic deletion is highlighted in pink.

In order to confirm the genomic deletion detected by WGM, the entire genome of FPR3757 was sequenced, *de novo* assembled into sequence contigs by Velvet with an optimized *k*-mer [10] and compared to its corresponding published genome using MapSolver. Re-sequencing of FPR3757 showed that the deleted fragment was 13354 bp long (nucleotides 1630722 and 1644075) and comprised genes SAUSA300_1477 to SAUSA300_1489 (Figure 3B). The observed difference in prediction of the site of deletion by WGM and sequencing is likely to be due to the loss of small fragments during preparation of whole genome maps. Indeed, upon inspection of the FPR3757 map, the total number of restriction fragments was found to be remarkably lower than in the *in silico* map (219 and 244 fragments, respectively). Loss of fragments shorter than 2 kb can result in a certain degree of imprecision in prediction of the size of restriction fragments and localization of *NcoI* sites in the whole genome map, as compared to the *in silico*-generated map.

BLAST searches against *S. aureus* published genomes using the nucleotide sequence of the deleted fragment as a query identified an integrative conjugative element *ICE6013*, which was recently described as a novel MGE in a number of MRSA [11]. The MGE in FPR3757 showed 99% pairwise identity to the *ICE6053* of *S. aureus* strain HDG2 (accession number FJ231270), except

for the absence of a 6551-bp *Tn552* transposon. Comparison of the two ICEs employing blast2go and RAST functional comparison tools allowed to assign putative functions to the genes and to compare their genetic contents: the absence of *Tn552* in the ICE in FPR3757 resulted in the loss of a β-lactamase operon comprising *blaI*, *blaR1* and *blaZ* genes, whereas other genes typical for ICEs and potentially responsible for their transposition, replication and conjugative transfer were found to be intact [12, 13] (Figure 4, Additional file 2: Table S2).

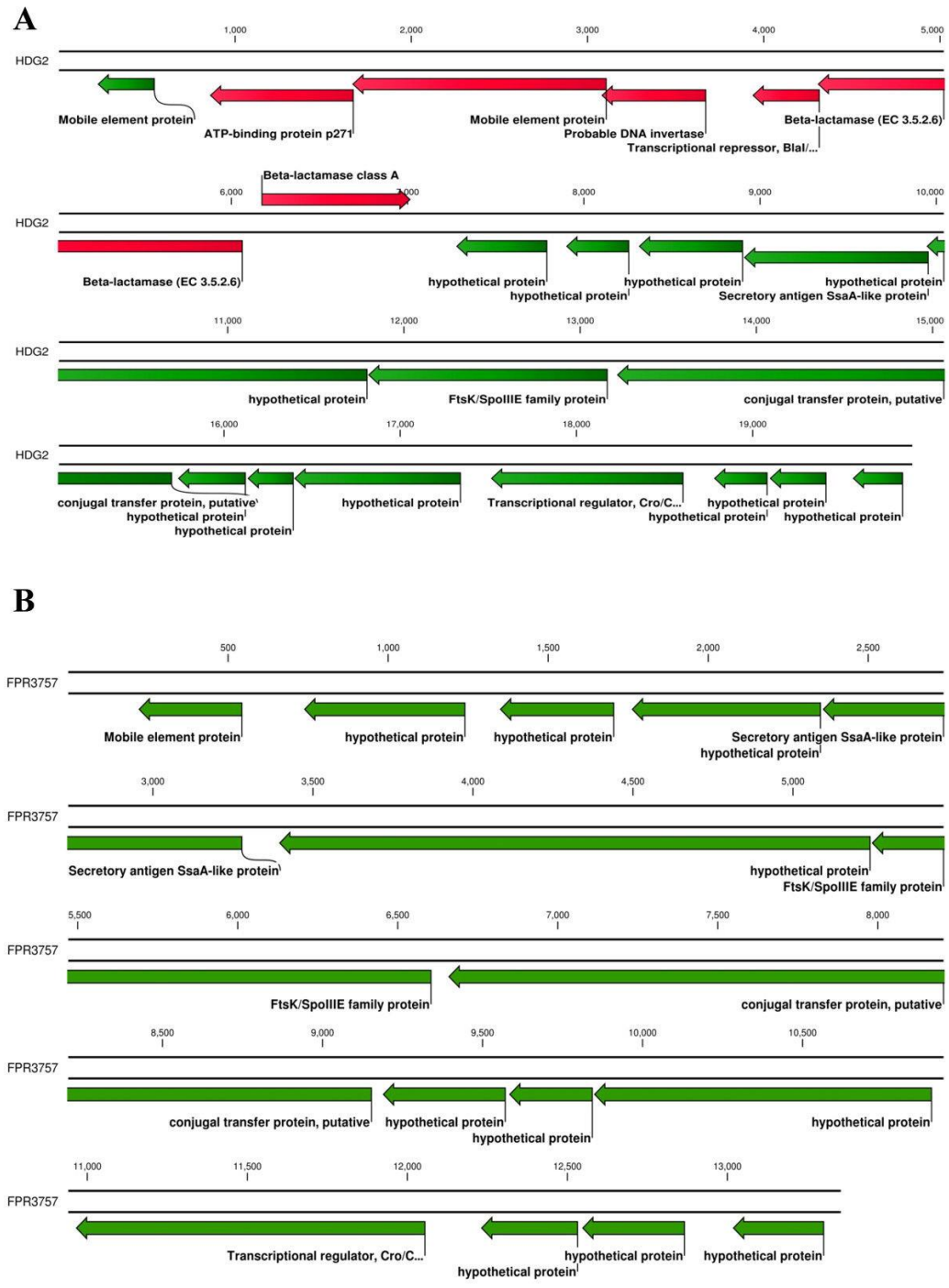


Figure 2.3.4 Detailed view of *ICE6053* of *S. aureus* strain HDG2 (A) and of *ICE6053* of *S. aureus* strain FPR3757 (B). Gene annotation was performed using RAST annotation server and blast2go. Genes highlighted in red in HDG2 represent those deleted in FPR3757.

2.3.4 Conclusions

In conclusion, WGM of sequenced *S. aureus* strains detected a large genomic deletion in one out of twelve strains studied here. Subsequent genome sequencing of the altered genome and inspection of the region identified the exact size, genomic location and the nature of this genomic deletion. WGM can thus be employed to rapidly assess genomic stability of strains over time during storage or after inter-laboratory transport followed by corroboration of genomic changes by whole genome sequencing.

Author contribution: BBX was involved in experimental work, data analysis, interpretation and drafting manuscript.

2.3.5 References

1. Darmon ELD: Bacterial genome instability. Microbiol Mol Biol Rev. 2014, 78 (1): 1-39. 10.1128/MMBR.00035-13.
2. Hiramatsu K, Sukuzi E, Takayama H, Katayama Y, Yokota T: Role of penicillinase plasmids in the stability of the mecA gene in methicillin-resistant *Staphylococcus aureus*. Antimicrob Agents Chemother. 1990, 34 (4): 600-604. 10.1128/AAC.34.4.600.
3. Schwartz DC, Li X, Hernandez LI, Ramnarain SP, Huff EJ, Wang YK: Ordered restriction maps of Saccharomyces cerevisiae chromosomes constructed by optical mapping. Science. 1993, 262: 110-114. 10.1126/science.8211116.
4. Kotewicz ML, Mammel MK, Le Clerc JE, Cebula TA: Optical mapping and 454 sequencing of Escherichia coli O157: H7 isolates linked to the US 2006 spinach-associated outbreak. Microbiology. 2008, 154 (11): 3518-3528. 10.1099/mic.0.2008/019026-0
5. Jackson SA, Kotewicz ML, Patel IR, Lacher DW, Gangiredla J, Elkins CA: Rapid genomic-scale analysis of Escherichia coli O104:H4 by using high-resolution alternative methods to next-generation sequencing. Appl Environ Microbiol. 2012, 78: 1601-1605. 10.1128/AEM.07464-11.
6. Petersen RF, Litrup E, Larsson JT, Torpdahl M, Sørensen G, Müller L, Nielsen EM: Molecular characterization of Salmonella Typhimurium highly successful outbreak strains. Foodborne Pathog Dis. 2011, 8 (6): 655-661. 10.1089/fpd.2010.0683.
7. Johnson PD, Ballard SA, Grabsch EA, Stinear TP, Seemann T, Young HL, Grayson ML, Howden BP: A sustained hospital outbreak of vancomycin-resistant Enterococcus faecium bacteremia due to emergence of vanB E. faecium sequence type 203. J Infect Dis. 2010, 202 (8): 1278-1286. 10.1086/656319
8. Onmus-Leone F, Hang J, Clifford RJ, Yang Y, Riley MC, Kuschner RA, Waterman PE, Lesho EP: Enhanced de novo assembly of high throughput pyrosequencing data using whole genome mapping. PLoS One. 2013, 8 (4): e61762-10.1371/journal.pone.0061762
9. Reslewic S, Zhou S, Place M, Zhang Y, Briska A, Goldstein S, Churas C, Runnheim R, Forrest D, Lim A, Lapidus A, Han CS, Roberts GP, Schwartz DC: Whole-genome shotgun optical mapping of rhodospirillum rubrum. Appl Environ Microbiol. 2005, 71: 5511-5522. 10.1128/AEM.71.9.5511-5522.2005
10. Xavier BB, Sabirova J, Pieter M, Hernalsteens JP, de Greve H, Goossens H, Malhotra-Kumar S: Employing whole genome mapping for optimal de novo assembly of bacterial genomes. BMC Res Notes. 2014, 7: 484-10.1186/1756-0500-7-484.
11. Smyth DS, Robinson D: Integrative and sequence characteristics of a novel genetic element, ICE6013, in *Staphylococcus aureus*. J Bacteriol. 2009, 191 (19): 5964-5975. 10.1128/JB.00352-09.
12. Arends K, Celik EK, Probst I, Goessweiner-Mohr N, Fercher C, Grumet L, Soellue C, Abajy MY, Sakinc T, Broszat M, Schiwon K, Koraimann G, Keller W, Grohmann E: TraG encoded by the pIP501 type IV secretion system is a two-domain peptidoglycan-degrading enzyme essential for conjugative transfer. J Bacteriol. 2013, 195: 4436-4444. 10.1128/JB.02263-12.
13. Parsons JA, Bannam TL, Devenish RJ, Rood JI: TcpA, an FtsK/SpoIIIE homolog, is

essential for transfer of the conjugative plasmid pCW3 in Clostridium perfringens. J Bacteriol. 2007, 189 (21): 7782-7790. 10.1128/JB.00783-07

14. Zerbino DR, Birney E: Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008, 18 (5): 821-829. 10.1101/gr.074492.107

2.4 EMPLOYING WHOLE GENOME MAPPING FOR OPTIMAL DE NOVO ASSEMBLY OF BACTERIAL GENOMES

Xavier BB[1], Sabirova J[1], Pieter M[1], Hernalsteen JP[2], Greve de H[3], Goossens H[1], Malhotra-Kumar S[1]

[1]Department of Medical Microbiology, Vaccine & Infectious Disease Institute, Universiteit Antwerpen, Antwerp, Belgium [2]Viral Genetics Research Group, Vrije Universiteit Brussel, Brussels, Belgium [3]Structural Biology Brussels, Flanders Institute for Biotechnology (VIB), Vrije Universiteit Brussel, Brussels, Belgium

Abstract

De novo genome assembly can be challenging due to inherent properties of the reads, even when using current state-of-the-art assembly tools based on de Bruijn graphs. Often users are not bio-informaticians and, in a black box approach, utilise assembly parameters such as contig length and N50 to generate whole genome sequences, potentially resulting in mis-assemblies. Utilising several assembly tools based on de Bruijn graphs like Velvet, SPAdes and IDBA, we demonstrate that at the optimal N50, mis-assemblies do occur, even when using the multi-k-mer approaches of SPAdes and IDBA. We demonstrate that whole genome mapping can be used to identify these mis-assemblies and can guide the selection of the best k-mer size which yields the highest N50 without mis-assemblies. We demonstrate the utility of whole genome mapping (WGM) as a tool to identify mis-assemblies and to guide *k-mer* selection and higher quality *de novo* genome assembly of bacterial genomes.

2.4.1 Findings

Genome assembly is often a primary step in the process of yielding results that lead to interpretation of biological data and hence sub-optimally assembled genomes might lead to faulty conclusions (Moffatt, et al. 2011). Factors causing such low quality genome assembly include sequence quality, presence of repetitive sequences, base composition, size and low genome coverage (Phillippy, et al. 2008; Salzberg and Yorke 2005), all of which complicate downstream data analysis using the available tools (Dark 2013). Currently, *de novo* assemblers based on de Bruijn graph are considered to yield the best results provided sufficient sequence quality and coverage are achieved. Such assembly tools based on de Bruijn graph algorithms, like Velvet (Zerbino and Birney 2008) and SPAdes (Bankevich, et al. 2012) use *k-mers* as building blocks, but as most users are not bio-informaticians, these tools are often considered as an encrypted black box with the quality of the assembly usually determined by statistics parameters such as the N50 and the size and number of contigs or scaffolds produced by the assemblers (Hunt, et al. 2013). However, the choice of the *k-mer* size is crucial as too low or too high *k-mer* sizes lead to sub-optimal assemblies. Indeed, low quality reads might produce false positive vertices, repeats lead to branching, while an uneven distribution of the reads results in gaps. The use of smaller *k-mers* reduces the problem associated with low quality reads and their uneven distribution, while larger *k-mer* sizes help to bridge repeat regions decreasing the branching problem (Peng, et al. 2010). In a balancing exercise, various *k-mer* sizes are usually selected, evaluating optimization by aiming for high N50 values and long, but fewer contigs. Whole Genome Mapping (WGM; Opgen Inc, Gaithersburg, MD, USA) is a relatively novel technique that generates high-resolution restriction maps of a genome based on the alignment of single DNA molecules cut with restriction enzymes and ordered with high resolution and accuracy (Ananiev, et al. 2008). WGM was proven helpful in mapping *de novo* assembled contigs against previously sequenced related genomes (Onmus-Leone, et al. 2013).

In this paper, we evaluated the utility of WGM for proper k-mer size selection and for optimization of parameters for *de novo* genome assembly. The whole genome sequence of a methicillin-resistant *Staphylococcus aureus* (MRSA) strain (E-MRSA15-CC22-SCC*mec*IV) was generated on an Illumina HiSeq-2000 via 2X150b paired end sequencing (Sabirova, et al. 2014a). Reads were *de novo* assembled using Velvet, SPAdes and IDBA-UD employing a range of *k-mers*. Velvet, using a single *k-mer* approach ranging from *k-mer* size 81 to 123 showed an initial increase in N50 (until *k-mer* size 115) and longest contig size and a concomitant decrease in the total number of contigs with increasing *k-mer* size with both these parameters positively influencing the assembly outcome (Table 2.4.1). Ensuing, using WGM, a whole genome map of *S. aureus* EMRSA-15 was

generated which, using MapSolver, was aligned with the assembly files corresponding to different *k-mer* sizes. Although the percentage of the genome covered by contigs increased with increasing *k-mer* size, several mis-assemblies were identified for the higher *k-mer* sizes (Table 2.4.1), revealing the best (without mis-assemblies) assembly was actually obtained using a *k-mer* size of 93 despite a higher N50 and fewer contigs as when for example utilizing a *k-mer* size of 115 (Table 2.4.1, Figure 2.4.1 A, B). In contrast, SPAdes, which allows to combine a range of *k-mer* sizes in a multi-*k-mer* approach did not yield any mis-assemblies on this sequence for the N50 based best assemblies (Figure 2.4.1, C). The same was true for IDBA, which similarly utilizes an iterative process including multiple *k-mer* sizes, while removing assembled sequences in subsequent rounds of analysis.

The general applicability of these results was investigated using two additional, similarly obtained, *S. aureus* sequences [UA-S391(accession#CP007690) and Mu50-CC5-SCC*mec*II (ATCC700699; previously sequenced and available under accession # NC_002758], again revealing mis-assemblies for Velvet at the highest N50 values, while error free assemblies could be obtained for lower *k-mer* sizes (data not shown). In addition, the sequence of *Klebsiella pneumoniae* ST258 was similarly generated using Illumina HiSeq-2000 via 2X150b paired end sequencing and was assembled using all three assembly tools. In this case, apart from mis-assemblies seen for Velvet, also SPAdes and IDBA were shown to produce mis-assemblies for certain *k-mer* sizes (Figure 2.4.1D), further demonstrating the potential of WGM to identify mis-assemblies, even for assemblers utilizing multi-*k-mer* approaches.

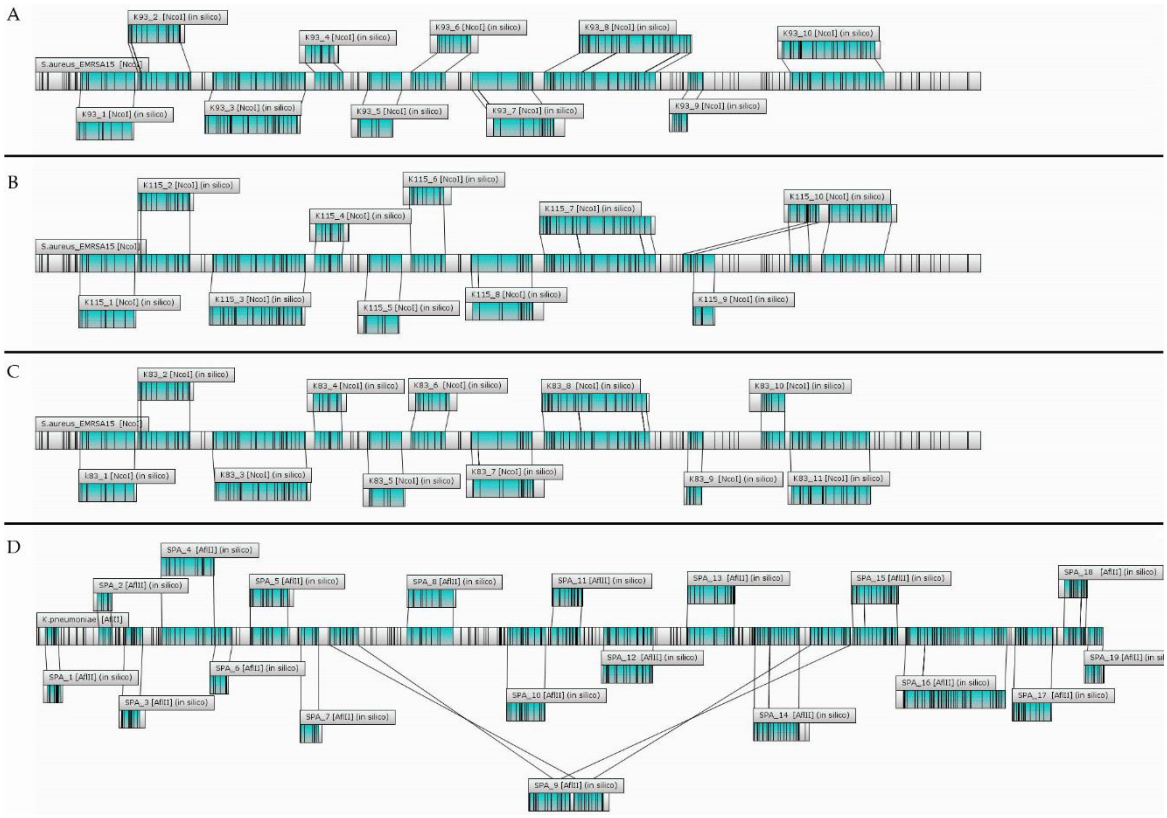


Figure 2.4.1 Alignment of contigs to the corresponding whole genome map: A) Velvet derived assembly using *k-mer* size 93, revealing no mis-assemblies; B) Velvet derived assembly using *k-mer* size 115, corresponding to the highest N50, but revealing mis-assemblies; C: SPAdes derived assembly using a multi-*k-mer* approach up to k-mer size 83, yielding the optimal N50 for this sequence and showing no mis-assemblies; D: SPAdes derived assembly using a multi-*k-mer* approach up to k-mer size 77, yielding the optimal N50 for this sequence, but showing mis-assemblies.

Table 2.4.1 Assembly statistics of Velvet applied on *Staphylococcus aureus* (MRSA) strain E-MRSA15-CC22-SCC*mec*IV showing an increase in contig size and N50 when using higher *k-mer* sizes but revealing mis-assemblies starting from *k-mer* size 97 using whole genome mapping.

K-mer Size	N50	Total number of contigs	Longest contig size	Num. of Mis-assemblies on mapped contigs (approx. nt involved in mis-assemblies
Velvet				
81	162295	40	340060	0
83	170447	38	351373	0
85	170449	37	351321	0
87	173763	33	351326	0
89	173765	33	351394	0
91	173767	33	351330	0
93	173769	35	340092	0
97	175770	33	365247	1 (~118928 nt)
99	175776	33	365260	1 (~118928 nt)
101	187438	32	365623	1 (~118928 nt)
103	187448	32	365625	1 (~118928 nt)
105	187458	32	365638	1 (~118928 nt)
107	187465	33	365647	1 (~118928 nt)
109	212189	32	365656	1 (~118928 nt)
111	212287	33	349286	1 (~118928 nt)
113	212292	34	349288	1 (~118928 nt)
115	212294	34	349290	1 (~118928 nt)
117	174074	35	349419	1 (~118928 nt)
119	174076	35	349423	1 (~118928 nt)
121	170642	37	349435	1 (~118928 nt)
123	170654	38	340456	1 (~118928 nt)

2.4.2 Conclusion

Genome assembly based on de Bruijn graphs potentially yields mis-assemblies when only considering standard parameters such as total number and length of the contigs and N50. However, Whole Genome Mapping provides a powerful tool to identify such mis-assemblies and to select the optimal *k-mer* sizes to produce optimally assembled genomes. Despite its additional cost, the biological need for error-free and complete genomes makes WGM an indispensable technique during the process of genome assembly and its validation.

Author contribution: BBX was involved in experimental work, data analysis, interpretation and drafting manuscript.

2.4.3 References

1. Li Y, Zheng H, Luo R, Wu H, Zhu H, Li R, Cao H, Wu B, Huang S, Shao H, et al: Structural variation in two human genomes mapped at single-nucleotide resolution by whole genome de novo assembly. *Nat Biotech* 2011, 29:723-730.
2. Salzberg SL, Yorke JA: Beware of mis-assembled genomes. *Bioinformatics* 2005, 21:4320-4321.
3. Phillippy AM, Schatz MC, Pop M: Genome assembly forensics: finding the elusive mis-assembly. *Genome Biol* 2008, 9: R55.
4. Dark M: Whole-genome sequencing in bacteriology: state of the art. *Infection and Drug Resistance* 2013:115.
5. Zerbino DR, Birney E: Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008, 18:821-829.
6. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al: SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012, 19:455-477.
7. Hunt M, Kikuchi T, Sanders M, Newbold C, Berriman M, Otto T: REAPR: a universal tool for genome assembly evaluation. *Genome Biology* 2013, 14:R47.
8. Peng Y, Leung HM, Yiu SM, Chin FL: IDBA – A Practical Iterative de Bruijn Graph De Novo Assembler. In *Research in Computational Molecular Biology. Volume* 6044. Edited by Berger B: Springer Berlin Heidelberg; 2010: 426-440: *Lecture Notes in Computer Science*].
9. Ananiev GE, Goldstein S, Runnheim R, Forrest DK, Zhou S, Potamousis K, Churas CP, Bergendahl V, Thomson JA, Schwartz DC: Optical mapping discerns genome wide DNA methylation profiles. *BMC Mol Biol* 2008, 9:68.
10. Onmus-Leone F, Hang J, Clifford RJ, Yang Y, Riley MC, Kuschner RA, Waterman PE, Lesho EP: Enhanced De Novo Assembly of High Throughput Pyrosequencing Data Using Whole Genome Mapping. *PLoS ONE* 2013, 8:e61762.
11. Sabirova JS, Xavier BB, Hernalsteens J-P, De Greve H, Ieven M, Goossens H, Malhotra-Kumar S: Complete Genome Sequences of Two Prolific Biofilm-Forming Staphylococcus aureus Isolates Belonging to USA300 and EMRSA-15 Clonal Lineages. *Genome Announcements* 2014, 2.

2.5 WHOLE-GENOME TYPING AND CHARACTERIZATION OF BLAVIM19-HARBOURING ST383 *KLEBSIELLA PNEUMONIAE* BY PFGE, WHOLE-GENOME MAPPING AND WGS

Sabirova JS*[1], **Xavier BB***[1], Coppens J[1], Zarkotou O[2], Lammens C[1], Janssens L[1], Burggrave R[3], Wagner T[3], Goossens H[1], Malhotra-Kumar S[1]
[1]Laboratory of Medical Microbiology, Vaccine & Infectious Disease Institute, Universiteit Antwerpen, Antwerp, Belgium. [2]Department of Microbiology, Tzaneio General Hospital, Piraeus, Greece. [3]OpGen, Inc., Gaithersburg, MD, USA.

Abstract

We utilized single-molecule, fluorescence-based restriction-mapping (whole-genome mapping-WGM), and whole-genome sequencing (WGS), to characterize 12 carbapenem-resistant *Klebsiella pneumoniae* isolated from patients at a Greek hospital during 2010-2013 (TGH1-TGH12). All strains were screened for carbapenemase genes by PCR, and typed by MLST, PFGE (*XbaI*), and WGM (*AflII)* according to manufacturer's protocols (Opgen, USA). WGS (Illumina) was performed on TGH8 and TGH10 and reads were *de novo* assembled and annotated (SPAdes, RAST). Contigs were aligned directly, and after *in silico AflII* restriction, with corresponding whole-genome maps (MapSolver, OpGen; BioNumerics v7.5, Applied Maths). All 12 strains were ST383. Eleven of the 12 strains were carbapenem-resistant of which 7 harboured $bla_{\text{KPC-2}}$ and 11, $bla_{\text{VIM-19}}$. Varying the parameters used to assign WGM clusters showed that these were comparable to the ST type, and to the 8 PFGE (sub)types (≥ 3-band difference). A similarity coefficient of 95% assigned all 12 WGMs to a single cluster while a more stringent analysis (similarity coefficient 99% or a ≥ 10 unmatched fragment difference) assigned the 12 WGMs to 8 (sub)clusters. Based on a ≥3-band difference between PFGE profiles, the Simpson's diversity index (SDI) of WGM (0.94, Jackknife pseudo-values CI: 0.883-0.996) and PFGE (0.93, Jackknife pseudo-values CI: 0.828-1.000) were similar (p = 0.649). However, discriminatory power of WGM was significantly higher (SDI: 0.94, Jackknife pseudo-values CI: 0.883-0.996) than PFGE profiles typed on a ≥7-band difference (SDI: 0.53, Jacknife pseudo-values CI: 0.212-0.849) (p = 0.007). Comparison of experimental and *in-silico* maps of two strains showed 11% and 1.4-2.1% losses of <5kb and >5kb fragments, respectively. This study demonstrates the application of whole genome mapping to understanding the epidemiology of hospital-associated *K. pneumoniae*. Utilizing a combination of WGM and WGS, we also present here the first longitudinal genomic characterization of the highly dynamic carbapenem-resistant ST383 *K. pneumoniae* clone that is rapidly gaining importance in Europe.

2.5.1 Introduction

The spread of multidrug-resistant *K. pneumoniae* strains in hospitals constitutes a pressing global health problem. The increasing prevalence of plasmid-encoded carbapenem-hydrolysing enzymes in *K. pneumoniae* is of particular concern due to their ability to hydrolyse almost all β-lactam antibiotics, as well as their genetic association with transferable multidrug resistance.(Goren, et al. 2010; Mathers, et al. 2011; Stapleton, et al. 1999) Infections due to carbapenem-resistant *K. pneumoniae* are not only difficult to treat due to limited therapeutic options, but such clones could potentially cause hospital epidemics if not promptly detected and contained.(Saidel-Odes and Borer 2013)

Molecular typing techniques are effective surveillance tools to monitor the dynamics of multidrug-resistant clones circulating in hospitals during non-outbreak situations and to detect early signs of an outbreak. Currently used techniques are based on amplification of marker genes followed by sequencing (multilocus sequence typing, MLST), or targeting entire genomes by pulsed-field gel electrophoresis (PFGE) or by whole genome mapping (WGM), with the restriction fragments analysed on a gel or in a microfluidic device, respectively. Of the three techniques, MLST is most commonly utilized for bacteria strain typing. While it is highly reproducible and relatively inexpensive, the resolution achieved by clustering strains based on sequenced segments of seven or more housekeeping genes is not high enough to study inter- and intra-clonal genetic diversity. Compared to PFGE, WGM has the advantage of being less labour intensive and allowing same day results and, importantly, inter-lab reproducibility. WGM is also technologically less challenging than whole genome sequencing (WGS) and does not require bioinformatics expertise. Furthermore, the genetic content of variable genome regions identified by WGM can be extracted by utilizing the vast number of publicly available whole genome sequences. These can be utilized to develop *in silico* restriction maps using the same enzymes as used for experimentally-generated maps, allowing comparison of restriction patterns and eventual identification of the genetic content of the variable regions in the strains of interest.(Miller 2013a)

In this study, we demonstrate the utility of WGM in conjunction with WGS for typing, characterizing, and dissecting the genomic features of carbapenem-resistant *K. pneumoniae* isolated at the Tzaneio General (TG) hospital, Piraeus, Greece. Carbapenem-resistant *K. pneumoniae* producing VIM-type metallo-β-lactamases have been endemic in Greek hospitals since the early 2000s.(Miyakis, et al. 2011) Many strains with VIM-1–producing *K. pneumoniae* have also been described.(Poulou, et al. 2010) From 2007, KPC-type carbapenemases became prevalent and even

caused outbreaks,(Pournaras, et al. 2009), (Maltezou, et al. 2009) followed by emergence of strains coproducing KPC and VIM.(Giakkoupi, et al. 2009) We studied VIM- and KPC-+VIM-producing *K. pneumoniae* isolated from colonized or infected patients during 2010-2013 at the TG hospital.

2.5.2 Materials and Methods

2.5.2.1 Strain collection

Twelve *K. pneumoniae* (TGH1-TGH12), harbouring bla_{VIM} and/or bla_{KPC}, and isolated from patients admitted to the intensive care unit (n=8) or surgical ward (n=4) at the TG hospital during 2010-2013 were studied. Clinical data and strain characteristics are outlined in Table 1.

2.5.2.2 Antimicrobial resistance profiling

All 12 strains were screened for resistance to 17 antibiotics, including β-lactams with and without β-lactamase inhibitors, by disk diffusion (Table 1). MICs of carbapenems (ertapenem, imipenem and meropenem) were determined by Etest (bioMérieux Inc., Durham, NC), and results interpreted according to Clinical and Laboratory Standards Institute cut-offs. (Clinical and Laboratory Standards Institute 2014) Strains were screened for presence of extended-spectrum β-lactamase (ESBL) and carbapenemase genes by PCR and Sanger sequencing as described previously.(Mabilat, et al. 1990; Monstein, et al. 2007; Poirel, et al. 2011a).

2.5.2.3 MLST and PFGE

MLST was performed as described previously for seven marker genes (*gapA, infB, mdh, pgi, phoE, rpoB, tonB*),(Diancourt, et al. 2005) and sequence types were assigned using the Institute Pasteur database (www.pasteur.fr/mlst). PFGE was performed as follows. Briefly, cells from an overnight blood agar culture were washed, adjusted to a density of 1.0 OD at 600nm in EC lysis buffer (100 mM EDTA, 5M NaCl, 0.5% Brij-58, 0.2% deoxycholate, 10% N-laurylsarcosine, 6 mM Tris–HCl pH 7.6), and after centrifugation, resuspended into 200µL EC lysis buffer with 10 µL proteinase K (20mg/mL). Cell suspension was mixed with equal volume of 1.0% (w/v) SeaKem Gold Agarose (Westburg) to form plugs. These were incubated in 2 mL of EC lysis buffer and 10 µL of proteinase K (20 mg/mL) for 2h at 55°C. The plugs were washed five times at 55°C for 15 min with sterile water and digested overnight at 37°C with 50 U of *XbaI* (Life Technologies). Plug slices were placed on the well comb, and tempered agarose was poured in the gel mould. The gel was run at 6.0 V/cm with an initial switch time of 5 s to a final switch time of 35 s at 14°C in 0.5× TBE (Tris-borate-

EDTA) running buffer for 24 h. DNA band profiles were stained with ethidium bromide and visualized and digitized by the Quantity One documentation system (Bio-Rad). Conversion, normalization and analysis of patterns was carried out using GelCompar software version 4.0 (Applied Maths, Kortrijk, Belgium) and pattern analyses were performed as described. (Tenover, et al. 1995)

2.5.2.4 WGM

The complete genomes of all 12 strains were mapped employing the Argus® Whole Genome Mapping System (Opgen Inc, Gaithersburg, USA). DNA extraction, DNA quality control, DNA restriction using *AflII*, and loading on a MapCard were done according to manufacturer's protocols. Briefly, *K. pneumoniae* colonies grown overnight at 37°C on Mueller-Hinton agar plates were immobilized in agarose plugs (as described for the PFGE protocol) and subjected to in-plug gentle lysis, followed by thoroughly washing plugs in TE buffer at 42ºC and enzymatic treatment with β-agarose (New England Biolabs Inc, Ipswich, USA). For WGM, dilutions were prepared with dilution buffer, and DNA was checked for quality and presence of high molecular weight DNA molecules (Argus® QCard kit, Opgen) and subsequently loaded on the MapCard (Argus® MapCard II kit, Opgen, Inc). *De novo* assembly of restricted DNA fragments was performed using MapManager software (Opgen). For editing, maps were adjusted in orientation and in their replication point employing an *in silico* map generated from *K. pneumoniae* KPNIH31 (accession number CP009876.1) using built-in function with default parameters in MapSolver (Opgen Inc.). All WGMs were analysed by filtering out fragment sizes smaller than 5 kb from the analysis and using three different set of parameters that allowed clustering, sub-clustering, and discrepant analysis. Firstly, pattern search was performed with a relative tolerance of 5%, an absolute tolerance of 2000 (bp) with 1 mismatch and secondary criteria with most identical matches, and a similarity co-efficient set at 95% was utilised. Secondly, relative tolerance of 1%, absolute tolerance of 1000 bp, pattern length search with 8 fragments to generate a dendrogram with most identical matches, and a similarity co-efficient of 99% was utilised. Thirdly, the same parameters as the second analysis were utilised, but instead of similarity, we studied the 'absolute number of unmatched fragments' between two WGMs assigning ≥ 10 unmatched fragments as a cut-off to assign a new cluster. All pattern-search cluster-analysis was performed using Bionumerics v7.5 (Applied Maths, Kortrijk, Belgium) employing UPGMA, and similarity matrix of clusters was defined by Cophenetic Correlation Coefficient (CCC).(Boopathi 2013)

2.5.2.5 *In silico* restriction mapping using *AflII*

For two strains, TGH8 and TGH10, which were whole genome sequenced (see below), we performed *in silico* restriction mapping in order to quantify the fragment losses and other differences observed with experimentally-generated mapping using WGM. Comparison of *in silico* and experimentally-generated maps of the clinical strains was performed using MapSolver and BioNumerics v7.5.

2.5.2.6 WGS and comparative genome analysis

Two *K. pneumoniae* isolates (TGH8 and TGH10) were whole genome sequenced. Briefly, genomic DNA was extracted using MasterPure^{TM} DNA Purification Kit (Epicentre Technologies Corp). WGS was performed using Nextera XT DNA Library Preparation Kit followed by sequencing via 2 X 150 bp paired end sequencing (Illumina Inc.). The sequence reads of strains were *de novo* assembled using SPAdes, (Bankevich, et al. 2012) and annotated using Rapid Annotation Subsystem technology (RAST) online server. (Aziz, et al. 2008b; Overbeek, et al. 2014) *De novo* assembled contigs were aligned against corresponding whole genome maps in MapSolver in order to generate pseudo chromosomes and also to identify variable regions and genomic content. To precisely localize the site and characteristics of the genomic change, sequence contigs corresponding to the regions of genome divergence were further analysed in CLC Genomics Workbench v7.5.1 (CLCbio, Denmark). Similarity search of variable regions was performed using NCBI BLAST at nucleotide and protein level. (Altschul, et al. 1990; Cosentino, et al. 2013; States and Gish 1994) Integron (In) sorting and analysis was done as follows: *de novo* assembled contigs were sorted by length and coverage i.e., >1 kb and > 250-fold coverage, the raw reads were extracted separately, and *de novo* assembled using SPAdes. The reads were mapped again in order to validate mis-assemblies using CLC Genomics Workbench (CLCbio), and the contigs annotated and validated using online databases RAC: (Repository of Antibiotic Resistance Cassettes http://rac.aihi.mq.edu.au/rac/) and INTEGRALL.(Tsafnat, et al. 2011b)·(Moura, et al. 2009) Identification and typing of integrative and conjugative elements (ICEs) was performed using web-based resource ICEberg. (Bi, et al. 2012) Plasmid sequence analysis was done as follows. Firstly, the contigs were screened for plasmid origin by using the online tool "Plasmid finder" (https://cge.cbs.dtu.dk/services/PlasmidFinder/).(Carattoli, et al. 2014) Next, the generated plasmid-specific contigs were used as a reference template and raw reads were mapped. Lastly, *de novo* assembly was performed on the mapped reads using CLC Genomics Workbench v7.5.1

(CLCbio, Denmark) with default parameters. The accession numbers for the integron (*In4863*-like) and whole genome sequences deposited in Genbank are as follows; *In4863*-like (KT820212), TGH8 (CP012743) and TGH10 (CP012744).

2.5.3 Results

2.5.3.1 Phenotypic and genotypic characterization of *K. pneumoniae*

Twelve of the 11 strains were carbapenem-resistant of which 7 harboured $bla_{\text{KPC-2}}$ and 11 $bla_{\text{VIM-19}}$ (Table 1). The ST383 strains exhibited five carbapenemase/ESBL combinations: 1) $bla_{\text{VIM-19}}$, $bla_{\text{KPC-2,}}$ and $bla_{\text{CTX-M-15}}$ (TGH1, TGH6, TGH8, TGH9), 2) $bla_{\text{VIM-19}}$, and $bla_{\text{KPC-2}}$ (TGH2, TGH4, andTGH5), 3) $bla_{\text{VIM-19}}$, and $bla_{\text{CTX-M-15}}$ (TGH3), 4) $bla_{\text{CTX-M-15}}$ (TGH7) and 5) $bla_{\text{VIM-19}}$ (TGH10, TGH11 and TGH12) (Table 2.5.1).

Table 2.5.1 Clinical data and characteristics of strains under study

Strain ID	Clinical information			Molecular typing								Resistance profile											
	Unit	Site of isolation	Isolation date	PFGE*	ST type	WGM Type	*bla*- genes					MIC by E test						Sensitivity by Disk Diffusion**					
							bla_{VIM}	bla_{KPC}	bla_{CTX-M}	bla_{SHV}	bla_{TEM}	Meropenem		Ertapenem		Imipenem		Fosfomycin		Cefoxitin		Gentamicin	
TGH1	ICU	Blood	Jan-10	3a	ST383	C1a	VIM-19	KPC-2	CTX-M-15	SHV-1	TEM-1	R	>32	R	>32	R	>32	S	19	R	0	R	0
TGH2	S	Wound Swab	Feb-10	1a	ST383	C2b	VIM-19	KPC-2	-	SHV-1	TEM-1	R	>32	R	>32	R	>32	S	19	R	0	S	21
TGH3	ICU	Urinary Catheter	Apr-10	1b	ST383	C1a	VIM-19	-	CTX-M-15	SHV-1	-	R	>32	R	>32	R	>32	S	19	R	0	R	0
TGH4	S	Urine	May-10	1a	ST383	C2b	VIM-19	KPC-2	-	SHV-1	TEM-1	R	>32	R	>32	R	>32	S	20	R	0	S	21
TGH5	ICU	Blood	May-10	3a	ST383	C1b	VIM-19	KPC-2	-	SHV-1	TEM-1	R	>32	R	>32	R	>32	S	19	R	0	S	22
TGH6	ICU	CVC	Aug-10	1d	ST383	C1b	VIM-19	KPC-2	CTX-M-15	SHV-1	-	R	>32	R	>32	R	>32	S	21	R	0	R	0
TGH7	S	Urine	Nov-10	1c	ST383	C2c	-	-	CTX-M-15	SHV-1	-	S	0.023	S	0.023	S	0.19	S	21	I	15	R	0
TGH8	ICU	CVC	Jun-11	2a	ST383	S3	VIM-19	KPC-2	CTX-M-15	SHV-1	TEM-1	R	>32	R	>32	R	>32	R	0	R	0	R	0
TGH9	ICU	Blood	Aug-11	3b	ST383	S4	VIM-19	KPC-2	CTX-M-15	SHV-1	TEM-1	R	>32	R	>32	R	>32	S	18	R	0	R	0
TGH10	ICU	Wound Swab	Mar-13	1b	ST383	C2a	VIM-19	-	-	SHV-1	-	R	>32	R	>32	R	>32	S	18	R	0	R	0
TGH11	ICU	Rectal	Apr-13	1e	ST383	C1c	VIM-19	-	-	SHV-1	-	R	>32	R	>32	R	>32	S	19	R	0	R	0
TGH12	S	Blood	Apr-13	1b	ST383	C2a	VIM-19	-	-	SHV-1	-	R	>32	R	>32	R	>32	S	19	R	0	R	0

*To define a PFGE type, the 7-band difference cut-off was utilized. The number of band differences between the subtypes is 1, which indicates that the organisms were genetically closely related

** All strains were resistant to the following antibiotics by disc diffusion: amoxicillin, piperacillin, amoxicillin-clavulanic acid, piperacillin-tazobactam, cefepime, cefotaxime, ceftazidime, aztreonam, ciprofloxacin, nitrofurantoin, trimethoprim- sulfamethoxazole. Abbreviations: SS - surgical site, ICU - intensive care unit, CVC - central venous catheter, R – resistant, S - susceptible.

2.5.3.2 Genetic diversification within *K. pneumoniae* ST383 identified by WGM

Strains TGH1-TGH12 belonged to ST383 and were divided into 3 PFGE types based on a ≥ 7-band difference. Subtypes were delineated based on a ≥ 3-band difference between profiles belonging to the same PFGE type (Figure 2.5.1). Utilizing a similarity co-efficient cut-off of 95% and TGH1, the oldest strain in our collection, as the reference map, WGMs of all 12 ST383 strains were found to form 1 cluster (Figure 2). TGH8 and TGH9 showed maximum dissimilarity (4.7%) compared to the other 10 strains (Figure 2.5.2). Differences between strains were mainly due to presence of mobile elements such as ICEs, prophages, and transposons. To add further granularity to our data and observe sub-clusters, we analysed the WGMs using more stringent parameters (Figure 3A and B). Figure 3A shows clustering based on a similarity co-efficient of 99% and Figure 3B using a diversity co-efficient (no. of unmatched fragments). Both parameters showed similar (sub) clustering of WGMs, identifying two clusters, C1 and C2, and two singletons, C3 and C4 (Figure 3B), with C1 and C2 each divided into two sub clusters and a singleton (Figure 3B). In this analysis, TGH9 was found to be the most dissimilar strain showing only 77.4% similarity or a 124-unmatched fragment difference with the rest of the WGMs (Figure 2.5.3A and B). Insertions in TGH8 and TGH11, identified by WGM analysis, accounted for ~1.8% and ~1.2 % genomic expansion (Figure 2.5.3A), respectively, compared to the other ST383 strains studied here.

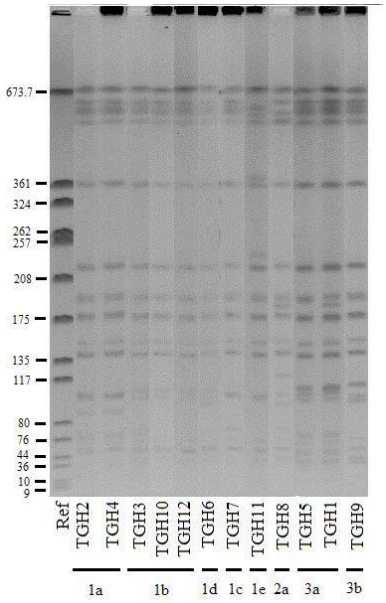


Figure 2.5.1 PFGE types (≥ 7-band difference) and subtypes (≥ 3-band difference) of *K. pneumoniae* ST383.

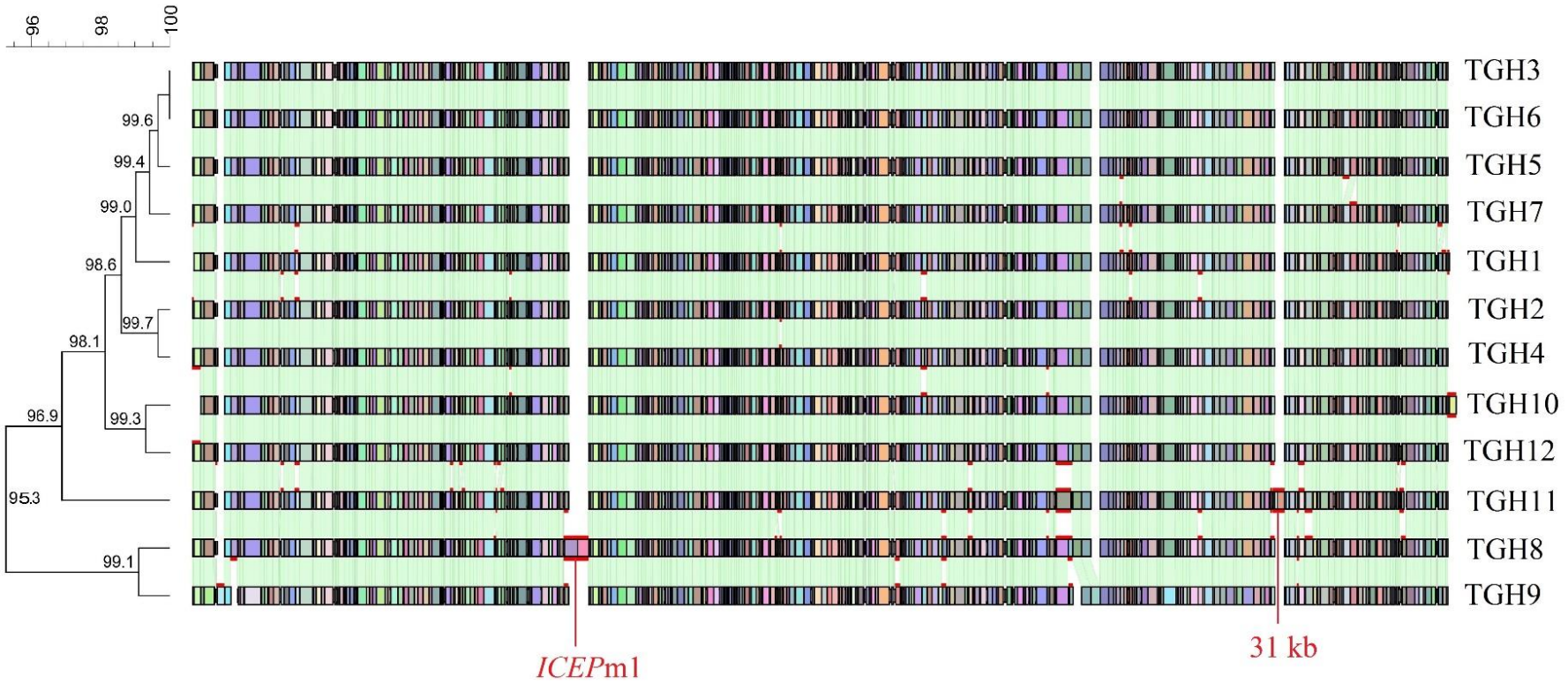


Figure 2.5.2 Comparison of whole genome maps of ST383 strains. Green shaded areas indicate identical restriction patterns among the maps and red horizontal marks represent variations. A similarity co-efficient of 95% was utilized which assigned the strain to a single cluster.

Figure 2.5.3 Comparison of whole genome maps of ST383 strains using a similarity coefficient of 99% (A), and a '≥ 10 unmatched fragments' criterion (B). Both parameters showed similar (sub) clustering of WGMs, identifying two clusters (C1 and C2) and two singletons (C3 and C4) with C1 and C2 each divided into two sub clusters and a singleton.

2.5.3.3 Comparison of discriminatory power and congruence between WGM and PFGE

We utilized the adjusted Wallace co-efficient to compare partitions, and the Simpson's diversity index (SDI) to compare the discriminatory power of WGM and PFGE.(Carriço, et al. 2006) As all strains studied here belonged to one ST type (ST383), no. of partitions achieved with MLST was 1 and hence was not included in this analyses. Based on the parameter 'no. of unmatched fragments' ≥10 (or distance co-efficient) (Figure 3B), we obtained 8 (sub) clusters with WGM that were compared to the corresponding 3 PFGE types (≥ 7-band difference) and 8 subtypes (≥ 3-band difference). Based on a ≥3-band difference between PFGE profiles, no significant difference was observed in the SDI of WGM (0.94, Jackknife pseudo-values CI: 0.883-0.996) and PFGE (0.93, Jackknife pseudo-values CI: 0.828-1.000) (p = 0.649). The adjusted Wallace co-efficient was also similar for PFGE (0.36, 95% CI: 0.000-0.816, ≥ 3-band difference) and WGM (0.46, 95% CI: 0.076-0.842). However, discriminatory power of WGM was significantly higher (SDI: 0.94, Jackknife pseudo-values CI: 0.883-0.996) than PFGE based on a ≥7-band difference (SDI: 0.53, Jacknife pseudo-values CI: 0.212-0.849) (p = 0.007).

2.5.3.4 Comparing experimentally derived WGM with in-silico restriction mapping using *AflII*

In order to assess the reliability and accuracy of the experimentally derived WGMs, we compared these with *in silico AflII* restriction maps generated from the whole genome sequencing data of TGH8 and TGH10 (Figure 4A). *AflII*, the manufacturer recommended and optimized enzyme for *K. pneumoniae* produces 442 (TGH8) and 437 (TGH10) fragments ranging from 78 kb-14 nt. After removal of the <5 kb fragment differences, we observed a similarity co-efficient of 80-84% between the experimental and *in silico* maps (Figure 2.5.4A). A similar comparison between PFGE and *XbaI in silico* restriction mapping was not possible because of the basic difference in map generation; the former being molecular size-based fragmentation and the latter an ordered genome map. Nonetheless, *XbaI in silico* restriction mapping of TGH8 and TGH10 produced approximately 45 and 37 fragments respectively (data not shown), while their respective PFGE profiles showed 17 and 15 fragments, respectively, with the < 36 kb fragments lost to follow-up on PFGE gels (Figure 2.5.1). Comparison of TGH8 experimental and *in silico* maps showed that there were missing restriction cuts in the former (Figure 2.5.4B). The TGH8 *in silico* map generated a total of 442 *AflII*-restricted fragments of which 284 were >5 kb, while the experimental map showed a total of 386

fragments of which 278 were > 5 kb. Similarly, the TGH10 *in silico* map generated a total of 437 *AflII*-restricted fragments of which 282 were >5 kb, while the experimental map showed a total of 385 fragments of which 278 were > 5 kb.

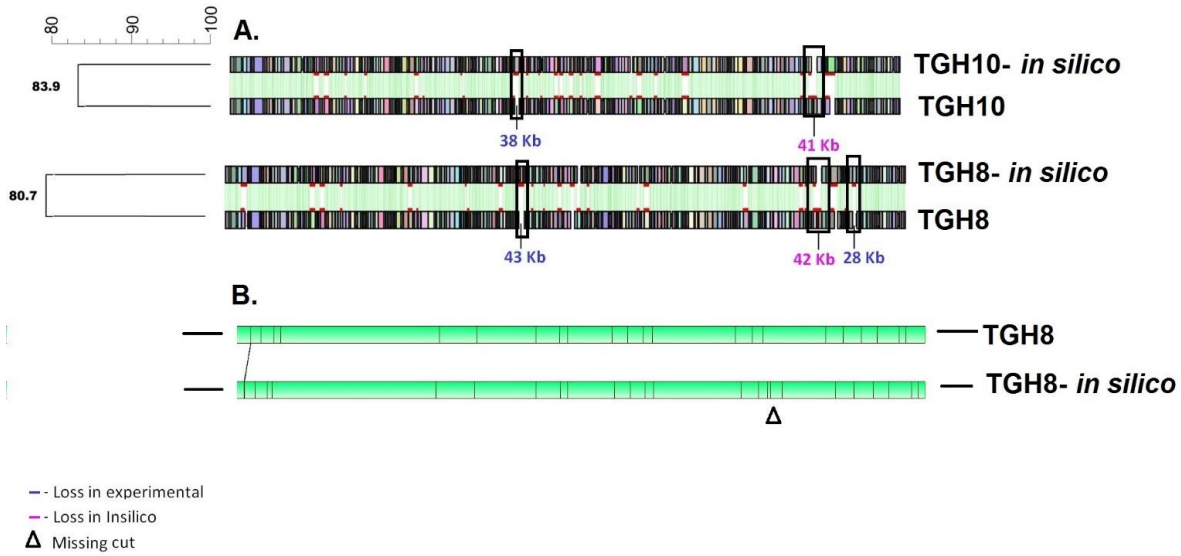


Figure 2.5.4 Comparison of experimental and *in silico* restriction maps using *AflII* that show fragment losses ranging from 28 kb to 43 kb in both experimental and *in silico* maps (A). Comparison of zoomed-in experimental and *in silico* restriction map of TGH8 showing missing restriction enzyme cuts in the former (B).

2.5.3.5 Intra-cluster comparison of ST383 whole genome maps

Compared to the other ST383 maps, TGH11 and TGH8 showed two unique insertions of ~31 kb and ~110 kb, respectively (Figure 2). These genomic insertions were unique in our mapped strains as deduced from comparative analysis of TGH8 and TGH11 maps with *in silico* maps generated from six previously sequenced *K. pneumoniae* available on NCBI (data not shown). By aligning the pseudo chromosomes of TGH8 and TGH10 to their corresponding WGMs, it was possible to pinpoint the sequence region harbouring the insertion in TGH8. The TGH8 insertion (from 773,512, to 884,869 kb) was identified as 99% homologous to a region present in *Proteus mirabilis* strain HI4320 (2,793,662 to 2,886,224 kb, accession number: AM942759). In TGH8, the ~110 kb genomic insertion lies in a region of high plasticity as evidenced by presence of flanking phage and mobile element remnants such as genes encoding phage capsids, phage-associated hypothetical proteins as well as transposases. Our search of the ~110 kb insertion against the ICEberg database identified an unclassified ~94 kb *ICEPm1* element, which harbours genes encoding a putative signal transducer/ampG/MFS, and virulence-associated genes such as F17 fimbrial protein precursor, iron acquisition yersiniabactin synthesis enzyme, soluble lytic murine transglycosylase, type IV

secretory pathway VirD4/B4, and interestingly, a lipid A export ATP-binding protein/MsbA. Similar searches for the ~ 31 kb insertion in TGH11 did not identify any pathogenic or virulence-related determinants in this region but rather genes encoding for glucan biosynthesis protein D precursor, permeases of drug/metabolite transporter (DMT), tellurite resistance proteins (TehA/TehB), benzyl alcohol dehydrogenase, and an uncharacterized membrane lipoprotein (data not shown).

2.5.3.6 Analysis of integrons harbouring blaVIM variants

Integron analysis of our sequenced strains (TGH8 and TGH10) showed that *bla*$_{\text{VIM-19}}$ in both TGH8 and TGH10 was carried on the class 1 integron, *In4863*, which showed 99% nucleotide level similarity between TGH8 and TGH10 (Figure 5). However, in contrast to the previously sequenced *In4863* (accession number KF894700), the element in TGH8 and TGH10 harboured a variant promoter (PcW-TGN-10),(Jové, et al. 2010) showing polymorphism at the 2nd base of the -35 sequence (T**T**/**G**ACA). In addition, Int1 also showed two predicted amino acid changes (Pro32Arg and Asn39His) compared to the published *In4863* sequence (Figure 5). *In silico* analysis of plasmid specific contigs using Plasmid finder and comparative sequence analysis showed that *bla*$_{\text{KPC-2}}$, *bla*$_{\text{CTX-M}}$ and *bla*$_{\text{VIM-19}}$ were might be carried on three different plasmids (IncFII(K), IncFIB and IncA/C2) in TGH8.

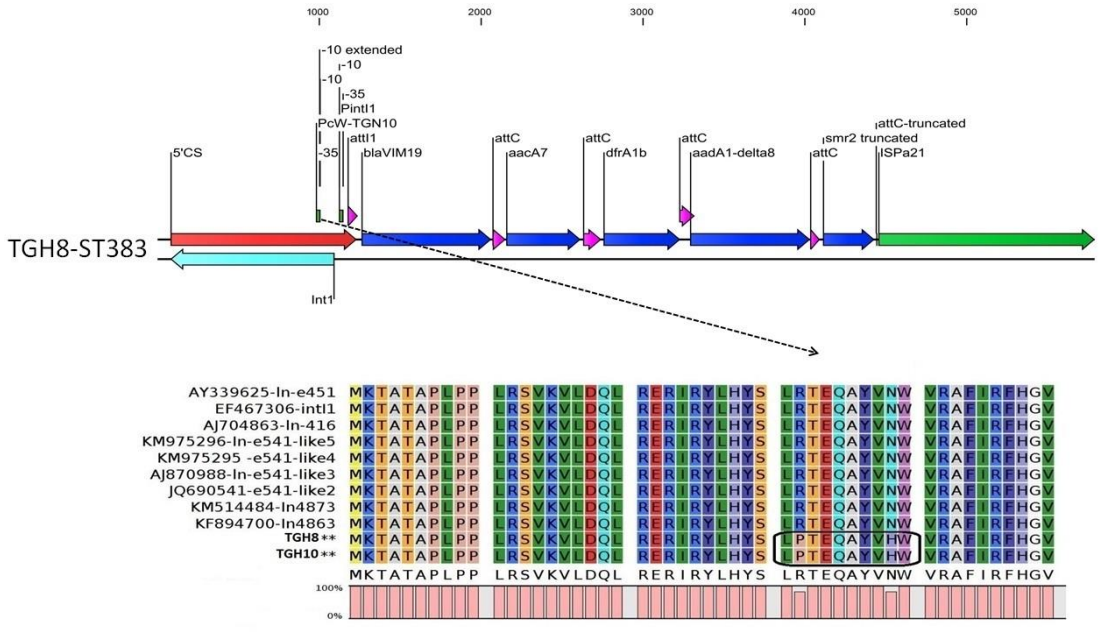


Figure 2.5.5 Gene cassette arrangement in Class 1 integrons harboured by ST383 (*In4863*-like) and Int1 protein alignment. Amino acids boxed in black represent variation in the Int1 protein between TGH8 and TGH10 ST383 compared to published Class 1 integrases.

2.5.5 Discussion

In this study, we utilized a collection of carbapenem-resistant *K. pneumoniae* isolated during 2010-2013 at the TG hospital, Greece in order to compare currently utilized gene- and genome-based typing methods as well as to better understand the molecular epidemiology of carbapenem-resistant *K. pneumoniae* at TGH.

MLST assigned all strains studied here to ST383. In concordance, a cut-off of 95% similarity co-efficient also assigned all WGMs to a single cluster. To allow comparisons with PFGE, which assigned the ST383 to 3 types based on a ≥ 7-band difference and to 8 subtypes based on a ≥ 3 band difference, we utilized a '≥ 10 unmatched fragments' criteria to assign WGM clusters. Compared to a ≥ 7-band difference in PFGE profiles, WGM showed a significantly higher discriminatory power while a 3- band difference criteria showed a similar SDI for PFGE and WGM. Different restriction enzymes had to be employed for these methods; *AflII* is the manufacturer recommended enzyme for whole genome mapping of *K. pneumoniae*, however fragment sizes with this enzyme range from 78 kb to14 nt. As a large number of *AflII*-generated fragments fall below the resolution of PFGE gels (<36 kb fragments in our hands, Figure 1), we utilized *XbaI* for the latter method. *In silico* mapping of TGH8 with *AflII* and *XbaI* generated 442 and 45 (556 kb to 197 nt) fragments, respectively. However, on-gel *XbaI* PFGE profiles consisted of, on average, 14 bands. Comparison of experimental and *in silico* maps also highlighted the challenges WGM faces for scoring of small fragments. Of the 56 and 52 missing fragments in the TGH8 and TGH10 experimental WGMs compared to their *in silico* maps, 50 and 48 (11% for both) were < 5 kb, respectively. This corresponds to a small fragment loss of 32%. In contrast, fragment loss rate of > 5 kb fragments were only 1.4-2.1% for both TGH8 and TGH10 WGMs.

Despite lack of evident differences in discriminatory power between PFGE and WGM, which might be due to the fact that the number of strains analysed here were limited and were closely related, a major advantage of WGM is that the technique produces an ordered genome map that allows comparison to previously sequenced genomes for identifying larger (> 5 kb) insertions/deletions. For instance, prior to sequencing, we had already identified the insertion observed in the WGM of TGH8 as an *ICEPm1* element by comparing with *in silico* maps generated from six previously sequenced *K. pneumoniae* available on NCBI. This element has been shown to originate from *P. mirabilis*, and is also highly conserved in other uropathogens such as *Providencia stuartii* and *Morganella morganii*.(Flannery, et al. 2009) Interestingly, the *ICEPm1* element is known to transfer in a site-specific manner, using phenylalanine tRNA genes as an integration site,

(Flannery, et al. 2011) and may contribute to fine tuning and adaptation of *K. pneumoniae* towards preferred infection or colonization pathways.(Chen, et al. 2010)

ST383 is a recently described clone that was first detected in Greek hospitals during 2009-2010.(Papagiannitsis, et al. 2010) Majority of the ST383 strains circulating in various Greek hospitals during 2009-2010 reportedly co-harboured *bla*$_{\text{VIM-4}}$, *bla*$_{\text{KPC-2}}$ and *bla*$_{\text{CMY-4}}$ β-lactamases.(Papagiannitsis, et al. 2010) Recent studies of 1-3 isolates of ST383 *K. pneumoniae* recovered during 2008-2010 have also reported presence of *bla*$_{\text{VIM-19}}$ in this ST type.(Papagiannitsis, et al. 2015; Samuelsen, et al. 2011) ST383 strains isolated at TGH in 2010-11 harboured the *bla*$_{\text{VIM-19}}$, *bla*$_{\textit{KPC-2}}$ and *bla*$_{\textit{CTX-M-15}}$ plasmids in various combinations, while one isolate that was carbepenem-susceptible did not harbour a carbapenemase. On the other hand, strains isolated in 2013 (TGH10, TGH11 and TGH12) harboured only *bla*$_{\text{VIM-19.}}$ These data underscore the remarkable plasticity of ST383 in terms of the accessory genome. Subject to the limited high-level MIC resolution allowed by Etest, carbapenem resistance remained high in the TGH10, TGH11 and TGH12 isolates despite loss of *bla*$_{\textit{KPC-2}}$ and potentiates the possibility that plasmid loss might have benefitted the ST383 strains in terms of fitness and transmissibility.

The *bla*$_{\text{VIM-19}}$ carbapenemase is also a new metallo-β-lactamase (MBL) gene variant isolated in Algiers in 2009.(Robin, et al. 2010) It is known to be derived from *bla*$_{\text{VIM-1}}$, differing from the former by two substitutions: Ser228Arg and Asn215Lys,(Robin, et al. 2010) which confer higher resistance to the carbapenems in comparison to *bla*$_{\text{VIM-1}}$.(Rodriguez-Martinez, et al. 2010) Accordingly, all *bla*$_{\text{VIM-19}}$ harbouring ST383 in this study, irrespective of the presence of *bla*$_{\text{KPC-2}}$, showed high-level resistance to meropenem, imipenem and ertapenem. Pournaras *et al.* have reported presence of *bla*$_{\text{VIM-19}}$ in a *K. pneumoniae* clinical strain co-producing *bla*$_{\text{KPC-2}}$ carbapenemase, *bla*$_{\text{CMY-2}}$ cephalosporinase and *bla*$_{\text{CTX-M-15}}$ extended-spectrum β-lactamase.(Pournaras, et al. 2010b) They found the *bla*$_{\text{VIM-19}}$ gene to be associated with a new class 1 integron with a structure similar to that carrying the close variant gene *bla*$_{\text{VIM-4}}$ in an *Enterobacter cloacae* isolate from Greece.(Pournaras, et al. 2010b) The *bla*$_{\text{VIM-19}}$ gene cassette was located downstream of the *attI1* recombination site, followed by an *aacA6* cassette, a *dfrA1* cassette, an *aadA1* cassette and the 3′-CS, containing *qacEΔ1* and *sulI*.(Pournaras, et al. 2010b) Another study has shown *bla*$_{\text{VIM-19}}$ to be harboured on *In4863*.(Papagiannitsis, et al. 2015) In our TGH8 and TGH10 strains, the *bla*$_{\text{VIM-19}}$ gene cassette was carried on an *In4863*-like element that differed from the previously described *In4863* by two (predicted) amino acid substitutions in the integrase and one nucleotide change at the 2$^{\text{nd}}$ position (T>G) in the integron promoter sequence. Interestingly, a recent study showed that these changes result in increased integrase activity and a concomitant

decrease in promoter strength,(Jové, et al. 2010) which hypothetically would increase the frequency of recombination events and horizontal transfer eventually causing a more rapid dissemination of the *In4863*-like element under antibiotic pressure. Also, the proximity of the gene cassette to the integron promoter influences its expression; closer the gene cassette to the promoter higher the expression. An *in vitro* study analyzed the impact of chloramphenicol pressure on gene cassette rearrangements in a class 1 integron harbouring the chloramphenicol-resistance encoding *catB9* gene, and found a wide variety of rearrangements under chloramphenicol pressure all leading to increased proximity of *catB9* to the integron promoter.(Barraud and Ploy 2015) Remarkably, all known MBL gene cassettes, including those harbouring $bla_{\text{VIM-19}}$, have been found to be consistently placed next to the integron's promoter.(Papagiannitsis, et al. 2015; Zhao and Hu 2015) Sustained carbapenem use and selection pressure in hospital environments are likely responsible for maintenance of MBL gene cassettes in this priority position.

To conclude, this study demonstrated the application of whole genome mapping to understanding the epidemiology of hospital-associated *K. pneumoniae*. Additionally, a combination of whole genome mapping and sequencing provided novel insights on the genomic features of the multi-drug resistant ST383 *K. pneumoniae* clone that is rapidly gaining importance in terms of prevalence and clinical significance in Europe.

Author contribution: BBX was involved in experimental work, data analysis, interpretation and drafting manuscript.

2.5.6 References

1 Mathers AJ, Cox HL, Kitchel B et al. Molecular Dissection of an Outbreak of Carbapenem-Resistant Enterobacteriaceae Reveals Intergenus KPC Carbapenemase Transmission through a Promiscuous Plasmid. *mBio* 2011; **2**: e00204-11.

2 Stapleton PD, Shannon KP, French GL. Carbapenem resistance in Escherichia coli associated with plasmid-determined CMY-4 beta-lactamase production and loss of an outer membrane protein. *Antimicrobial agents and chemotherapy* 1999; **43**: 1206-10.

3 Goren MG, Carmeli Y, Schwaber MJ et al. Transfer of carbapenem-resistant plasmid from Klebsiella pneumoniae ST258 to Escherichia coli in patient. *Emerging infectious diseases* 2010; **16**: 1014-7.

4 Saidel-Odes L, Borer A. Limiting and controlling carbapenem-resistant Klebsiella pneumoniae. *Infection and drug resistance* 2013; **7**: 9-14.

5 Miller JM. Whole-genome mapping: a new paradigm in strain-typing technology. *J Clin Microbiol* 2013; **51**: 1066-70.

6 Miyakis S, Pefanis A, Tsakris A. The Challenges of Antimicrobial Drug Resistance in Greece. *Clinical Infectious Diseases* 2011; **53**: 177-84.

7 Poulou A, Spanakis N, Pournaras S et al. Recurrent healthcare-associated community-onset infections due to Klebsiella pneumoniae producing VIM-1 metallo-β-lactamase. *Journal of Antimicrobial Chemotherapy* 2010; **65**: 2538-42.

8 Pournaras S, Protonotariou E, Voulgari E et al. Clonal spread of KPC-2 carbapenemase-producing Klebsiella pneumoniae strains in Greece. *Journal of Antimicrobial Chemotherapy* 2009; **64**: 348-52.

9 Maltezou HC, Giakkoupi P, Maragos A et al. Outbreak of infections due to KPC-2-producing Klebsiella pneumoniae in a hospital in Crete (Greece). *Journal of Infection* 2009; **58**: 213-9.

10 Giakkoupi P, Pappa O, Polemis M et al. Emerging Klebsiella pneumoniae Isolates Coproducing KPC-2 and VIM-1 Carbapenemases. *Antimicrobial agents and chemotherapy* 2009; **53**: 4048-50.

11 Clinical and Laboratory Standards Institute. *Performance Standards for Antimicrobial Susceptibility Testing: 24nd Informational Supplement M100-S24.* CLSI, Wayne, PA, USA, 2014.

12 Monstein HJ, ÖStholm-Balkhed Å, Nilsson MV et al. Multiplex PCR amplification assay for the detection of blaSHV, blaTEM and blaCTX-M genes in Enterobacteriaceae. *APMIS* 2007; **115**: 1400-8.

13 Poirel L, Walsh TR, Cuvillier V et al. Multiplex PCR for detection of acquired carbapenemase genes. *Diagnostic microbiology and infectious disease* 2011; **70**: 119-23.

14 Mabilat C, Goussard S, Sougakoff W et al. Direct sequencing of the amplified structural gene and promoter for the extended-broad-spectrum beta-lactamase TEM-9 (RHH-1) of Klebsiella pneumoniae. *Plasmid* 1990; **23**: 27-34.

15 Diancourt L, Passet V, Verhoef J et al. Multilocus sequence typing of Klebsiella pneumoniae nosocomial isolates. *J Clin Microbiol* 2005; **43**: 4178-82.

16 Tenover FC, Arbeit RD, Goering RV et al. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol* 1995; **33**: 2233-9.
17 Boopathi NM, & SpringerLink (Online service). . *Genetic mapping and marker assisted selection: Basics, practice and benefits*: Springer India, 2013.
18 Bankevich A, Nurk S, Antipov D et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology : a journal of computational molecular cell biology* 2012; **19**: 455-77.
19 Aziz RK, Bartels D, Best AA et al. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 2008; **9**: 75-.
20 Overbeek R, Olson R, Pusch GD et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research* 2014; **42**: D206-D14.
21 Altschul SF, Gish W, Miller W et al. Basic local alignment search tool. *Journal of molecular biology* 1990; **215**: 403-10.
22 States DJ, Gish W. Combined use of sequence similarity and codon bias for coding region identification. *Journal of computational biology : a journal of computational molecular cell biology* 1994; **1**: 39-50.
23 Cosentino S, Voldby Larsen M, Møller Aarestrup F et al. PathogenFinder - Distinguishing Friend from Foe Using Bacterial Whole Genome Sequence Data. *PloS one* 2013; **8**: e77302.
24 Tsafnat G, Copty J, Partridge SR. RAC: Repository of Antibiotic resistance Cassettes. *Database* 2011; **2011**.
25 Moura A, Soares M, Pereira C et al. INTEGRALL: a database and search engine for integrons, integrases and gene cassettes. *Bioinformatics* 2009; **25**: 1096-8.
26 Bi DX, Xu Z, Harrison EM et al. ICEberg: a web-based resource for integrative and conjugative elements found in Bacteria. *Nucleic Acids Research* 2012; **40**: D621-D6.
27 Carattoli A, Zankari E, García-Fernández A et al. In Silico Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. *Antimicrobial agents and chemotherapy* 2014; **58**: 3895-903.
28 Carriço JA, Silva-Costa C, Melo-Cristino J et al. Illustration of a Common Framework for Relating Multiple Typing Methods by Application to Macrolide-Resistant Streptococcus pyogenes. *Journal of Clinical Microbiology* 2006; **44**: 2524-32.
29 Jové T, Da Re S, Denis F et al. Inverse Correlation between Promoter Strength and Excision Activity in Class 1 Integrons. *PLoS Genet* 2010; **6**: e1000793.
30 Flannery EL, Mody L, Mobley HLT. Identification of a Modular Pathogenicity Island That Is Widespread among Urease-Producing Uropathogens and Shares Features with a Diverse Group of Mobile Elements. *Infection and Immunity* 2009; **77**: 4887-94.
31 Flannery EL, Antczak SM, Mobley HLT. Self-Transmissibility of the Integrative and Conjugative Element ICEPm1 between Clinical Isolates Requires a Functional Integrase, Relaxase, and Type IV Secretion System. *J Bacteriol* 2011; **193**: 4104-12.
32 Chen N, Ou H-Y, van Aartsen J et al. The pheV Phenylalanine tRNA Gene in Klebsiella pneumoniae Clinical Isolates Is an Integration Hotspot for Possible Niche-Adaptation Genomic Islands. *Curr Microbiol* 2010; **60**: 210-6.

33 Papagiannitsis CC, Giakkoupi P, Vatopoulos AC et al. Emergence of Klebsiella pneumoniae of a novel sequence type (ST383) producing VIM-4, KPC-2 and CMY-4 beta-lactamases. *International journal of antimicrobial agents* 2010; **36**: 573-4.
34 Samuelsen Ø, Toleman MA, Hasseltvedt V et al. Molecular characterization of VIM-producing Klebsiella pneumoniae from Scandinavia reveals genetic relatedness with international clonal complexes encoding transferable multidrug resistance. *Clinical Microbiology and Infection* 2011; **17**: 1811-6.
35 Papagiannitsis CC, Izdebski R, Baraniak A et al. Survey of metallo-β-lactamase-producing Enterobacteriaceae colonizing patients in European ICUs and rehabilitation units, 2008–11. *Journal of Antimicrobial Chemotherapy* 2015; **70**: 1981-8.
36 Robin F, Aggoune-Khinache N, Delmas J et al. Novel VIM Metallo-β-Lactamase Variant from Clinical Isolates of Enterobacteriaceae from Algeria. *Antimicrobial agents and chemotherapy* 2010; **54**: 466-70.
37 Rodriguez-Martinez J-M, Nordmann P, Fortineau N et al. VIM-19, a Metallo-β-Lactamase with Increased Carbapenemase Activity from Escherichia coli and Klebsiella pneumoniae. *Antimicrobial agents and chemotherapy* 2010; **54**: 471-6.
38 Pournaras S, Poulou A, Voulgari E et al. Detection of the new metallo-β-lactamase VIM-19 along with KPC-2, CMY-2 and CTX-M-15 in Klebsiella pneumoniae. *Journal of Antimicrobial Chemotherapy* 2010; **65**: 1604-7.
39 Barraud O, Ploy M-C. Diversity of Class 1 Integron Gene Cassette Rearrangements Selected Under Antibiotic Pressure. *J Bacteriol* 2015.
40 Zhao W-H, Hu Z-Q. Acquired metallo-β-lactamases and their genetic association with class 1 integrons and ISCR elements in Gram-negative bacteria. *Future Microbiology* 2015; **10**: 873-87.

2.6. COMPARISON OF TYPING METHODS FOR HOSPITAL-ASSOCIATED ACINETOBACTER BAUMANNII CLONES ISOLATED IN GREECE AND INDIA

Basil Britto Xavier[1], Jasmine Coppens[1], Lavanya Vanjari[2], Julia S. Sabirova[1], Christine Lammens[1], Lakshmi Vemu[2], Trevor Wagner[3], Spyros Pournaras[4], Herman Goossens[1], Surbhi Malhotra-Kumar[1]

[1]Vaccine & Infectious Disease Institute, Universiteit Antwerpen, Antwerp, Belgium; [2]Department of Microbiology, Nizam's Institute of Medical Sciences, Hyderabad, India; [3]OpGen Inc., Gaithersburg, Maryland, United States of America; [4]Department of Microbiology, Tzaneio General Hospital, Piraeus, Greece.

Abstract

A set of *A. baumannii* strains isolated at Tzaneio General Hospital (TGH) at Piraeus in Greece and primary care hospital in South India were compared using different typing methods PFGE, MLST and whole genome mapping (WGM) to search molecular patterns in order to discriminate closely related strains from baseline hospital infections. Strains were clustered based on number of unmatched fragments to assign different cluster WGM sub-types, with minimum cut-off ≥10.0 distance coefficient to assign WGM sub-types, which coincides with similarity matrix UPGMA clustering, WGM shows more discriminatory power than results obtained by pulse field gel electrophoresis (PFGE). WGM might be useful typing method with higher discriminatory ability for the most diverged strains *A. baumannii*. Emergence of new STs is reported from India and Greece.

2.6.1 Introduction

Hospital outbreaks caused by multidrug-resistant *Acinetobacter baumannii* epidemic global clones such as GC1 and GC2 are a major problem. Whole genome mapping (WGM) generating *in situ* restriction maps surmounts other existing typing techniques in as far as a high degree of resolution of linearly ordered restriction fragments is achieved via immobilization and *in situ* digestion of genomic DNA molecules. Such high resolution maps generated in a fast and reproducible manner can be useful for typing and differentiating microbial pathogens in clinical settings, such as outbreak events in hospitals, and indeed, WGM has been evaluated for genomic typing of various microbial pathogens (Alexander, et al. 2012; Hang, et al. 2012; Kotewicz, et al. 2008), yielding higher discriminatory results than obtained by pulse field gel electrophoresis (PFGE), yet with degree of resolution high enough to allow detection of more subtle genomic changes (Miller 2013b).

WGM of Gram-negative pathogens such as *Escherichia coli*, *Salmonella typhimurium*, and *Acinetobacter baumannii* proved to be a typing technique of choice in terms of capturing the enormous genome plasticity of these pathogens (Kotewicz, et al. 2008; Petersen, et al. 2011; Shukla, et al. 2009) as major genomic variations such as deletions, insertions, duplications and inversions are well detectable by WGM (Watanabe, et al. 2007) Deletions/insertions most often reflect the loss or the acquisition of phages, transposons, and integrative conjugative elements (Sabirova, et al. 2014c). Of particular clinical and epidemiological interest are genomic changes exhibited by specific pathogen strains that are associated with outbreak situations, as these could be used for the identification and characterization of outbreak-promoting pathogenic strains. Such changes should be retrievable by comparing outbreak-specific to phylogenetically linked isolates preceding or following outbreak periods.

A.baumannii is an emerging multidrug resistant (MDR) pathogen with a reputation of being a major cause of nosocomial infections and of frequent hospital outbreaks (Fournier and Richet 2006). Tracking chromosomal changes associated with outbreak-causing *A. baumannii* strains is of clinical and epidemiological importance, as recent studies (Seputiene, et al. 2012) indicate that newly acquired pathogenic features of this emerging pathogen, such as the emergence of resistance islands encoding resistance to aminoglycosides, aminocyclitols, tetracycline, chloramphenicol and carbapenemase antibiotics, are accompanied by major genomic rearrangements (Hamidian, et al. 2015; Ramirez, et al. 2011).

We utilized three different typing methods pulsed-field gel electrophoresis (PFGE), multilocus sequence typing (MLST), and whole genome mapping (WGM) to understand the genetic diversity of hospital-associated *A. baumannii* isolated from two geographically distinct regions (Greece and India) over seven years.

2.6.2 Materials and Methods

2.6.2.1 Clinical information on the *A. baumannii* strains

In this study, twenty-three *A. baumanni* strains were isolated from patients admitted to ICU (n=19), surgery (n=2) and pathology (n=2) wards (Table 2.6.1). The patients were hospitalised in Greece (Tzaneio Hospital (n=13), General hospital Serres (n=1) and University hospital Larissa (n=2)) and Nizam's Institute of Medical Sciences (NAB) in Hyderabad, India (n=6) in the period between 2008 and 2014.

2.6.2.2 Antimicrobial resistance profiling and strain typing

All 23 strains were screened for resistance to 21 antibiotics by disk diffusion and results were interpreted according to Clinical and Laboratory Standards Institute (CLSI) cut-offs. (Table 1)(Institute 2014). Multi-locus sequence typing (MLST) of *A. baumannii* was performed using both schemes pubMLST and Pasteur Institute with their respective set of seven housekeeping genes (*gltA*, *gyrB*, *gdhB*, *recA*, *cpn60*, *rpoD* and *gpi*) and (*fusA*, *gltA*, *pyrG*, *recA*, *cpn60*, *rpoB*, and *rplB*) according to their available protocols available in the *A. baumannii* MLST databases (http://pubmlst.org/abaumannii/) (Bartual, et al. 2005) and Pasteur. eBURST (version 3, http://eburst.mlst.net/) was used study diversity of the STs. The strains were screened for presence of carbapenemase presence of bla_{KPC}, $bla_{\text{OXA-48,}}$ $bla_{\text{OXA-23,}}$ $bla_{\text{OXA-58,}}$ bla_{SME}, bla_{IMP}, bla_{VIM}, $bla_{\text{SPM-1}}$, $bla_{\text{GIM-1}}$, $bla_{\text{SIM-1}}$, $bla_{\text{AIM-1}}$, bla_{NDM}, $bla_{\text{DIM-1}}$, and $bla_{\text{BIC-1}}$ in 3 multiplex PCRs (Poirel, et al. 2011b) PFGE was used to determine clonal relatedness of the isolates and was performed as described previously. PFGE band patterns were analysed using GelCompare software version 4.0 (Applied Maths, Kortrijk, Belgium).

Table 2.6.1 Clinical and strain characteristics the typed *A. baumannii* strains. next to general clinical information also resistance profiles for FEP, SXT and GM, carbapenemase resistance (OXA 23, OXA 58 and NDM), MLST (Pasteur and PUBMLST), PFGE and WGM (Combined, Greece and Indian) types. Abbreviations: ICU - intensive care unit, PU - pathology unit, SS - surgical site, CVC – central venous catheter, Wash – Bronchial wash, ETA - , FEP- Cefepime , SXT – trimethoprim-sulfamethoxazole, GM - gentamicin, R-resistant, I - intermediate, S – sensitive

Strain ID	Outbreak	Isolation date	Isolation place	Ward	Origin	FEP	SXT	GM	OXA 23	OXA-58	NDM	ST (PUBMLST)	ST (Pasteur)	PFGE	WGM
2402	YES	16/07/2008	Athens, Greece	ICU	Blood	R	R	S	Neg	Pos	negative	437	2	1e	C6c
501	NO	7/01/2012	Athens, Greece	PU	Blood	R	R	R	Pos	Neg	negative	436	645	1a	C7f
272	NO	00/05/2011	Serres, Greece	ICU	bronchial	R	R	R	Neg	Pos	negative	437	2	1j	C3
347	NO	30/01/2009	Larissa, Greece	ICU	bronchial	R	R	S	Neg	Pos	negative	350	2	1g	C2
249	NO	26/08/2008	Larissa, Greece	ICU	Pus	R	R	S	Neg	Pos	negative	437	2	1f	C4
3560	NO	25/07/2009	Athens, Greece	ICU	Sputum	R	R	R	Neg	Pos	negative	452	2	1h	C1b
4375	NO	14/06/2010	Athens, Greece	SS	Drainage	I	R	R	Neg	Pos	negative	452	2	1i	C5
2462	NO	4/08/2008	Athens, Greece	ICU	Wound	I	R	R	Neg	Pos	negative	452	2	1d	C1a
2595	YES	5/09/2008	Athens, Greece	ICU	CVC	R	R	I	Neg	Pos	negative	437	2	1e	C6b
2503	YES	14/08/2008	Athens, Greece	ICU	bronchial	I	R	S	Neg	Pos	negative	437	2	1e	C6a
3427-1	No	1/11/2013	Athens, Greece	ICU	Pleural	R	R	R	Pos	Neg	negative	425	2	1a	C7a
5801	NO	20/09/2013	Athens, Greece	SS	Wound	R	S	R	Pos	Neg	negative	436	2	1a	C7c
5636	NO	13/05/2013	Athens, Greece	ICU	Wound	R	I	R	Pos	Neg	negative	851	2	1a	C8
5392	NO	5/11/2012	Athens, Greece	PU	Blood	R	I	R	Pos	Neg	negative	208	2	1a	C7b
9733	NO	11/11/2013	Athens, Greece	ICU	Rectal	R	S	R	Pos	Neg	negative	436	2	1a	C7d
5504	NO	2/02/2013	Athens, Greece	ICU	Bronchial	R	S	R	Pos	Neg	negative	436	2	1a	C7e
NAb16	NO	17/12/2013	India	ICU	Wash	R	R	R	ND	ND	negative	195	2	1c	C11
NAb17	NO	26/12/13	India	ICU	ETA	R	R	S	ND	ND	negative	849	646	4a	C12
NAb5	NO	12/09/2013	India	ICU	Blood	R	R	R	ND	ND	NDM1/2	847	85	2a	C13
NAb7	NO	15/12/2013	India	ICU	ETA	R	R	R	ND	ND	negative	848	2	1b	C9a
NAb9	NO	14/12/2013	India	ICU	Wash	R	R	R	ND	ND	negative	848	2	1b	C9b
NAb12	NO	1/01/2014	India	ICU	ETA	R	R	R	ND	ND	negative	848	2	1b	C10
NAb15	NO	28/12/2013	India	ICU	Blood	R	R	R	ND	ND	negative	447	82	3a	C14

2.6.2.3 Whole genome mapping

The entire genomes of the 23 strains were mapped using the Argus™ Whole Genome Mapping (Opgen Inc, Gaithersburg, USA). DNA extraction, DNA quality control, DNA restriction using *NcoI*, and loading on a mapcard were done according to manufacturer's protocols (Opgen, Inc). Two methods of DNA extraction were used, agarose plug method for 13 strains and the other 10 strains by isolation of high molecular weight (HMW). Briefly, *A. baumannii* colonies grown overnight at 37°C on Muller-Hinton agar plates were immobilised in agarose plugs and subjected to gentle lysis, followed by washing and enzymatic treatment with agarose were used for isolation of high molecular weight (HMW) DNA using Argus®HMW DNA isolation kit (Opgen, Inc). DNA preps recovered from plugs were checked for the presence of high molecular weight DNA molecules by using Argus® QCard kit (Opgen, Inc) and subsequently used for mapping (Argus® MapCard II kit, Opgen, Inc). The assembly of restricted DNA molecules and identification of *NcoI* restriction sites was performed using MapManager software (Opgen, Inc)(Sabirova, et al. 2014c). Visualization, editing (adjusting starting point) and analysis of maps were performed using Bionumerics v7.5.1 (Applied Maths, Belgium) and by filtering out fragment sizes smaller than 5 kb from the analysis and an absolute tolerance of 2400 (bp), relative tolerance of 10% and mismatch size with 5000 the alignment was performed followed by clustering analysis. The dendrogram is generated by Clustering analysis, which was performed using UPGMA and similarity matrix of clusters was defined. Comparing partitions analysis was performed using online tool http://darwin.phyloviz.net/ComparingPartitions/index.php?link=Tool for Simpson's index diversity, adjusted rand and adjusted Wallace(Carriço, et al. 2006; Pinto, et al. 2008; Severiano, et al. 2011).

2.6.3 Results

Antibiotic susceptibility test showed that multi-drug resistance (resistance to β-lactams with and without β-lactamase inhibitors) was detected among all 23 *A. baumannii* strains. Of the 16 Greek strains, $bla_{\text{OXA-23}}$ was detected in 7 and $bla_{\text{OXA-58}}$ in 9 strains. $Bla_{\text{NDM-1}}$ was detected in 1 Indian strain, Nab5. MLST (Pasteur) assigned 19 of the 23 strains to ST2 (GC2) and the other 4 isolates to 4 different STs (645, 85, 646 and 82). On the other hand, PubMLST assigned the

23 strains to 13 STs. ST2 (Pasteur) strains were assigned to 10 different PubMLST types. PFGE results showed four different types. Based on a ≥ 7-band difference, all Greek and 4 Indian strains (Nab7, Nab9, Nab12 and Nab16) belonged to 1 PFGE type and were divided into 9 subtypes (1a-1i, ≥ 3-band difference). The remaining three Indian strains (Nab5, Nab15 and Nab17) were singletons (Figure 2.6.1) Based on a 95% similarity co-efficient, WGMs of the 16 Greek strains were divided into 3 clusters and 5 singletons. Overall similarity between the Greek strains was 71.3% (Figure 2.6.2a). On the other hand, WGMs of Indian strains isolated over a 5-month period during 2013-14 and from a single hospital were highly diverse. These 6 strains from one cluster and 4 singletons with an overall similarity of 36.2%, (Figure 2.6.2b). STs assigned by Pasteur MLST correlated well with PFGE

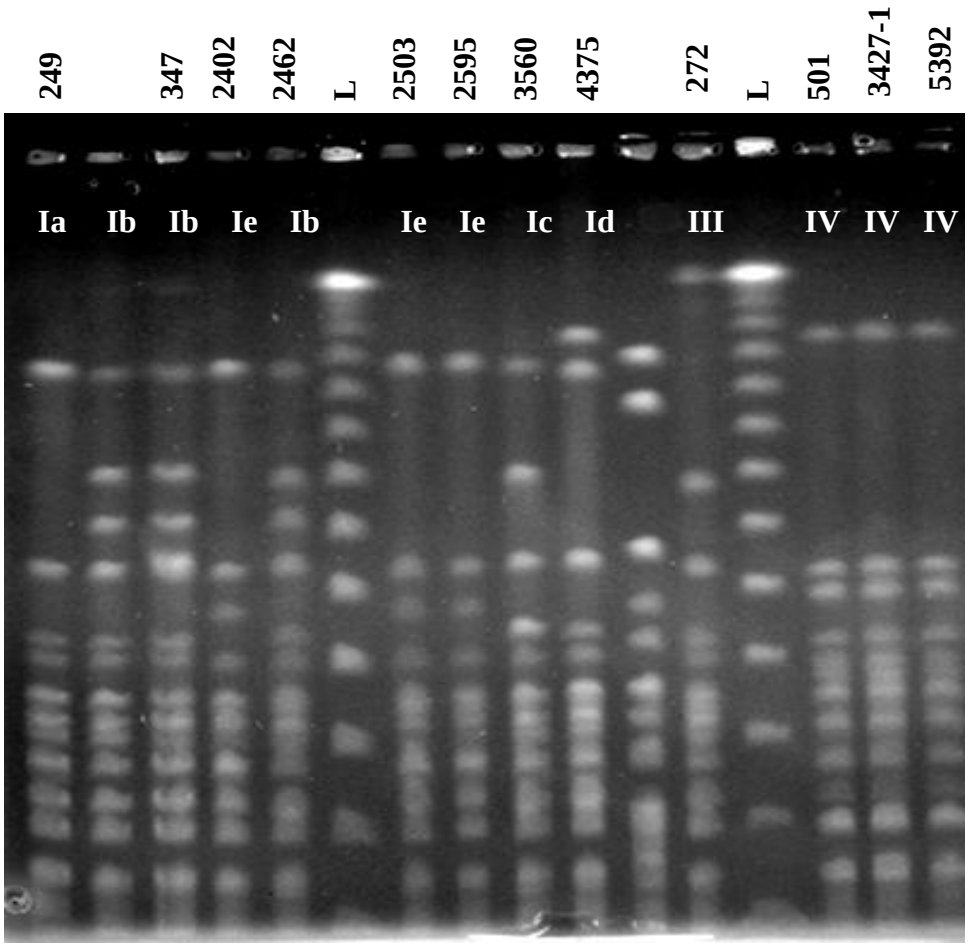


Figure 2.6.1. PFGE (>7 band differences) pattern of all analysed strains

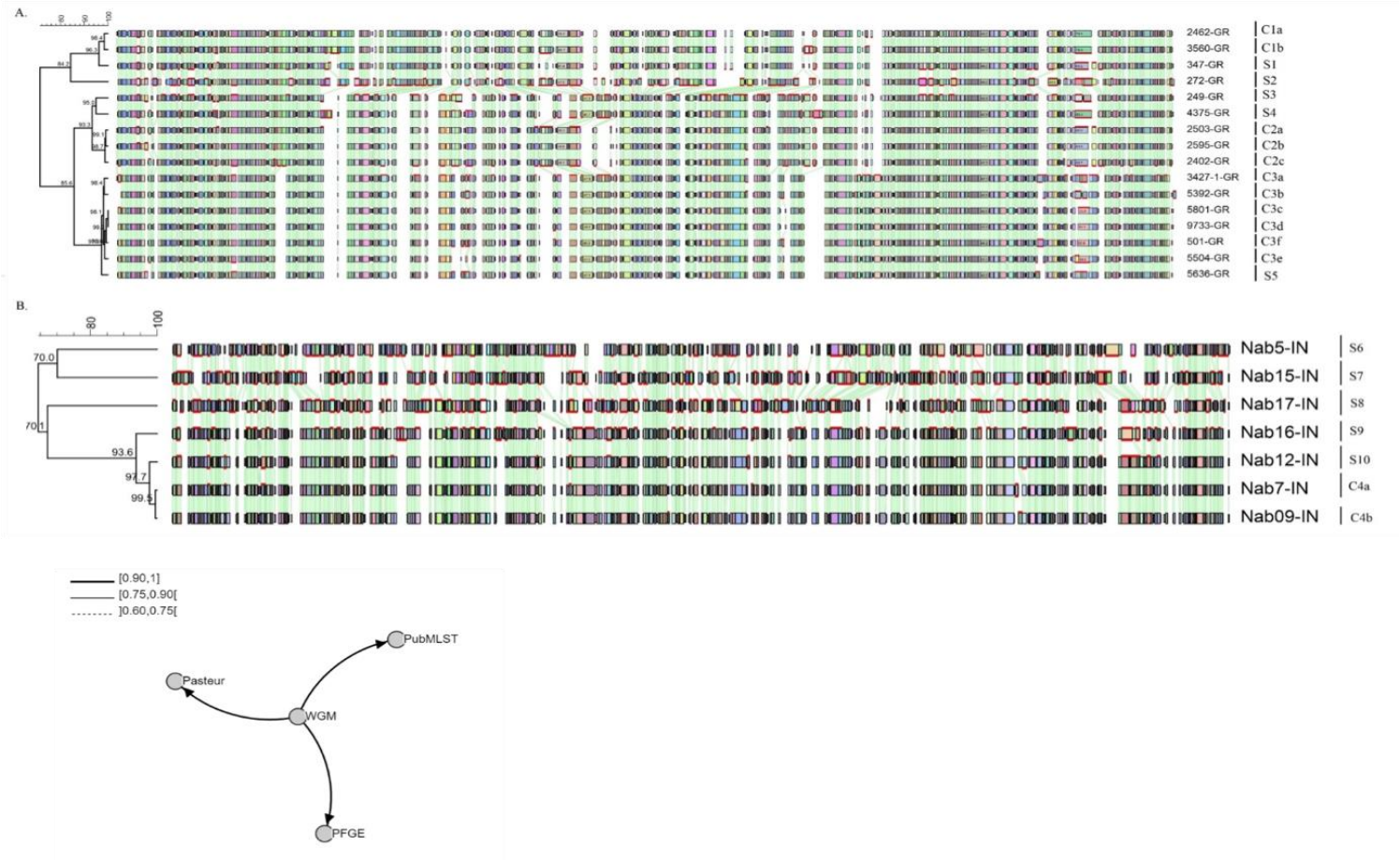


Figure 2.6.2. Map similarity cluster generated from the whole genome maps by employing Unweighted Pair Group Method with Arithmetic Mean (UPGMA). A. for Greece strains B. for Indian strains. Both alignment of WGM and a map similarity cluster were generated by Bionumerics excluding restriction fragments smaller than 3 kb and correcting the maps for the presence of inversions for the sake of overall alignment of restriction pattern similarities. C. A representation of the overall relationships between the typing methods WGM, MLST and PFGE by adjusted Wallace.

2.6.3.1 Phenotypic and genotypic characterization of *A. baumannii*

Antibiotic susceptibility test showed that severe resistance was detected among all 23 *A. baumannii* strains. Most of the typed *A. baumanni* strains shared multidrug resistance phenotype with resistance to all antibiotics tested, except 2503, 2402 249, 347 and Nab17 exhibiting sensitivity to aminoglycosides and 5504,5801 and 9733 exhibiting sensitivity to trimethoprim/ sulfamethoxazole. Intermediate phenotypes were detected within strains 4375, 2462 and 2503 had for cefepime, strains 5636 and 5392 for trimethoprim/ sulfamethoxazole and 2595 for aminoglycoside. Screening of *carbapenemase* genes shows negative results for bla_{KPC}, $bla_{OXA-48,}$ bla_{SME}, bla_{IMP}, bla_{VIM}, bla_{SPM-1}, bla_{GIM-1}, bla_{SIM-1}, bla_{AIM-1}, bla_{NDM}, and bla_{BIC-1}. But, Nab5 strain from India was positive for $bla_{NDM-1.}$ Furthermore, seven isolates with $bla_{OXA-23-like}$ genes and nine isolates with bla_{OXA-58} genes were also positive (Table 2.6.1).

MLST results for Pasteur MLST scheme, assigned 19 strains belonged to ST 2 and the 4 other isolates were divided into 4 sequence types (501 assigned to ST 645, Nab5 assigned to ST 85, Nab15 assigned to ST 82, Nab17 assigned to ST 646) (Table 1). Among them ST645 (Cpn60-2, FusA-103, GltA-2, PyrG-2, RecA-2, RplB2, RpoB-2) and ST646 (Cpn60-1, FusA-1, GltA-96, PyrG-1, RecA-5, RplB1, RpoB-2) are new STs that were identified in this study. Similarly, PubMLST scheme for all 23 strains identified 12 different ST types, ST437 (n=5); 436 (n=4); 350 (n=1); 452 (n=3); 425 (n=1); 208 (n=1); 195 (n=1); 851 (n=1); 849 (n=1); 847 (n=1); 848 (n=3) and 447 (n=1). Out of these ST847 (GltA-28, GyrB-38, GdhB-45, RecA-1,-Cpn60-30, Gpi-66, RpoD-2), ST848 (GltA-1, GyrB-15, GdhB-3, RecA-2,-Cpn60-2, Gpi-142, RpoD-3) ST849 (GltA-64, GyrB-53, GdhB-4, RecA-11,-Cpn60-4, Gpi-142, RpoD-5) and ST851 (GltA-1, GyrB-3, GdhB-3, RecA-2,-Cpn60-4, Gpi-97, RpoD-2) are new STs identified in this study. Our e-burst analysis indicated that all new ST types evolved from CC-92 and longest distance ST with 8 sub-clusters evolved ST847 (Nab5), which was apparently positive for bla_{NDM-1} and isolated from India. E-burst analysis indicated that 16 strains isolated from three different hospitals in Greece, all were originated from CC-92 with three sub-clusters (ST208, ST218 and ST684). The new ST identified from Greece isolates are evolved from sub-cluster ST208. However, seven strains from India isolated from single hospital have more diverged and multiple sub-clusters. For instance, new ST847 strain Nab5 carrying NDM-1 gene evolved from 8 sub-clusters originated from CC-92. Similarly, other new ST types ST849 (Nab17) evolved from ST109 sub-cluster, ST848 (evolved from ST451), which has three

strains belong to this new type. The sub-clusters were evolving more quickly in ST451 than other STs (Figure 3).

PFGE results show four different type profiles. All Greece and 4 Indian strains (Nab7, Nab9, Nab12 and Nab16) belongs to 1 PFGE type and can be divided in 9 subgroups (1a-1i). While the remaining three Indian strains (Nab5, Nab15 and Nab17) had different PFGE types (Table 1). The PFGE profiles coincides with the Pasteur ST types, expect for strain 501. The PFGE type 1, including all Greece and 4 Indian strains (Nab7, Nab9, Nab12 and Nab16), has 8 different subtypes. However, PubMLST based ST types diverged comparing to PFGE subtypes, suggesting that PubMLST had more discriminatory capacity than Pasteur database allelic profile.

2.6.3.2 e-burst analysis

In order to know the diversity of strains we have done e-burst based clustering based on the allelic profile database in pubMLST database. Based on our analysis, we have identified sub-strains linked to clusters were quite diverged even though all the different STs originated from same clonal complex (CC) CC92 according to pubMLST. The Indian strains belonged to different ST208, ST109, ST451 and ST1089 associated with different clusters. This suggest that heterogeneous populations from single hospital. The novel ST847 (Nab05) highly diverged in the CC92.

2.6.3.3 Comparison of WGMs generated from Greece strains.

Alignment of whole genome maps from Greece shows overall 71.33% similarity and 16 maps divided into three major clusters based on the similarity matrix (Figure 2.6.3.a). The range of similarity level was between 84.2% to 97.3%, the largest cluster with 7 strains (3427-1, 5392, 5801, 9733, 501, 5504 and 5636) among these 5636 with new ST type (ST 851 with less similar with 97.32 in this cluster), the second largest cluster (249, 4375, 2503, 2595 and 2402) with 93.25% similarity level and third cluster (2462, 3560, 347 and 272) with 84, 2% similarity level.

Figure 2.6.3. Pattern-based alignment of all WGMs of *A. baumannii* strains isolated from Greece Hospital. The alignment done using Bionumerics Restriction fragments are multi-colored to infer restriction patterns. The identical restriction patterns indicated as green vertical lines, while regions with divergence are indicated by red lines.

2.6.3.4 Comparison of WGMs generated from Indian strains

The seven WGMs from Indian strains have overall similarity of 64.3%, forming two clusters based on the similarity matrix (figure 2.6.4). Where one cluster contains (Nab09, Nab07, Nab12 clustered with 97.7% similarity, two singletons (Nab16 and Nab17) forms with 93.6% and 67.8% similarity level to other strains in the node respectively. The other two strains (Nab5 and Nab15) forms separate cluster with 70% similarity and differ from previous cluster 36% divergent level.

2.6.3.5 Combining both WGMs from Greece and India

When we combine both WGMs from India and Greece strains shows 63.9% similarity and forms 4 different clusters based on similarity matrix and one wgm's with singleton (Figure 2.6.2). As we expected strains from Greece forms in one cluster and Indians strains form another cluster. Interesting observation was strain Nab16 (ST 195 (pubMLST) and ST2 (Pasteur) was clustering with Greece strains 97.8% similarity (most of the strains in the node was ST436). The largest distance matrix among these strains was Nab5 and Nab15 from India with 194.3.

2.6.3.6 Comparing partitions analysis of different typing methods.

Application of the adjusted Wallace index showed that WGM clusters were strongly predictive of both MLSTs and PFGE types (Wallace co-efficient of 1.00 for both) (Figure 2.6.2C). Simpson's diversity index showed that discriminatory power of WGM (1.000, 95% CI: 0.976-0.99) was higher than that of PFGE (0.893, 95% CI: 0.801-0.986) and PubMLST (0.913, 95% CI: 0.857-0.970).

Figure 2.6.4. Pattern-based alignment of all WGMs of *A. baumannii* strains isolated from Indian Hospital. The alignment done using Bionumerics Restriction fragments are multi-colored to infer restriction patterns. The identical restriction patterns indicated as green vertical lines, while regions with divergence are indicated by red lines.

2.6.4 Discussion

A. baumannii has become one of the quickly spreading strain in the hospital environment across the world. The development of multi-drug resistance level in a strain is determined mainly by the genes acquired, genes transferred between clones and the ability of the strains that spread quickly within population from integration of the resistance genes into the chromosomes. The most recent plasmid borne resistance gene, NDM-1 was first identified in *A. baumannii* and was found to be integrated into the chromosome suggesting the genotypic plasticity of *A. baumannii*. The predominant epidemic clone GC2 expanded rapidly and disseminated worldwide with wider differences within the CC-92 complex and many sub-clusters emerged within this GC2 clone. Molecular typing methods such as PFGE and MLST are still crucial for strain differentiation as the existence of known Pasteur MLST typing scheme genes are not conserved anymore. Similarly, different set of genes introduced by Oxford pubMLST also lacks evidence to discriminate the clonality of the strains. Since the genomic evolution rate is higher in *A. baumannii*, the gene-based approach is no longer stable. Therefore, new typing approach like WGM define a new era in molecular typing and epidemiological surveillance. WGM preserves the chromosome order and the large number of fragments generated can be compared and the genomic variations can be easily determined, as is not the case with PFGE where the band size and number of fragments generated has an impact on defining the PFGE subtypes.

In this study, we have characterised clonality of *A. baumannii* strains isolated from different hospitals in Greece and one hospital from India. The strains were compared with different typing methods such as PFGE and MLST. The results were compared to high resolution whole genome mapping (WGM) and the strains were discriminated at different levels using both the methods Comparison of the high-resolution typing technique WGM to other typing methods (MLST and PFGE) showed its superior discriminatory power for analysis of rapidly evolving hospital-associated *A. baumannii* global clones. The high diversity among the Indian strains studied here was noteworthy.

Author contribution: BBX was involved in experimental work, data analysis, interpretation and drafting manuscript.

2.6.5 References

1. Alexander DC, Hao W, Gilmour MW et al. Escherichia coli O104:H4 infections and international travel. *Emerging infectious diseases* 2012; **18**: 473-6.
2. Hang J, Clifford RJ, Yang Y et al. Genome sequencing of pathogenic Rhodococcus spp. *Emerging infectious diseases* 2012; **18**: 1915-6.
3. Kotewicz ML, Mammel MK, LeClerc JE et al. Optical mapping and 454 sequencing of Escherichia coli O157 : H7 isolates linked to the US 2006 spinach-associated outbreak. *Microbiology* 2008; **154**: 3518-28.
4. Miller JM. Whole-Genome Mapping: a New Paradigm in Strain-Typing Technology. *J Clin Microbiol* 2013; **51**: 1066-70.
5. Shukla SK, Kislow J, Briska A et al. Optical Mapping Reveals a Large Genetic Inversion between Two Methicillin-Resistant Staphylococcus aureus Strains. *J Bacteriol* 2009; **191**: 5717-23.
6. Petersen RF, Litrup E, Larsson JT et al. Molecular characterization of Salmonella Typhimurium highly successful outbreak strains. *Foodborne pathogens and disease* 2011; **8**: 655-61.
7. Watanabe S, Ito T, Morimoto Y et al. Precise excision and self-integration of a composite transposon as a model for spontaneous large-scale chromosome inversion/deletion of the Staphylococcus haemolyticus clinical strain JCSC1435. *J Bacteriol* 2007; **189**: 2921-5.
8. Sabirova JS, Xavier BB, Ieven M et al. Whole genome mapping as a fast-track tool to assess genomic stability of sequenced Staphylococcus aureus strains. *BMC Research Notes* 2014; **7**: 704.
9. Fournier PE, Richet H. The epidemiology and control of Acinetobacter baumannii in health care facilities. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2006; **42**: 692-9.
10. Seputiene V, Povilonis J, Suziedeliene E. Novel variants of AbaR resistance islands with a common backbone in Acinetobacter baumannii isolates of European clone II. *Antimicrobial agents and chemotherapy* 2012; **56**: 1969-73.
11. Ramirez MS, Adams MD, Bonomo RA et al. Genomic Analysis of Acinetobacter baumannii A118 by Comparison of Optical Maps: Identification of Structures Related to Its Susceptibility Phenotype. *Antimicrob Agents Ch* 2011; **55**: 1520-6.
12. Hamidian M, Holt KE, Hall RM. Genomic resistance island AGI1 carrying a complex class 1 integron in a multiply antibiotic-resistant ST25 Acinetobacter baumannii isolate. *The Journal of antimicrobial chemotherapy* 2015; **70**: 2519-23.
13. Institute CaLS. Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically; approved standard—24th ed. CLSI document M100-S24. *CLSI, Wayne, PA, USA, 2014* 2014.
14. Bartual SG, Seifert H, Hippler C et al. Development of a multilocus sequence typing scheme for characterization of clinical isolates of Acinetobacter baumannii. *J Clin Microbiol* 2005; **43**: 4382-90.

15. Poirel L, Walsh TR, Cuvillier V et al. Multiplex PCR for detection of acquired carbapenemase genes. *Diagn Micr Infec Dis* 2011; **70**: 119-23.
16. Severiano A, Pinto FR, Ramirez M et al. Adjusted Wallace Coefficient as a Measure of Congruence between Typing Methods. *J Clin Microbiol* 2011; **49**: 3997-4000.
17. Pinto FR, Melo-Cristino J, Ramirez M. A Confidence Interval for the Wallace Coefficient of Concordance and Its Application to Microbial Typing Methods. *Plos One* 2008; **3**.
18. Carriço JA, Silva-Costa C, Melo-Cristino J et al. Illustration of a Common Framework for Relating Multiple Typing Methods by Application to Macrolide-Resistant Streptococcus pyogenes. *Journal of Clinical Microbiology* 2006; **44**: 2524-32.

2.7 COMPLETE GENOME SEQUENCES OF NITROFURANTOIN-SENSITIVE AND RESISTANT *ESCHERICHIA COLI* ST540 AND ST2747 STRAINS

Basil Britto Xavier,[a] Jascha Vervoort,[a] Andrew Stewardson,[b] Niels Adriaenssens,[a] Samuel Coenen,[a] Stephan Harbarth,[b] Herman Goossens,[a] and Surbhi Malhotra-Kumar[a]

[a]Laboratory of Medical Microbiology, Vaccine & Infectious Disease Institute, Universiteit Antwerpen, Antwerp, Belgium [b]University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland

Abstract

Widespread multidrug resistance in *Escherichia coli* has necessitated the reintroduction of older antibiotics, such as nitrofurantoin. However, mechanisms by which resistance to nitrofurantoin emerges in *E. coli* are not well elucidated. Toward this aim, we sequenced two nitrofurantoin-sensitive *E. coli* sequence types (ST540 and ST2747) and their four nitrofurantoin-resistant derivatives generated *in vitro* under aerobic and anaerobic growth conditions.

2.7.1 Genome Announcement

Escherichia coli is the most common cause of urinary tract infections (UTI) in the community (1). Multidrug resistance in *E. coli* has necessitated the reintroduction of older antibiotics, like nitrofurantoin. Nitrofurantoin acts by undergoing reduction by bacterial nitroreductases to generate toxic derivatives that attach to ribosomal proteins (2), causing defective transcription and translation in bacteria (3). First described in the 1970s, nitrofurantoin resistance in *E. coli* involves loss-of-function mutations in two genes encoding nitroreductases, *nfsA* and *nfsB* (4, 5). Although these mechanisms are well known, no previous studies have focused beyond the *nfs* genes.

We generated nitrofurantoin-resistant isolates *in vitro* under aerobic and anaerobic conditions from two nitrofurantoin-sensitive *E. coli* strains (multilocus sequence types [MLST] ST540 and ST2747, for which MICs of nitrofurantoin are 16 and 4 µg/ml, respectively) derived from stool samples from two Belgian outpatients with UTI. The strains were subjected to three stepwise plating's on Mueller-Hinton agar supplemented with increasing nitrofurantoin concentrations (0.5- to 4-fold MIC for the parent strain). Whole-genome sequencing (Pacific Biosciences, Menlo Park, CA, USA) was done on the two parental and four nitrofurantoin-resistant strains generated under aerobic conditions (strains ST540-A and ST2747-A, with nitrofurantoin MICs of 256 and 128 µg/ml, respectively) and anaerobic conditions (strains ST540-AN and ST2747-AN, with nitrofurantoin MICs of 64 and 32 µg/ml, respectively). Genomic DNA was isolated with the Master Pure complete DNA and RNA purification kit (Epicentre, Madison, WI, USA), according to the manufacturer's protocol. Library preparation and sequencing reactions were performed using the PacBio DNA template prep kit 2.0 (3 kb to 10 kb) and the PacBio DNA sequencing kit 2.0 with C2 chemistry. Sequence runs for six single-molecule real-time (SMRT) cells were performed on the PacBio RS II sequencer with a 1 × 180-minute movie time/SMRT cell. The SMRT Analysis portal version 2.0 was used for filtering the reads and sub reads, with default parameters, and post filtered data of ~320 Mb on each cell/per strain were considered for assembly. The six genomes were assembled using the Hierarchical Genome Assembly Process (HGAP), which is available with the SMRT Analysis packages and accessed through the SMRT Analysis Portal version 2.0. All six strains were sequenced with ~28-fold coverage spanning ~14 scaffolds for each strain. The scaffolds were ordered against an *in silico* whole-genome map of *E. coli* ATCC 8739 (GenBank accession no. CP000946) using the MapSolver software (Whole Genome Mapping; OpGen, Gaithersburg, MD) to derive a complete gapless chromosome (B. B. Xavier, J. Sabirova, P. Moons, J.-P.

Hernalsteens, H. De Greve, H. Goossens, and S. Malhotra-Kumar, submitted for publication). The chromosome sizes of the ST540 strains are 4,758,628 bp, 4,807965 bp, and 4,875,674 bp. The ST2747 strains have chromosome sizes of 5,054,424 bp, 4,998,910 bp, and 5,090,326 bp, with a G+C content of ~51%. We used the Rapid Annotations using Subsystems Technology (RAST) server (6) and NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) (7) for function-based annotation for all strains. The genomes of ST540 and ST2747 contain 4,587 and 4,859 protein-coding genes and 114 and 119 RNA genes, respectively, of which 89 encode tRNAs and 8 encode rRNAs. In addition, our genome annotation confirmed the duplication of the mobile elements and other single-nucleotide polymorphisms (SNPs) in ST540 and ST2747, which were identified using an inbuilt tool in the CLC Genomics Workbench 6.5.1 (CLC, Inc., Aarhus, Denmark).

2.7.2 Nucleotide sequence accession numbers.

The genome sequences for all six strains have been deposited at DDBJ/EMBL/GenBank under the accession no. CP007265, CP007390, CP007391, CP007392, CP007393, and CP007394. The sequences described in this paper are the first versions.

Author contribution: BBX was involved in experimental work, data analysis, interpretation and drafting manuscript.

2.7.3 References

1. Manges AR, Johnson JR, Foxman B, O'Bryan TT, Fullerton KE, Riley LW. 2001. Widespread distribution of urinary tract infections caused by a multidrug-resistant *Escherichia coli* clonal group. N. Engl. J. Med. 345:1007–1013. 10.1056/NEJMoa011265
2. McOsker CC, Fitzpatrick PM. 1994. Nitrofurantoin: mechanism of action and implications for resistance development in common uropathogens. J. Antimicrob. Chemother. 33(Suppl A):23–30. 10.1093/jac/33.suppl_A.23
3. McCalla DR. 1964. Effects of some nitrofurans on DNA synthesis and prophage induction. Can. J. Biochem. 42:1245–1247. 10.1139/o64-134
4. Asnis RE. 1957. The reduction of furacin by cell-free extracts of furacin-resistant and parent-susceptible strains of *Escherichia coli*. Arch. Biochem. Biophys. 66:208–216. 10.1016/0003-9861(57)90551-9
5. McCalla DR, Kaiser C, Green MH. 1978. Genetics of nitrofurazone resistance in *Escherichia coli*. J. Bacteriol. 133:10–16
6. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75. 10.1186/1471-2164-9-75
7. Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity G, Kodira CD, Kyrpides N, Madupu R, Markowitz V, Tatusova T, Thomson N, White O. 2008. Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation. OMICS 12:137–141. 10.1089/omi.2008.0017

CHAPTER III

APPLICATIONS OF NEXT GENERATION SEQUENCING (NGS) FOR ELUCIDATING COLISTIN RESISTANCE MECHANISM

Chapter III is further divided into the following sub-chapters:

3.1 Identification of a novel plasmid-mediated colistin resistant gene *mcr-2* in *E. coli*. (**Xavier BB**, Lammens C, Ruhal R, Butaye P, Kumar-Singh S, Goossens H, Malhotra-Kumar S. *Euro surveillance. 2016*. http://dx.doi.org/10.2807/1560-7917.ES.2016.21.27.30280. **IF:7.2**)

3.2 A systems biology approach identifies pathways mediating colistin resistance in *Klebsiella pneumoniae* (**Xavier BB**, Plantinga N, Coppens J, Minh N, Zarkotou, O. Janssens L, Lammens C, Lal S, Kumar-Singh S, Bonten M, Tsakris A, Pournaras S, Goossens H, Malhotra-Kumar S) (*In preparation*).

3.3 Complete sequence of an IncFII plasmid harbouring the colistin resistance gene *mcr-1* isolated from Belgian pig farms. (**Xavier BB**, Lammens C, Butaye P, Goossens H, Malhotra-Kumar S. *Journal of Antimicrobial Chemotherapy*. 2016 June 3, 2016. **IF:5.313.)**

3.4 Colistin resistance gene *mcr-1* harboured on a multidrug-resistant plasmid. (Malhotra-Kumar S, **Xavier BB**, Das AJ, Lammens C, Butaye P, Goossens H. *The Lancet Infectious Diseases*. 2016. **IF:22.433**)

3.5 Colistin-resistant *Escherichia coli* harbouring *mcr-1* isolated from food animals in Hanoi, Vietnam. (Malhotra-Kumar S, **Xavier BB,** Das AJ, Lammens C, Hoang HTT, Pham NT, et al. *The Lancet Infectious Diseases*. 2016. **IF: 22.433**).

3.6 Identification of *mcr-1* gene in *Escherichia coli* ST88 from human isolate harboured in IncHI2 plasmid. (**Xavier BB**, Bonten M, Goossens H, Malhotra-Kumar S. (*In preparation*).

3.1. IDENTIFICATION OF A NOVEL PLASMID-MEDIATED COLISTIN-RESISTANCE GENE, MCR-2, IN *ESCHERICHIA COLI,* BELGIUM, JUNE 2016

Xavier BB[1], Lammens C[1], Ruhal R[1], Butaye P [2,3,4], Kumar-Singh S[3], Goossens H[1], Malhotra-Kumar S[1].
[1]Laboratory of Medical Microbiology, University of Antwerp, Wilrijk, Belgium. [2]Faculty of Veterinary Medicine, Ghent University, Ghent, Belgium [3]CODA-CERVA, Brussels, Belgium [4]Ross University School of Veterinary Medicine, ST Kitts and Nevis, West Indies. [5]Laboratory of Molecular Pathology, Cell Biology and Histology, Universiteit Antwerp, Antwerp, Belgium

Abstract

We identified a novel plasmid-mediated colistin-resistance gene in porcine and bovine colistin-resistant *Escherichia coli* that did not contain *mcr-1*. The gene, termed *mcr-2*, a 1,617 bp phosphoethanolamine transferase harboured on an IncX4 plasmid, has 76.7% nucleotide identity to *mcr-1*. Prevalence of *mcr-2* in porcine colistin-resistant *E. coli* (11/53) in Belgium was higher than that of *mcr-1* (7/53). These data call for an immediate introduction of *mcr-2* screening in ongoing molecular epidemiological surveillance of colistin-resistant Gram-negative pathogens.

3.1.1 Rapid Communication

Following the report of of *mcr-1* detection in China in November 2015 (Liu et al. 2015), we screened 105 colistin-resistant *Escherichia coli* (colistin minimum inhibitory concentration (MIC) 4–8 mg/L (EUCAST, 2016) isolated during 2011–12 from passive surveillance of diarrhoea in 52 calves and 53 piglets in Belgium (Malhotra-Kumar et al. 2016). *mcr-1* was detected in 12.4% (n = 13) of the *E. coli* isolates, of which six and seven were from calves and piglets, respectively (Malhotra-Kumar et al. 2016; Xavier et al. 2016). In the present study, we analysed porcine and bovine colistin-resistant *Escherichia coli* isolates that did not show the presence of *mcr-1* and identified a novel plasmid-mediated colistin resistance-conferring gene, *mcr-2*.

Of 92 porcine and bovine colistin-resistant *Escherichia coli* isolates not harbouring *mcr-1*, 10 were randomly selected for further analysis. Plasmid DNA was isolated (PureLink HiPure Plasmid Miniprep Kit, Invitrogen, Carlsbad, CA, United States), sequenced by Illumina (2 x 250 bp) (Nextera XT sample preparation kit, MiSeq), de novo assembled and annotated using SPAdes (v3.8.1) and RAST (Bankevich et al. 2012; Aziz et al. 2008). Plasmids from three of the 10 *E. coli* isolates showed the presence of a gene for a putative membrane protein, which was identified as a phosphoethanolamine transferase (sulfatase) using pfam and Interproscan protein databases (Finn et al. 2016; Quevillon et al. 2005) The *mcr-2* gene, as we termed it, is 1,617 bp long, nine bases shorter than *mcr-1* (1,626 bp), and shows 76.75% nt identity to *mcr-1* (supplementary material (Xavier, Lammens, Ruhal, et al. 2016).

The entire *mcr-2* gene was amplified (PCR primers: MCR2-F 5′ TGGTACAGCCCCTTTATT 3′; MCR2-R 5′ GCTTGAGATTGGGTTATGA 3′), cloned (vector pCR 2.1, TOPO TA Cloning kit, Invitrogen) and electroporated into DH-5 α *E. coli*. Transformants exhibited colistin MICs of 4–8 mg/L (E-test, bioMerieux, Marcy l'Etoile, France), which were reconfirmed by macrobroth dilution (European Committee on Antimicrobial Susceptibility Testing (EUCAST) guidelines (EUCAST, 2016)).

3.1.2 *mcr-2* is harboured on IS1595 with likely origins in *Moraxella* spp.

mcr-2-harbouring plasmids from all three *E. coli* isolates were analysed. The mobile element harbouring *mcr-2* was identified as an IS element of the IS1595 superfamily, which are distinguished by the presence of an ISXO2-like transposase domain (Siguier, Gagnevin, and Chandler 2009).

We also identified a 297 bp open reading frame downstream of *mcr-2* on this element, which encodes a PAP2 membrane-associated lipid phosphatase with 41% identity to *Moraxella osloensis* phosphatidic acid phosphatase (71% query coverage). Interestingly, a blastn search of the IS1595 backbone, after removal of the *mcr-2* and *pap2* phosphatase gene sequences, identified a single hit to *Moraxella bovoculi* strain 58069 (GenBank accession number CP011374) genomic region (1,531,602 to 1,532,255 bp) with 75% identity and 100% query coverage.

3.1.3 *mcr-2* is harboured on an IncX4 incompatibility-type plasmid in *E. coli* ST10

The three *mcr-2* plasmids-harbouring *E. coli* isolates belonged to ST10 (n = 2, porcine) and ST167 (n = 1, bovine). All three plasmids belonged to IncX4 incompatibility type; all three *mcr-2* genes showed 100% homology.

Plasmid pKP37-BE isolated from one of the porcine ST10 *E. coli* isolates was found to have a size of 35,104 bp, 41.3% GC content and 56 protein-encoding gene sequences (RAST) (Figure 1); European Nucleotide Archive accession numbers PRJEB14596 (study) and LT598652 (plasmid sequence).

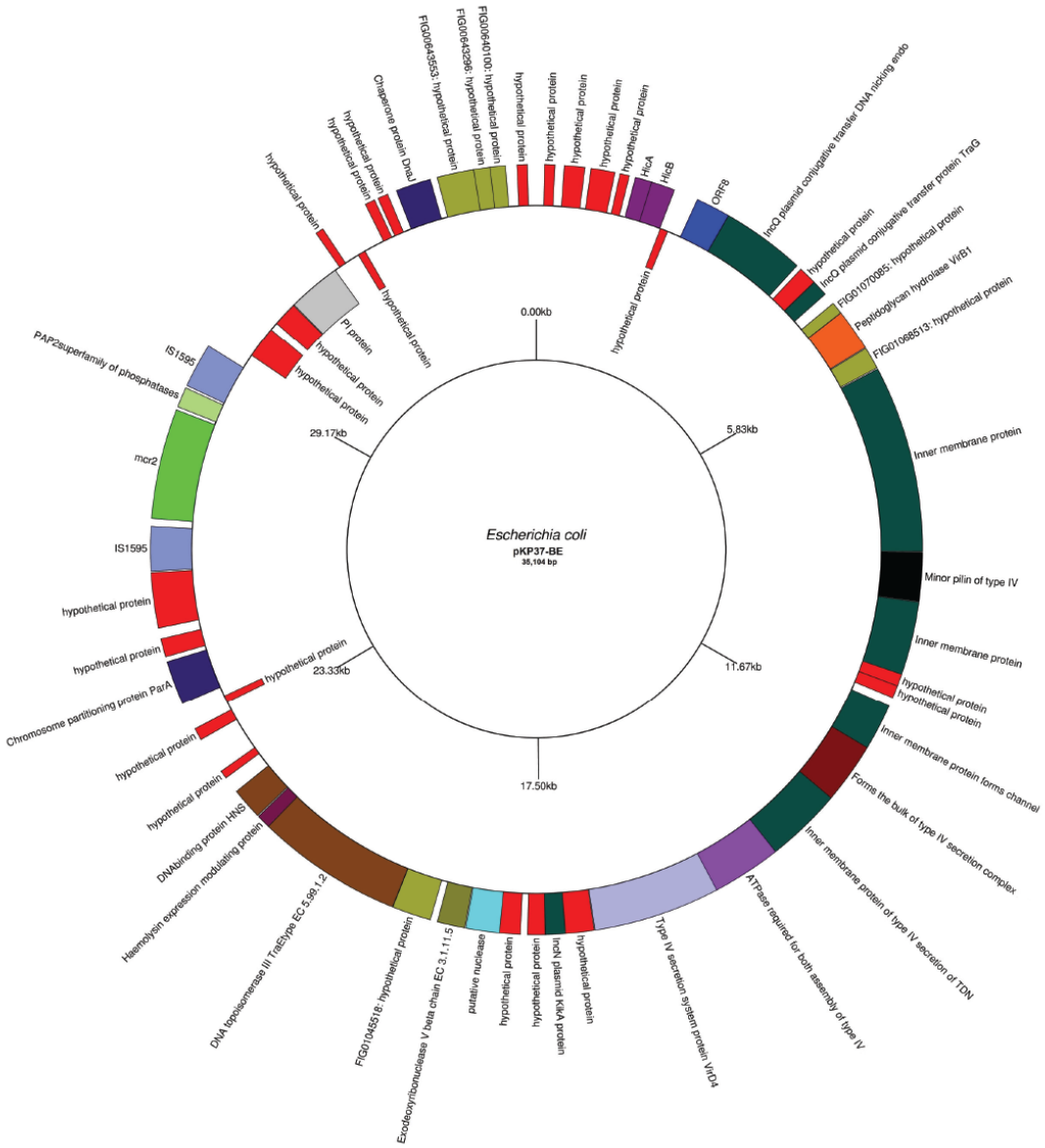


Figure 3.1.1 Genetic organisation and structure of the *mcr-2*-harbouring plasmid pKP37-BE from a colistin-resistant *Escherichia coli* isolate not harbouring *mcr-1*, Belgium, June 2016. The plasmid map was generated using GenomeVx (http://wolfe.ucd.ie/GenomeVx/).

Apart from IS1595, pKP37-BE did not carry any other resistance genes and the plasmid backbone was highly similar to *Salmonella enterica* subsp. *enterica* serovar Heidelberg plasmid pSH146_32 (GenBank accession number JX258655), with 98% identity and 90% query coverage. Several *Salmonella*-associated virulence genes were found on pKP37-BE, including

virB/*D4* that encodes a type 4 secretion system (Gokulan et al. 2013).

Conjugation experiments using a rifampicin-resistant *E. coli* recipient (A15) showed an approximately 1,200-fold higher transfer frequency of the *mcr-2*-harbouring pKP37-BE (1.71×10^{-3}) compared with the *mcr-1* harbouring IncFII plasmid, pKP81-BE (1.39×10^{-6}) Xavier et al. 2016). Both *mcr-1* and *mcr-2* transconjugants exhibited colistin MICs of 4–8 mg/L (macrobroth dilution).

3.1.4 Structure predictions and phylogenetic analyses of the MCR-2 protein

MCR-2 protein was predicted to have two domains, with domain 1 (1 to 229 residues) as a transporter and domain 2 (230 to 538 residues) as a transferase domain (Figure 2).

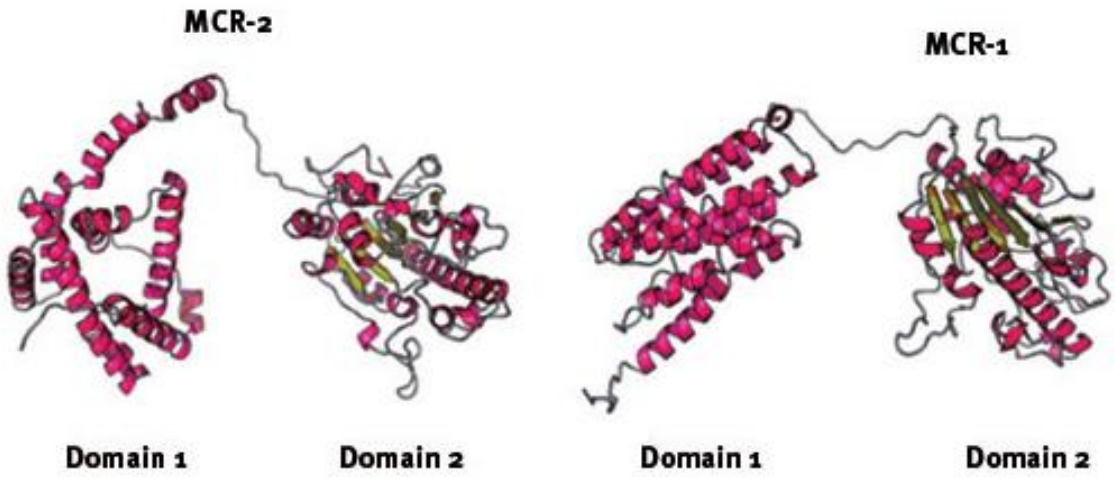


Figure 3.1.2 MCR-2 and MCR-1 predicted tertiary structures. RaptorX (Källberg et al. 2014) was used to generate the structures. For both MCR-2 and MCR-1, domain 1 was predicted to be a transporter and domain 2 a phosphoethanolamine transferase (sulfatase).

The best template for domain 1 was 4HE8, a secondary membrane transport protein with a role in transferring solutes across membranes (Berrisford, Baradaran, and Sazanov 2016). The best-fit template for domain 2 was 4kav ($p = 4.13$ e-13), a lipooligosaccharide phosphoethanolamine transferase A from *Neisseria meningitides*, also previously shown to be the best-fit template for MCR-1 (Liu et al. 2015). 4kav belongs to the YhjW/YjdB/YijP superfamily and its role in conferring polymyxin resistance has been experimentally validated

(Wanty et al. 2013). Overall, the un-normalised global distance test (uGDT) was 318 (GDT: 58) and all 538 residues were modelled (Figure 2).

MCR-1 and MCR-2 proteins showed 80.65% identity (supplementary material (Xavier, Lammens, Ruhal, et al. 2016). In addition, MCR-2 showed 64% identity to the phosphoethanolamine transferase of *Moraxella osloensis* (WP_062333180) with 99% sequence coverage, and 65%, 65%, and 61% identity to that of *Enhydrobacter aerosaccus* (KND21726), *Paenibacillus sophorae* (WP_063619495) and *Moraxella catarrhalis* (WP_003672704), respectively, all with 97% query coverage.

We also carried out blastp searches of the two domains of MCR-2 separately. The identity level of domain 1 between MCR-1 and MCR-2 was low (72%) compared with that for domain 2 (87.4%). Other blastp hits for the domain 2 transferase were *Enhydrobacter aerosaccus* and *Moraxella osloensis* (69% identity; 100% query coverage) followed by *Paenibacillus sophorae* (68% identity; 100% query coverage) and *Moraxella catarrhalis* (68% identity; 99% query coverage). Phylogenetic analysis showed that MCR-2 might have originated from *Moraxella catarrhalis* (56% bootstrap value) (Figure 3).

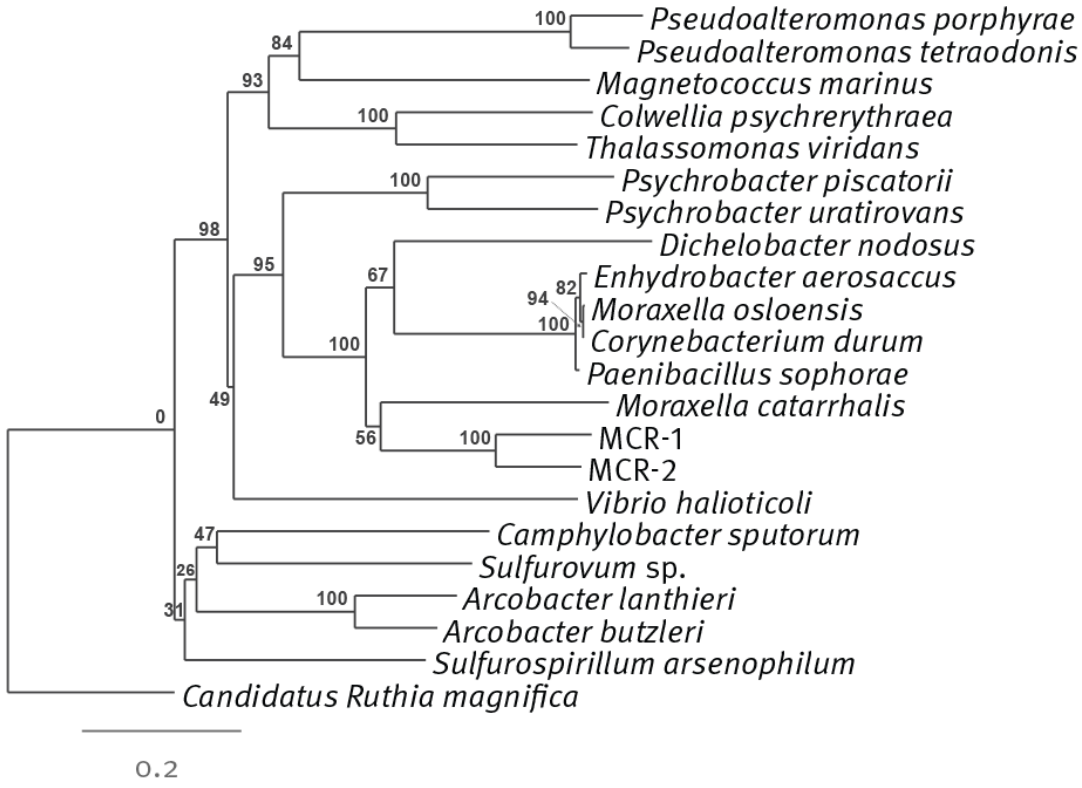


Figure 3.1.3 Phylogenetic analysis of the entire MCR-2 protein sequence

Maximum likelihood tree generated by bootstrap analysis using 1,000 replicates. The analysis was carried out using CLC Genomics workbench v9.0.1 (clcbio, Qiagen) in-built tool for this evolutionary relationship with other related sequences. Branch length is proportional to the number of substitutions per site. Bootstrap values are indicated in the nodes.

3.1.5 PCR-based screening identified a higher prevalence of *mcr-2* than of *mcr-1* in porcine *E. coli* in Belgium

We screened our entire collection of porcine and bovine colistin-resistant *E. coli* isolates (n = 105) using an *mcr-2*-specific PCR approach using primers MCR2-IF 5' TGTTGCTTGTGCCGATTGGA 3' and MCR2-IR 5' AGATGGTATTGTTGGTTGCTG 3', and the following cycling conditions: 33 cycles of 95 °C × 3 min, 65 °C × 30 s, 72 °C × 1 min, followed by 1 cycle of 72 °C × 10 min. We found *mcr-2* in 11/53 porcine and 1/52 bovine colistin-resistant *E. coli* isolates (overall prevalence of 11.4%).

3.1.6 Discussion

Identification of plasmid-mediated colistin resistance represents a paradigm shift in colistin-resistance mechanisms, which until recently were restricted to chromosomal mutations and vertical transmission. Since *mcr-1* conferring plasmid-mediated colistin resistance was first detected in China, *mcr-1* has been identified in 28 countries across five continents (Fernandes et al. 2016; Skov and Monnet 2016; Prim et al. 2016; Perreten et al. 2016) (Figure 4)

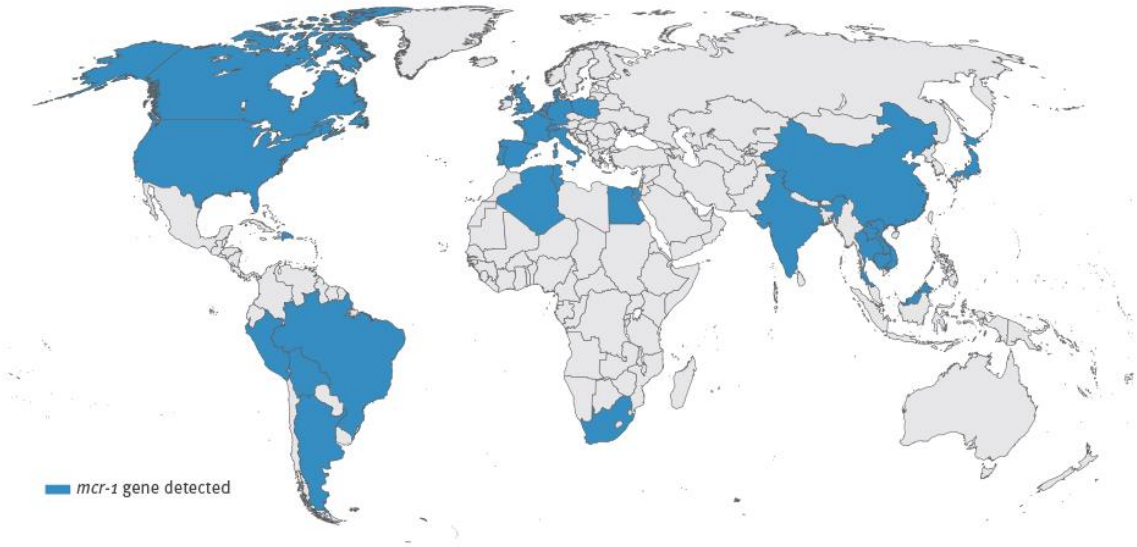


Figure 3.1.4 Countries (n = 32) reporting presence of *mcr-1* in samples of animal, environmental or human origin (data collected till 27 June 2016). Adapted from (Skov and Monnet 2016) updated using data from (Fernandes et al. 2016; Prim et al. 2016; Perreten et al. 2016; Rapoport et al. 2016; Izdebski et al. 2016; McGann et al. 2016).

Our analysis identified a novel plasmid-mediated phosphoethanolamine transferase-encoding gene, *mcr-2*, which was detected at an even higher prevalence than that of *mcr-1* among colistin-resistant porcine *E. coli* in our study. We were, however, limited by small sample numbers. It should also be noted that the calves and piglets were from different regions of the country (calves from Wallonia and piglets from Flanders).

Phylogenetic analysis of MCR-2 provided strong evidence that this protein was distinct from MCR-1, and that it might have originated from *Moraxella catarrhalis*. The latter set of data are further strengthened by the fact that *mcr-2* is co-harboured with a lipid phosphatase gene that shows the highest homology to a phosphatase from *Moraxella* spp., and that the genetic element IS1595 harbouring these two genes might itself have originated from *Moraxella* spp. While *Moraxella* spp. are not polymyxin producers, this bacterial genus is known to be intrinsically resistant to polymyxins (D R Storm, K S Rosenthal, and Swanson 1977), and potential intergeneric transfer of *mcr-2* from co-habiting *Moraxella* spp. of animal, human or environmental origin is therefore highly likely. Phosphoethanolamine transferases are housekeeping enzymes that catalyse the addition of the phosphoethanolamine moiety to the outer 3-deoxy-D-manno-octulosonic acid (Kdo) residue of a Kdo(2)-lipid A (Reynolds et al. 2005). The fact that we did not identify any chromosomal mutations in the known colistin resistance-conferring genes in our *E. coli* isolates (by whole-genome sequencing, data not shown) additionally supports the role of the acquired phosphoethanolamine transferase in conferring colistin resistance.

Finally, the high transfer frequency of the *mcr-2*-harbouring IncX4 plasmid might underlie the higher prevalence of *mcr-2* in our porcine isolates. In the three *mcr-2* harbouring isolates analysed, IS1595 showed the presence of direct repeats and a complete *tnpA* gene, while inverted repeats were not found (data not shown). However, the carrier plasmid IncX4 is itself highly transmissible, showing 10^2–10^5-fold higher transfer frequencies than, for instance, epidemic IncFII plasmids, as shown previously (Lo et al. 2014) as well as in our own transconjugation experiments. Importantly, a lack of fitness-burden of IncX4 carriage on bacterial hosts (Lo et al. 2014) makes this plasmid replicon a highly effective vehicle for dissemination of *mcr-2*. IncX4 plasmids have also been previously shown to harbour *mcr-1*

(Hasman et al. 2015) as well as extended spectrum beta-lactamase genes, *bla*$_{\text{CTX-M}}$ (Lo et al. 2014). Interestingly, the pKP37-BE backbone, which likely originated from *Salmonella* spp., harboured a battery of virulence genes including the *virB4/D4* genes encoding a type-IV secretion system that has been shown to mediate downregulation of host innate immune response genes and an increased bacterial uptake and survival within macrophages and epithelial cells (Gokulan et al. 2013). Outer membrane modifications leading to colistin resistance have been shown to attenuate virulence (Vila-Farrés et al. 2015): whether these co-harboured virulence genes are able to compensate the pathogenic abilities of colistin-resistant *E. coli* remains to be explored.

Taken together, these data call for immediate inclusion of *mcr-2* screening in ongoing molecular epidemiological surveillance to gauge the worldwide dissemination of *mcr-2* in both human and animal colistin-resistant Gram-negative bacteria of medical importance.

Author contribution: BBX was involved in data analysis, interpretation and drafting manuscript.

3.1.7 References

1. Liu Y-Y, Wang Y, Walsh TR, Yi L-X, Zhang R, Spencer J, et al. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. Lancet Infect Dis. 2016;16(2):161-8. http://dx.doi.org/10.1016/S1473-3099(15)00424-7 PMID:26603172
2. European Committee on Antimicrobial Susceptibility Testing (EUCAST). Breakpoint tables for interpretation of MICs and zone diameters. Version 6.0, 2016. EUCAST; 2016. Available from: http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/v_6.0_Breakpoint_table.pdf
3. Malhotra-Kumar S, Xavier BB, Das AJ, Lammens C, Butaye P, Goossens H. Colistin resistance gene mcr-1 harboured on a multidrug resistant plasmid. Lancet Infect Dis. 2016;16(3):283-4. http://dx.doi.org/10.1016/S1473-3099(16)00012-8 PMID:26774247/
4. Xavier BB, Lammens C, Butaye P, Goossens H, Malhotra-Kumar S. Complete sequence of an IncFII plasmid harbouring the colistin resistance gene mcr-1 isolated from Belgian pig farms. J Antimicrob Chemother. 2016; dkw191. http://dx.doi.org/10.1093/jac/dkw191 PMID:27261261/
5. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455-77. http://dx.doi.org/10.1089/cmb.2012.0021 PMID:22506599
6. 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(1):75. http://dx.doi.org/10.1186/1471-2164-9-75 PMID:18261238
7. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44(D1):D279-85. http://dx.doi.org/10.1093/nar/gkv1344 PMID:26673716
8. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, et al. InterProScan: protein domains identifier. Nucleic Acids Res. 20051;33(Web Server issue):W116-20. PMID:15980438
9. Xavier BB, Lammens C, Ruhal R, Kumar-Singh S, Butaye P, Goossens H, et al. Rapid Communication, Eurosurveillance, Identification of a novel plasmid-mediated colistin-resistant gene mcr-2 in E. coli. Supplementary material. [Accessed 7 Jul 2016]. Available from: https://www.uantwerpen.be/en/rg/lab-of-medical-microbiology/projects-and-publications/publications/key-publications/rapid-communication-eurosurveillance/
10. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 2006;34(Database issue):D32-6. http://dx.doi.org/10.1093/nar/gkj014 PMID:16381877
11. Gokulan K, Khare S, Rooney AW, Han J, Lynne AM, Foley SL. Impact of plasmids, including those encodingVirB4/D4 type IV secretion systems, on Salmonella enterica serovar Heidelberg virulence in macrophages and epithelial cells. PLoS One. 2013;8(10):e77866. http://dx.doi.org/10.1371/journal.pone.0077866 PMID:24098597

12. Berrisford JM, Baradaran R, Sazanov LA. Structure of bacterial respiratory complex I. Biochim Biophys Acta. 2016;1857(7):892-901. http://dx.doi.org/10.1016/j.bbabio.2016.01.012 PMID:26807915
13. Wanty C, Anandan A, Piek S, Walshe J, Ganguly J, Carlson RW, et al. The structure of the neisserial lipooligosaccharide phosphoethanolamine transferase A (LptA) required for resistance to polymyxin. J Mol Biol. 2013;425(18):3389-402. http://dx.doi.org/10.1016/j.jmb.2013.06.029 PMID:23810904
14. Fernandes MR, Moura Q, Sartori L, Silva KC, Cunha MP, Esposito F, et al. Silent dissemination of colistin-resistant Escherichia coli in South America could contribute to the global spread of the mcr-1 gene. Euro Surveill. 2016;21(17) 16;21(17):pii=30214.
15. Skov RL, Monnet DL. Plasmid-mediated colistin resistance (mcr-1 gene): three months later, the story unfolds. Euro Surveill. 2016;21(9):30155. http://dx.doi.org/10.2807/1560-7917.ES.2016.21.9.30155 PMID:26967914
16. Prim N, Rivera A, Rodriguez-Navarro J, Espanol M, Turbau M, Coll P, et al. Detection of mcr-1 colistin resistance gene in polyclonal Escherichia coli isolates in Barcelona, Spain, 2012 to 2015. Euro Surveill. 2016;21(13) (13):pii=30183.
17. Perreten V, Strauss C, Collaud A, Gerber D. Colistin Resistance Gene mcr-1 in Avian-Pathogenic Escherichia coli in South Africa. Antimicrob Agents Chemother. 2016;60(7):4414-5. http://dx.doi.org/10.1128/AAC.00548-16 PMID:27161625
18. Storm DR, Rosenthal KS, Swanson PE. Polymyxin and related peptide antibiotics. Annu Rev Biochem. 1977;46(1):723-63. http://dx.doi.org/10.1146/annurev.bi.46.070177.003451 PMID:197881
19. Reynolds CM, Kalb SR, Cotter RJ, Raetz CRH. A phosphoethanolamine transferase specific for the outer 3-deoxy-D-manno-octulosonic acid residue of Escherichia coli lipopolysaccharide. Identification of the eptB gene and Ca2+ hypersensitivity of an eptB deletion mutant. J Biol Chem. 2005;280(22):21202-11. http://dx.doi.org/10.1074/jbc.M500964200 PMID:15795227
20. Lo WU, Chow KH, Law PY, Ng KY, Cheung YY, Lai EL, et al. Highly conjugative IncX4 plasmids carrying blaCTX-M in Escherichia coli from humans and food animals. J Med Microbiol. 2014;63(Pt 6):835-40. http://dx.doi.org/10.1099/jmm.0.074021-0 PMID:24595536
21. Hasman H, Hammerum AM, Hansen F, Hendriksen RS, Olesen B, Agersø Y, et al. Detection of mcr-1 encoding plasmid-mediated colistin-resistant Escherichia coli isolates from human bloodstream infection and imported chicken meat, Denmark 2015. Euro Surveill. 2015;20(49):30085. http://dx.doi.org/10.2807/1560-7917.ES.2015.20.49.30085 PMID:26676364
22. Vila-Farrés X, Ferrer-Navarro M, Callarisa AE, Martí S, Espinal P, Gupta S, et al. Loss of LPS is involved in the virulence and resistance to colistin of colistin-resistant Acinetobacter nosocomialis mutants selected in vitro. J Antimicrob Chemother. 2015;70(11):2981-6. http://dx.doi.org/10.1093/jac/dkv244 PMID:26311838
23. Conant G, Wolfe K. GenomeVx. Dublin: University College Dublin. [Accessed 22 Jun 2016]. Available from: http://wolfe.ucd.ie/GenomeVx/

24. Källberg M, Margaryan G, Wang S, Ma J, Xu J. RaptorX server: a resource for template-based protein structure modeling. Methods Mol Biol. 2014;1137:17-27. http://dx.doi.org/10.1007/978-1-4939-0366-5_2 PMID:24573471
25. Rapoport M, Faccone D, Pasteran F, Ceriana P, Albornoz E, Petroni A, et al.; MCR Group. First Description of mcr-1-Mediated Colistin Resistance in Human Infections Caused by Escherichia coli in Latin America. Antimicrob Agents Chemother. 2016;60(7):4412-3. http://dx.doi.org/10.1128/AAC.00573-16 PMID:27090181
26. Izdebski R, Baraniak A, Bojarska K, Urbanowicz P, Fiett J, Pomorska-Wesołowska M, et al. Mobile MCR-1-associated resistance to colistin in Poland. J Antimicrob Chemother. 2016 Jun 20. pii: dkw261. 10.1093/jac/dkw261 PMID:27330064
27. McGann P, Snesrud E, Maybank R, Corey B, Ong AC, Clifford R, et al. Escherichia coli Harboring mcr-1 and blaCTX-M on a Novel IncF Plasmid: First Report of mcr-1 in the United States. Antimicrob Agents Chemother. 2016;60(7):4420-1. http://dx.doi.org/10.1128/AAC.01103-16 PMID:2723079

3.2 A SYSTEMS BIOLOGY APPROACH IDENTIFIES PATHWAYS MEDIATING COLISTIN RESISTANCE IN KLEBSIELLA PNEUMONIAE

Xavier BB[1], Plantinga N[3], Coppens J[1], Minh N[1], Zarkotou O[4], Janssens L[1], Lammens C[1], Lal S[2], Kumar-Singh S[1], Bonten M[3], Tsakris A[4], Pournaras S[4], Goossens H[1], Malhotra-Kumar S[1]

[1] Laboratory of Medical Microbiology, Vaccine & Infectious Disease Institute, Universiteit Antwerpen, Antwerp, Belgium; [2]Department of Microbiology, Aster MIMS, Calicut, India; [3]Department of Medical Microbiology and Epidemiology of Infectious Diseases, Julius Center for Health Sciences & Primary Care, UMC Utrecht, The Netherlands; [4]Department of Microbiology, Tzaneio General Hospital, Piraeus, Greece.

Abstract

Colistin is currently increasingly used to treat infections caused by Gram-negative bacteria and emergence of colistin resistance cannot be avoided. *K. pneumoniae* is one of the primary cause of hospital-acquired infections. In this study, we had collected 100 colistin-resistant *K. pneumoniae* strains from patients admitted to hospitals across Europe, Israel, Vietnam, and India. All strains were sequenced using Nextera XT kit via 2X250b paired-end sequencing (Miseq, Illumina). Additionally, three strains were sequenced using Pacbio long read technology followed by comparative genome analysis. RNA sequencing was performed on colistin-sensitive and resistant strains with different *mgrB* modifications. Transcriptomic analysis was carried out using CLC Genomics Workbench v9.5.3 (clcbio, Denmark). Colistin MICs ranged from 4, and 256 µg/ml for resistant strains and the strains belonged to different sequence type's ST383, ST101, ST512, ST258, ST147 and ST11. We show here that colistin pressure might trigger the transposition event leading to interruption in the *mgrB* gene and novel *mgrB* and *pmrB* variants were determined. Based on the transcriptomic analysis of colistin-resistant strains we suggest that different metabolic pathways are involved in colistin resistance mechanisms.

3.2.1 Introduction

The immense increase in antibiotic resistance among critical bacterial pathogens, driven by the uncontrolled use of antibiotics (1), is currently recognised as one of the most pressing threats to human health by the World Health Organization (WHO) (2). Notably, the last decade has seen a significant rise in infections caused by multidrug- and extremely drug-resistant (MDR and XDR) Gram-negative pathogens, such as *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Acinetobacter baumannii* (3, 4). In the last decade, the scarcity of antimicrobials and emergence of pan-drug resistant bacterial pathogens led to the reintroduction of the polymyxins (colistin) and nitrofurantoin as one of the last line antibiotics to treat nosocomial and other infections caused by MDR *K. pneumoniae, Acinetobacter* spp and *E. coli*. In some countries, colistin is the only drug to which these bacteria show sensitivity and therefore, any emergence of resistance to these antibiotics is of paramount importance (4-6). Inevitably, regular use of colistin has also resulted in the emergence of colistin resistance in *K. pneumoniae* and *E. coli*. Colistin, like other cationic antibiotics, targets the bacterial cell membrane and resistance develops through adaptive or mutational mechanisms, with almost complete cross-resistance existing between colistin and other polymyxins. Colistin acts by displacing divalent cations from the phosphate groups of the outer membrane lipids during interaction with the membrane lipopolysaccharide (LPS) and phospholipids, which have a net negative charge and are a primary target for colistin (7). In *K. pneumoniae,* modifications in *mgrB*, which encodes a short transmembrane protein that is involved in a negative feedback loop to the PhoP/PhoQ regulatory system. Interestingly, observations on *mgrB* modifications have shown to mediate colistin resistance (8). The capsular genes of *K. pneumoniae*, as well as an efflux pump *kpnEF* (9), might also be involved in potentiating colistin resistance by preventing colistin binding to target sites (7). Adding to these mechanisms, mobile colistin resistance (*mcr)* genes harboured on plasmids in *E. coli* heralded the identification of transferrable colistin resistance mechanisms (10) (11). These prior data underscore the variety and complexity of mechanisms involved in mediating colistin resistance and urge the use of multiple strategies to elucidate exact mechanistic of colistin resistance.

Therefore, utilized a combination of genomics, transcriptomics to elucidate impact in the transcriptomic profile and the involvement of multiple metabolic pathways impacted in the colistin-resistant strains using omics-based strategy.

3.2.2 Materials and Methods

3.2.2.1 Strains and clinical information

A total of 109 clinical *K. pneumoniae* strains (100 colistin-resistant and nine sensitive) were utilized in this study. Of these, 78 colistin-resistant *K. pneumoniae* strains were isolated from clinical samples [blood (n=2) and aspirate (n=25)] and screening samples [throat (n=2) and rectal swabs (n=49)] between August 2014 - March 2017 from patients admitted to intensive care units (ICUs) in Belgium (3 ICUs, n=30), Spain (2 ICUs, n=29) and Italy (2 ICUs, n=19) as part of a clinical trial during in an EU-FP7 project (Resistance in Gram-Negative Organisms: Studying Intervention Strategies, R-GNOSIS). Further, 19 *K. pneumoniae* were isolated from patients admitted to the ICU or surgical ward at the TG hospital, Greece during 2009-2013. Among these, both colistin sensitive (n=9) and –resistant (n=10) strains were isolated from clinical samples [blood (n=7), CVC (n=2), bronchial (n=2), wound (n=2), urine (n=5) and urinary catheter (n=1)] and also included isogenic 'clinical' pairs of colistin sensitive and resistant *K. pneumoniae* isolated from 4 patients before and after colistin therapy (clinical details of each patient are given as Supplementary information). Additionally, 8 colistin-resistant isolates were collected from clinical samples [blood (n=3) and wound (n=5)] at a tertiary care hospital in India (Malabar Institute of Medical Science-MIMS, Calicut) between May 2015─November 2016 from patients admitted to the surgical wards. Furthermore, colistin-resistant *K. pneumoniae* random hospitals from Israel (n=1), Vietnam (urine; n=2), and Belgium (endotracheal tube; n=1) were obtained.

Colistin sensitive strains (3319S & 3789S) from Greece were utilised to obtain resistance strains under colistin pressure. Of these, 2 (3319-4R & 3319-10R, MICs to colistin, 64 µg/mL) strains were sequenced and utilised to observe colistin resistance in a 'clean' background. Table 3.2.1 gives an inventory of the strain origin and other characteristics.

3.2.2.2 MIC phenotype & resistome profile

Colistin minimum inhibitory concentrations (MIC)s of all investigated strains was determined by macro broth tube method, and the results were interpreted according to the EUCAST guidelines, 2017 (http://www.eucast.org). WGS was utilised to determine resistance gene profiles, and those detected by WGS were confirmed by PCR and Sanger sequencing.

3.2.2.3 Generation and analysis of *in vitro* resistant clones under colistin pressure and stability studies

Two colistin sensitive *K. pneumoniae* (3319S, MIC 1 µg/mL and 3789S, MIC 2 µg/mL) were used to generate *in vitro* clones. The two strains were serially passaged under sub- and supra-MICs (0.5, 0.75, 1, 2, and up to 64 µg/ml) to colistin to generate resistant mutants *in vitro* as described (16). Briefly, sensitive strains were grown overnight in MHB with 0.5 µg/mL colistin. 200 µl of the overnight culture was inoculated into 2 ml of MHB containing 2 µg/mL colistin. This procedure was repeated by passaging daily into the fresh medium by gradually increasing colistin concentrations, reaching up to 64 µg/mL. Three serial passages (≈ 115 generations) were required for isolation of both single-step and multistep mutants (Figure 3.2.1). *In vitro* generated resistant clones from 3319S (2R, 4R, 8R and 10R) and 3789S (1R, 2R, 4R and 9R) were sub-cultured from plates with 64 µg/mL colistin concentrations. MICs were performed to confirm colistin resistance using broth macro dilution method. To assess the resistant phenotypic stability of the *in vitro* generated resistant clones, 20 colonies each were randomly selected from MH agar with colistin concentration 64 µg/ml along with *in vivo* resistant strain [3359-R (CL MIC 32 µg/ml)]. All the strains (n=41) were replicated in antibiotic-free MHA for ~18 hours (1 replication or passage ~ 38 generations) for 160 days (~6500 generations). From every ~100th generation, ten colonies were randomly picked, subcultured and stored in 2 µg/mL colistin broth. The stability of acquired resistance was checked for every ~ 200 generations, 20 colonies were selected randomly and were tested for colistin MICs.

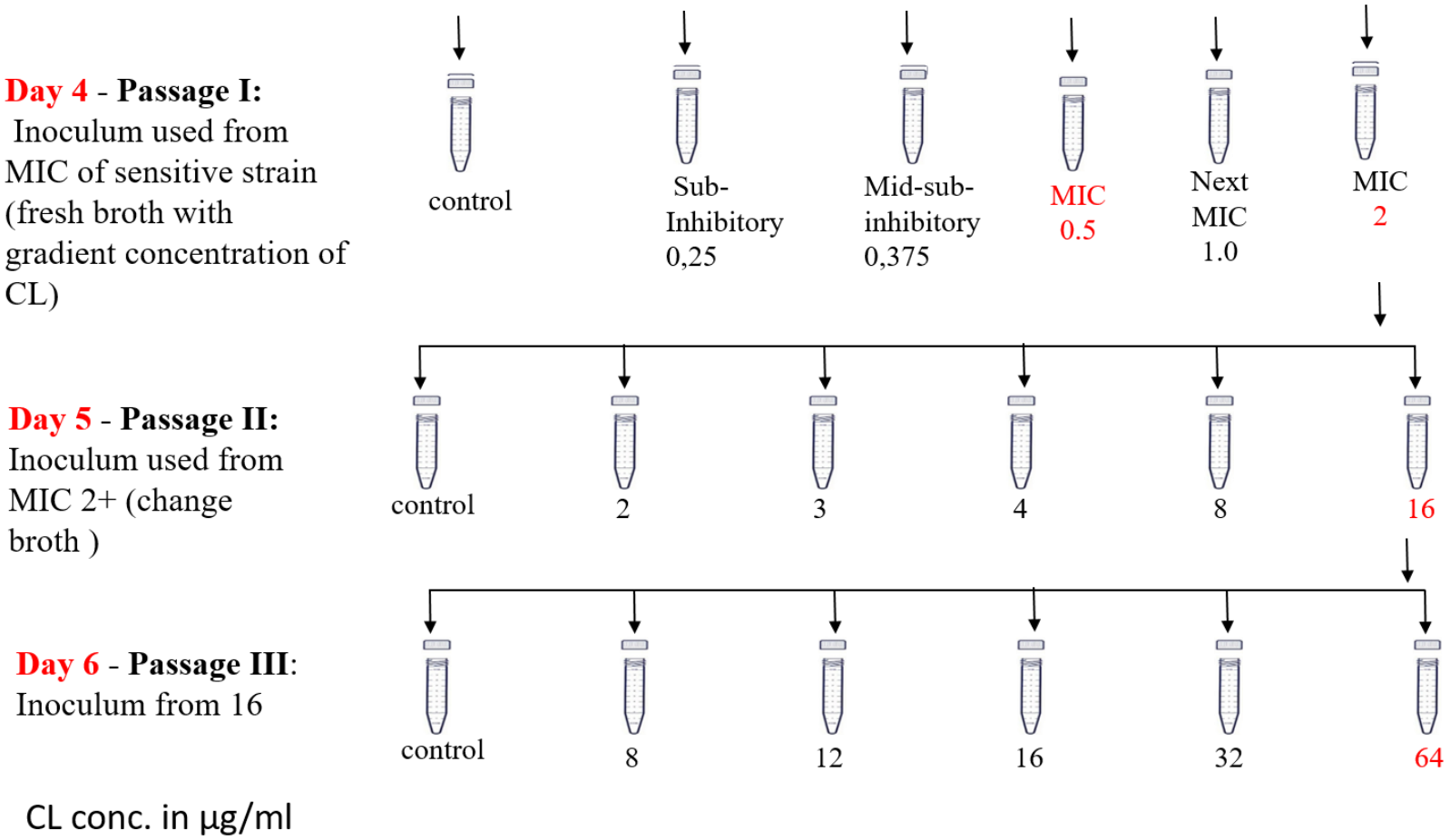


Figure 3.2.1 Flow diagram of *in vitro* CL resistant strains.

3.2.2.4 Population analysis of *in vitro*-generated colistin-resistant *K. pneumoniae*

In vitro clones from the strain, 3319S were further screened by population analysis to determine the frequency and type of transposition under colistin pressure. Briefly, 20 colonies were randomly picked from each colistin concentration (2, 3, 8, 12, 16, 32, 64 µg /mL) and sub-cultured on MHA plates. A total of 140 colonies were picked from different generations (G5, G6, G7, G24, G34, G55, G68, G75, G84, G96, G107, G107, G117, G118, G132, G151, G161) and *mgrB* modifications were determined using specific primer pairs by PCR. The following were the primers used for detecting the *mgrB* interruption, forward primer (5'TCTGAGTCCACAGCAACAGG3') and reverse primer (5'AGAGAAACTCCACCACTTTA 3'). For identification of *mgrB* promoter interruptions, *mgrB*-F1 (5'ATAACACCCCATAACCGTC3') and *mgrB*-R1 (5'AGAGAAACTCCACCACTTTA 3') primers were used.

3.2.2.5 Elucidation of biological cost of colistin resistance by competition assays

We studied fitness cost of colistin resistance in *in vivo* pairs (clinical isogenic) and *in vitro* generated resistant strains by competing with their parental strains on growth competition assay. Both, the clinical colistin-resistant (3359R) and *in vitro* colistin-resistant mutants (3319-2R, 4R, 8R and 10R) were competed against the clinical colistin sensitive (3319S) *K. pneumoniae* as described (17, 18). The competition assay was done in triplicates for every pair. Strains were thoroughly mixed and were plated on MHA plates with and without colistin (8 µg /mL). Serial dilutions were performed, and the number of colonies for each dilution was counted and compared among the pairs. Relative fitness of each competed strain pair was determined based on the ratio of the number of generations grown by both strains (Table 3.2.2).

3.2.2.6 Whole genome sequencing (WGS) and analysis

All the strains were subjected to WGS as follows: DNA was isolated, and after QC assessment of DNA, libraries were prepared using Nextera XT for short read and SMRT bell for long read (19, 20) (Figure 3.2.2).

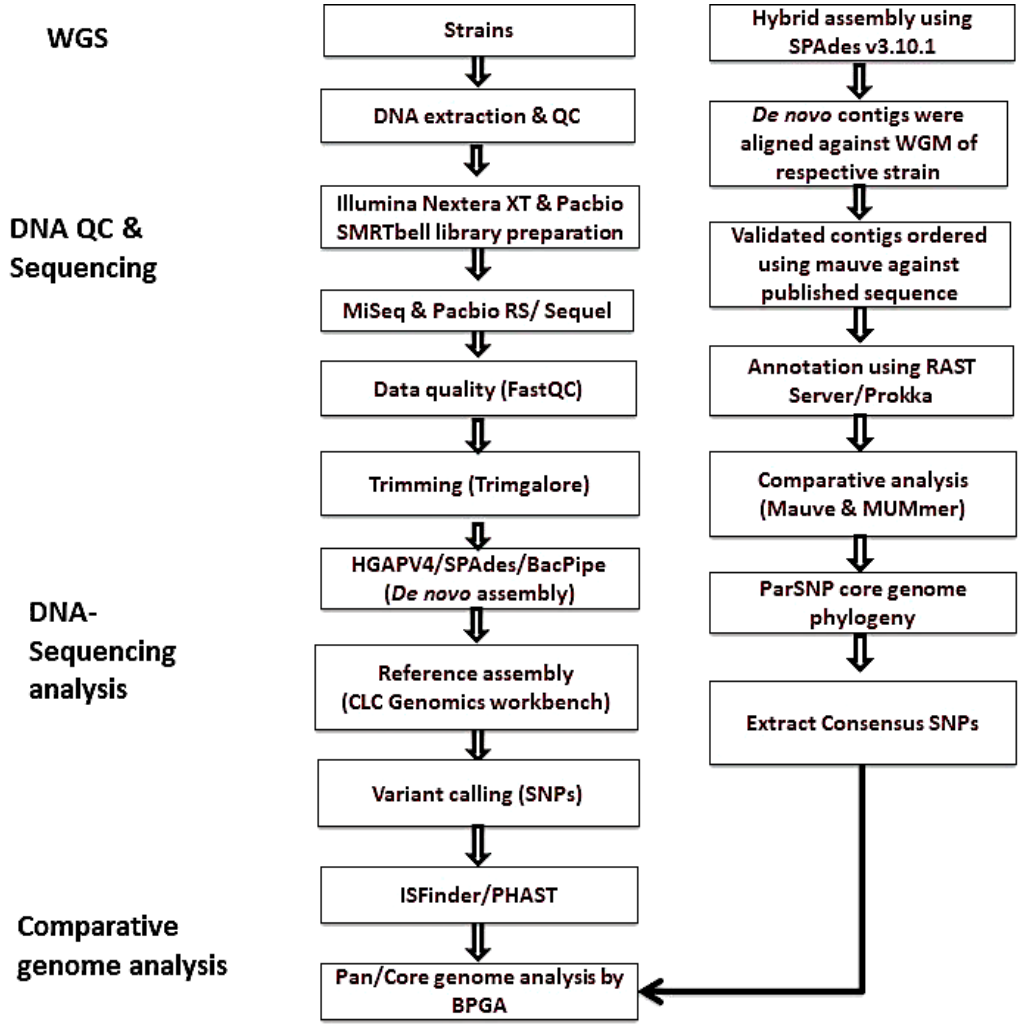


Figure 3.2.2 WGS analysis workflow for both short and long reads.

Briefly, for short read sequencing, genomic DNA was isolated all strains (n=109) using the MasterPure Complete DNA and RNA purification kit (Epicentre, Madison, WI, USA), according to the manufacturer's protocol. DNA quality was checked on Qubit™ DNA and RNA Assay (Life Technologies, Thermo Fisher Scientific Inc). Libraries were prepared using Nextera XT for short read, followed by Illumina (2 x 250 bp) paired-end sequencing. For long read sequencing (pacbio) [n=3 (3319S, 3359R and 3319- 10R)], genomic DNA was isolated using Qiagen® MagAttract® HMW kit (Qiagen) according to the manufacturer's protocol. Isolated genomic DNA was sheared using Covaris G-tubes to fragment size distributions around 8-20kb, and barcoded SMRT bell™ libraries were established following the Pacific Biosciences protocol. Then the libraries were pooled in equimolar amounts, subsequently annealed with SMRT Sequencing Primer v3 and complexed with Sequel polymerase version 2.0. The DNA/Polymerase complex was purified using a Chromaspin TE400 column (ClonTech), and

2.5 fmol complexes were loaded onto a Sequel v2 SMRT cell and sequenced using Sequel sequencing chemistry 2.0 (10h collection, Magbead loading). Raw sequences were processed through the PacBio Biosciences' software suite SMRTLink 4.0. Long read sequences were assembled using the HGAP4 and hybrid assembly with short reads were done using SPAdes v3.10.1 and further analysis was done using in-house developed BacPipe v1.0. Online tools for detecting insertion sequence (IS) elements (https://www-is.biotoul.fr/), genomic islands (www.pathogenomics.sfu.ca/islandviewer/) and prophage (phaster.ca/) (Figure 3.2.3) were used. The *mgrB* genetic modifications were analysed using CLC Genomics Workbench v9.5.3 (clcbio, Denmark) (Figure 3.2.3). Pan-genome analysis was done to delineate the number of core genes, accessory genes and Kyoto Encyclopaedia Genes and Genomes (KEGG) distribution across different ST types using Bacteria Pan Genome Analysis (BPGA) Pipeline (21). The provean (protein variation effect analyser) score was determined for every mutation in the *mgrB* gene (22).

3.2.2.7 Whole Genome Mapping for genome assembly validation

The whole genome maps of isogenic strains (3 pairs) and 2 *in vitro* clones were generated and utilised for validating short read genome assembly. Briefly, after DNA extraction and quality check, restriction digestion was performed using *AflII*, and the DNA was loaded onto a map card following the manufacturer's protocol (23). The scaffolds were ordered against *in silico* whole-genome maps of the strains aligned using the MapSolver software (Whole Genome Mapping; OpGen, Gaithersburg, MD) to validate genome assembly and a complete gapless chromosome of the selected strains was achieved.

3.2.2.8 Transcriptomic profiling of colistin-resistant strains by RNA-Sequencing

Twelve strains were selected for RNA sequencing (six colistin sensitive and six resistant). Further, to understand gene induction under colistin pressure, two colistin-resistant strains (MIC 16 and 32 µg/mL), each grown under colistin pressure (8 µg/ml) were subjected to RNA sequencing. All strains were grown to the log-phase at 37^{o}C in MHB broth. Total RNA was isolated with the MasterPure complete DNA and RNA purification kit (Epicentre, Madison, WI, USA), according to the manufacturer's protocol. Isolated RNA was quantified using Qubit™ DNA and RNA Assay (Life Technologies, Thermo Fisher Scientific Inc) followed by Ribo depletion, Stranded TruSeq RNA library preparation and sequencing was performed using NextSeq500, v2, 1x75bp (Illumina Inc.,). Data were analysed as follows; the quality of the read

was checked, and the adapters and low-quality reads were trimmed. A reference-based approach was used to assemble the reads, and differential gene expression analysis (DGE) was carried out using CLC Genomics Workbench v9.5.3 (clcbio, Denmark). The empirical analysis of differential gene expression was done using a Negative Binomial distribution using GLM model. The *p*-values were adjusted using the Benjamini & Hochberg method. False discovery rate (FDR) corrected *p*-value of 0.005 and $\log_2$ (fold change) of two was used as the threshold to generate the significantly differentially expressed gene list. The colistin sensitive strain 116S that belonged to ST383was used as the reference for RNA sequencing analysis (Figure 3.2.3).

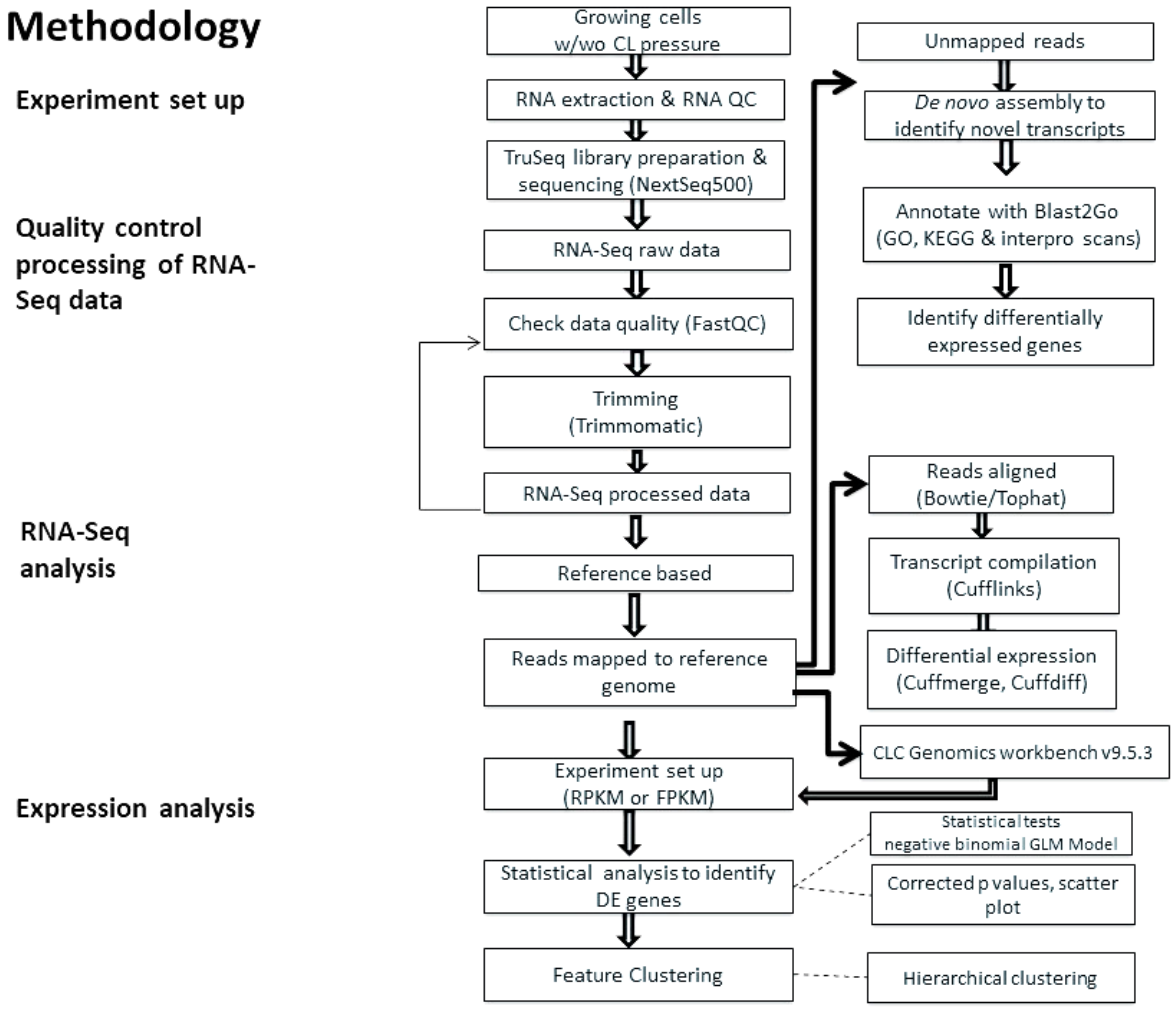


Figure 3.2.3 Overview of transcriptomic analysis

3.2.2.9 Validation of differentially expressed genes (DEGs) using qPCR

From each comparison, top 3 differentially expressed genes (DEGs) were selected and further validated by reverse transcription polymerase chain reaction (RT-PCR) performed in duplicate (StepOne Plus, Life Technologies, Merelbeke, Belgium). After normalisation against the calibrator, transcript levels were expressed as n-fold differences relative to levels obtained with sensitive vs resistant and antibiotic vs absence of antibiotic, respectively.

3.2.2.10 Molecular phylogenetic analysis

The core genome-based phylogenetic analysis was done utilising sequences representing phylogroup of *K. pneumoniae* as described (24). *K. pneumoniae* NJST258-2 (KPI) (accession no: CP006918), *K. quasipneumoniae* (KpII) (accession no: CP014696), *K. variicola* (KpIII) (accession no: NZ_MPC101000001), HS11286 (colistin sensitive reference strain) (accession no: NC_016845) and 109 study strains were used for this core genome analysis. The core genome SNPs were extracted using ParSNP (25). Phylogenetic reconstruction was carried out using MEGA7 (26). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbour-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach and then selecting the topology with superior log-likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 200.0000)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site (Figure 3.2.12).

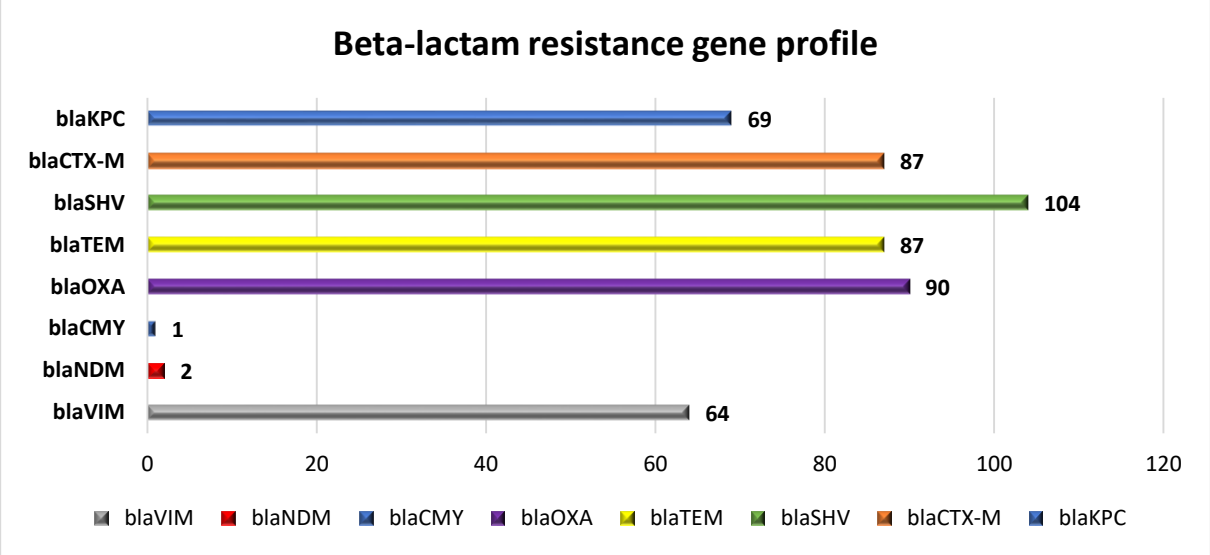
3.2.2.11 Pathway Analysis

The transcriptomics data from the three different phenotypes analysed (colistin-sensitive, -resistant, and-induced) was utilised to build a metabolic pathway. For this analysis, we had selected DEGs (FDR $p<0.005$) from every comparison. DEGs were translated to predicted proteins, and Blast2Go tool was used to predict the functional annotation (27) and EC (Enzyme Commission) numbers of the proteins. The EC numbers were submitted to the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway program (http://www.genome.jp/kegg/), and the metabolic pathway was determined. Protein network analysis was done using the STRING database v10.1 (https://string-db.org/).

3.2.3 Results

3.2.3.1 Strain Characteristics

The sequence type (ST) diversity of colistin-resistant strains was specific to the geographical region with high-level diversity identified in strains from Belgium. Out of the three Belgian hospitals, two showed high-level ST diversity (ST709 (n=2), ST101 (n=1), ST11 (n=3), ST15 (n=3), ST17 (n=2), ST323 (n=1) and ST405 (n=4)) and one hospital showed a predominant single ST (ST512 (n=11)). Hospitals in Italy showed a dominance of ST101 (n=13), and ST15 (n=1), ST307 (n=2), ST327 (n=1) and ST409 (n=1). Following were the STs found in Spain [ST147 (n=17), ST15 (n=1), ST405 (n=2) and ST437 (n=8)]; Greece [ST147 (n=2), ST258 (n=2) and ST383 (n=13)] and India [ST14 (n=1), ST2096 (n=4), ST231 (n=2) and ST307 (n=1)]. Interestingly, heterogeneous population among colistin-resistant strains from Belgium were identified with eight different STs and four novel STs compared to other countries such as Italy, Greece, Spain and India. Prevalence of beta-lactam genes among studied strains showed a high prevalence of *blaSHV* followed by *blaOXA*, *blaCTX-M*, *blaTEM* and *blaKPC* (Table 3.2.1, Figure 3.2.4 A & B and 3.2.5).



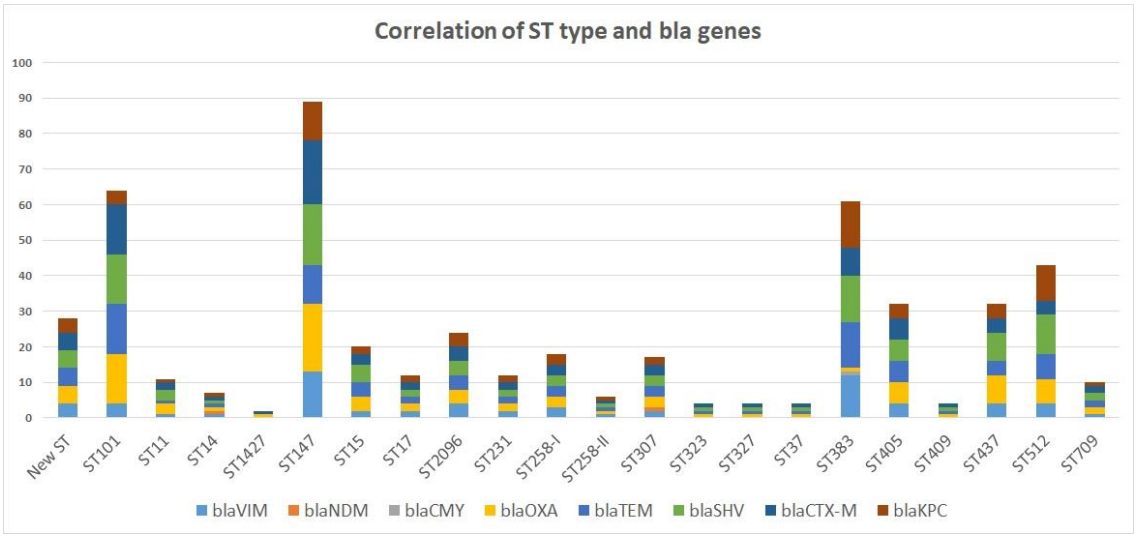


Figure 3.2.4 Overall prevalence of beta-lactam resistance genes among studied strains (A). The correlation between sequence type and beta-lactam genes among colistin-resistant *K. pneumonia* (B)

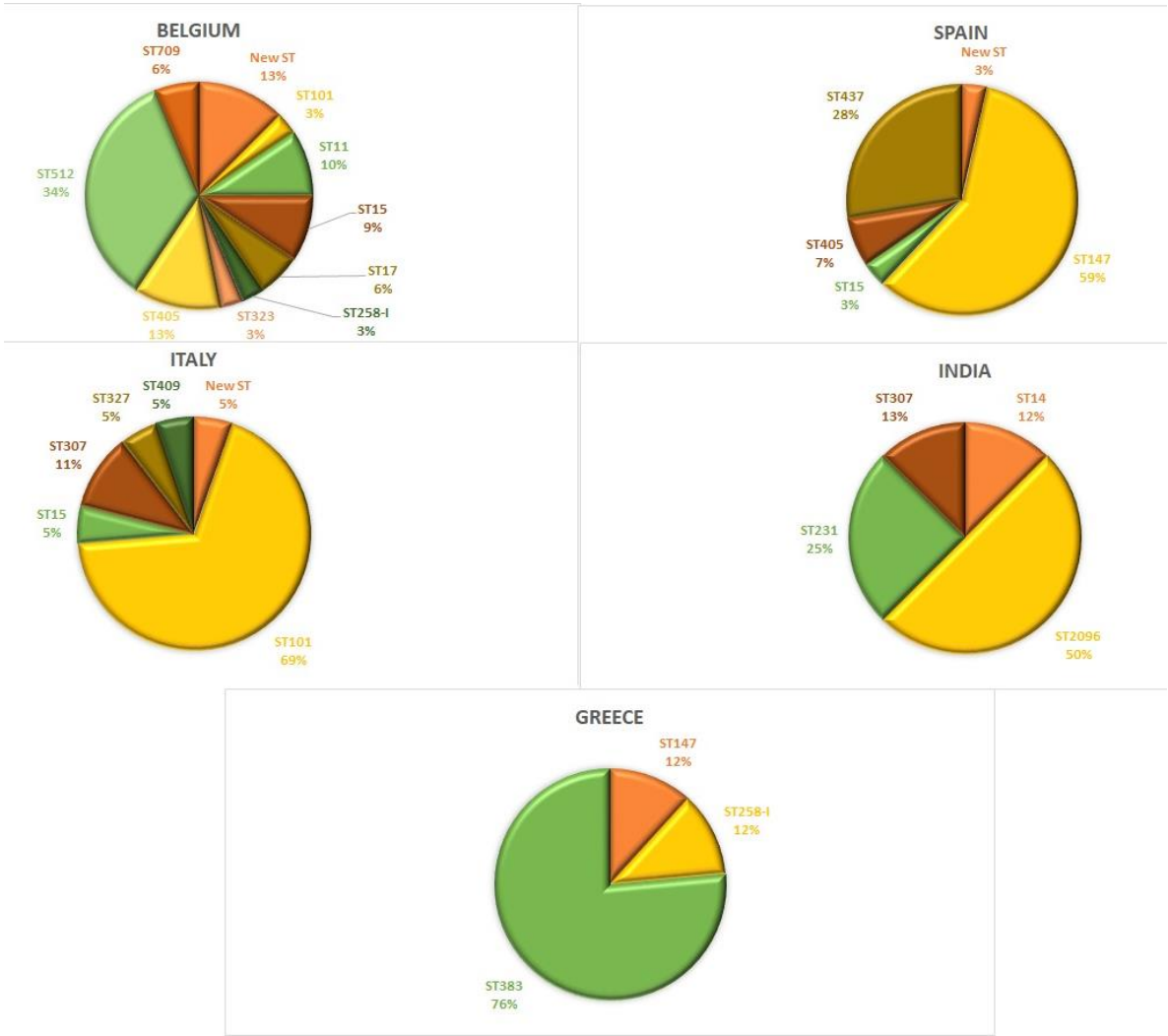


Figure 3.2.5 Genetic diversity of colistin-resistant *K. pneumoniae* in different countries

Table 3.2.1 Strain and clinical information of *K. pneumoniae* isolates used in this study

Strain ID	Site	Country	Project	DOI	MIC		Colistin resistance gene		ST type	Plasmids	Molecular typing							
							mgrB modifications	Others			*bla*-genes							
											bla VIM	*bla* KPC	*bla* CTX-M	*bla* SHV	*bla* TEM	*bla* OXA	*bla* CMY	*bla* NDM
1240	Blood	Greece	BOFCOL	Jan-10	R	64	IS1R interruption at promoter region		ST383	IncFII(K), IncA/C2, ColRNAI	VIM-	KPC-		SHV-1	TEM-			
1251	Wound	Greece	BOFCOL	Feb-10	S	0.5	Sensitive		ST383	IncFIB (pQil), IncFII(K), IncA/C2,	VIM-	KPC-	-	SHV-1	TEM-			
1262	Urine	Greece	BOFCOL	Apr-10	S	0.5	Sensitive		ST383	IncFII(K), IncA/C2,	VIM-	-	CTX-M-	SHV-1	-			
1264	Urine	Greece	BOFCOL	May-10	S	0.5	Sensitive		ST383	IncFII(K), IncA/C2,	VIM-	KPC-	-	SHV-1	TEM-			
1238	Blood	Greece	BOFCOL	May-10	R	64	del 48th nt C premature stopcodon		ST383	IncFIB (pQil), IncFII(K), IncA/C2,	VIM-	KPC-	CTX-M-	SHV-1	TEM-			
1274	CVC	Greece	BOFCOL	Aug-10	S	0.25	Sensitive		ST383	IncFIB (pQil), IncA/C2,	VIM-	KPC-		SHV-1	-			
1282	Urine	Greece	BOFCOL	Nov-10	S	0.5	Sensitive		ST383	IncFIB (pQil), IncFII(K), IncA/C2,	VIM-	-		SHV-1	-			
1292	Blood	Greece	BOFCOL	Aug-11	R	64	del 48th nt C premature stopcodon		ST383	IncFIB (pQil), IncFII(K), IncA/C2,	VIM-	KPC-	CTX-M-	SHV-1	TEM-			
287	Wound	Greece	BOFCOL	Mar-13	R	32	IS1R interruption at promoter region		ST383	IncFIB (pQil), IncFII(K), IncA/C2,	VIM-	KPC-	-	SHV-1	-			
110	Blood	Greece	BOFCOL	Jan-10	R	32	IS1R interruption at promoter region		ST383	IncFIB(pQil),IncA/C2	VIM-	KPC-		SHV-1	TEM-	OXA-9	CMY-4	
116	CVC	Greece	BOFCOL	Feb-10	S	2	Sensitive		ST383	IncFIB (pQil), IncFII(K), IncA/C2,		KPC-		SHV-1	TEM-			
117	Blood	Greece	BOFCOL	Apr-10	R	32	IS1R interruption at promoter region		ST383	IncFIB (pQil), IncFII(K), IncA/C2,	VIM-	KPC-		SHV-1	TEM-			
30S	Urine	Greece	BOFCOL		S	2	Sensitive		ST147	IncFIB(K),IncR,IncFIB(Mar),IncH1B,IncFIB(pKPHS1),IncFII	VIM-			SHV-11				
31R	Urine	Greece	BOFCOL		R	32		pmrB	ST147	IncFII(K),IncFIB(K),IncR,IncFIB(Mar),IncH1B,IncFIB(pKPHS1),IncFI	VIM-			SHV-11				
3712	Bronchi	Greece	BOFCOL	Sep-09	R	32	ND		ST258	IncFIB(K),IncFIB(pQil),ColRNAI,IncX3,IncFII(pRSB107)		KPC-		SHV-	TEM-	OXA-9		
3789	Bronchi	Greece	BOFCOL	Oct-09	S	2	Sensitive		ST258	IncFIB(K),IncFIB(pQil),ColRNAI,IncX3,IncFII(pRSB107)		KPC-		SHV-	TEM-	OXA-9		
3359R	Urine	Greece	BOFCOL	May-09	R	32	IS elment intrruption		ST258-I	IncFIB(K),IncFIB(pQil),ColRNAI,IncX3		KPC-		SHV-	TEM-	OXA-9		
3319S	Blood	Greece	BOFCOL	Apr-09	S	1	Sensitive		ST258-I	IncFII(K),IncFIB(K),IncFIB(pQil),ColRNAI,IncX3		KPC-		SHV-12	TEM-	OXA-9		
313	Blood	Greece	BOFCOL	Apr-13	R	32	delta mgrB		ST383	IncFIB (pQil), IncFII(K), IncA/C2,	VIM-	KPC-	CTX-M-	SHV-1	TEM-			
3319-4R	*In vitro*	Belgium	BOFCOL	Apr-12	R	64	IS elment intrruption		ST258-I	IncFIB(K),IncFIB(pQil),ColRNAI,IncX3,IncFII(pRSB107)		KPC-		SHV-	TEM-	OXA-9		
3319-10R	*In vitro*	Belgium	BOFCOL	Apr-12	R	64	IS elment intrruption		ST258-I	IncFIB(K),IncFIB(pQil),ColRNAI,IncX3,IncFII(pRSB107)		KPC-		SHV-	TEM-	OXA-9		
ANR1308	Rectum	Belgium	RGNOSIS	Aug-14	R	32	C to A at 117th premature stop codon (C44*)		ST709	IncFII(K)			CTX-M-	SHV-26	TEM-	OXA-1		
ANR3090	Rectum	Belgium	RGNOSIS	Aug-14	R	128	insertion of ISKpn26 (IS5 family) at 75th position		ST11	IncFIB(K),IncR,IncFII(K)			CTX-M-	SHV-11		OXA-1		
ANR3090	Rectum	Belgium	RGNOSIS	Aug-14	R	32	WT		ST323	IncFIB(K)			CTX-M-	SHV-99	TEM-	OXA-1		
LBR0016	Aspirat	Belgium	RGNOSIS	Sep-14	R	256	TA insertion at 124th position led to extension		ST512	IncFII(K),IncFIB(pQil),ColRNAI,IncFIB(K)		KPC-		SHV-11				
LBR0909	Aspirat	Belgium	RGNOSIS	Feb-15	R	128	TA insertion at 124th position led to extension		ST512	IncFII(K),IncFIB(pQil),ColRNAI,IncFIB(K)		KPC-		SHV-11				
LBR1161	Aspirat	Belgium	RGNOSIS	Mar-15	R	>128	TA insertion at 124th position led to extension		ST512	IncFII(K), IncFIB(pQil), IncFIB(K),ColRNAI		KPC-		SHV-11	TEM-	OXA-9		
LBR1174	Aspirat	Belgium	RGNOSIS	Apr-15	R	32	TA insertion at 124th position led to extension		ST512	IncFII(K),IncFIB(pQil), ColRNAI, IncFIB(K)		KPC-		SHV-11				
LBR1945	Aspirat	Belgium	RGNOSIS	Aug-15	R	64	WT		ST512	ColRNAI, IncFIB(K), IncFII(K)				SHV-11				
LBR7062	Aspirat	Belgium	RGNOSIS	Jan-15	R	64	IS eletment at 48th position (IS3 like (ISEc16) C to A		New-ST	IncFIB(K),Col156								
ANR1308	Rectum	Belgium	RGNOSIS	Aug-14	R	32	IS eletment insertion at 48th position (IS3 like		ST709	IncFII(K)			CTX-M-	SHV-26	TEM-	OXA-1		
ANR4736	Rectum	Belgium	RGNOSIS	Dec-15	R	64	13 nt segment deletion led elongation of the gene		ST11	IncFII(K),IncFIB(K),ColRNAI,IncL/M(pOXA-48),IncR,Col(BS512)				SHV-11		OXA-1,OXA-48		
LBR2393E	Rectum	Belgium	RGNOSIS	Dec-15	R	16	TA insertion at 124th position led to extension		New ST	IncFII(K),IncFII(pRSB107)			CTX-M-			OXA-1		
LBR2648E	Aspirat	Belgium	RGNOSIS	Nov-15	R	128	Interruption of IS1R (IS1 family) at 108th position		ST17	IncFII(K),IncFIA(HI1),IncR				SHV-11	TEM-			
LBR2648	Aspirat	Belgium	RGNOSIS	Nov-15	R	32	WT		ST17	IncFII(K),IncFIA(HI1),IncR				SHV-12	TEM-			
LBR2674	Aspirat	Belgium	RGNOSIS	Nov-15	R	32	TA insertion at 124th position led to extension		ST512	IncFII(K),IncFIB(K),IncFIB(pQil)		KPC-		SHV-11	TEM-	OXA-9		
LBR2708	Aspirat	Belgium	RGNOSIS	Jan-16	R	128	WT		ST512	IncFIB(pQil),ColRNAI		KPC-		SHV-11	TEM-	OXA-9		
LBR2993	Aspirat	Belgium	RGNOSIS	Jan-16	R	128	WT		ST512	IncFII(K),IncFIB(K),IncFIB(pQil)		KPC-		SHV-11	TEM-	OXA-9		
LBR3880E	Rectum	Belgium	RGNOSIS		R	32	Interruption of IS1R (IS1 family) at 108th position		ST405	IncFII(K),IncFIB(K)			CTX-M-	SHV-76	TEM-	OXA-1		
LBR4210E	Rectum	Belgium	RGNOSIS		R	64	WT		ST405	IncFII(K),IncFIB(K),IncFIB(Mar)			CTX-M-	SHV-76	TEM-			
LBR4435	Aspirat	Belgium	RGNOSIS		R	32	Interruption of IS1R (IS1 family) at 108th position		ST512	IncFII(K),IncFIB(K),IncFIB(pQil)		KPC-		SHV-11	TEM-	OXA-9		
LBR5420	Rectum	Belgium	RGNOSIS		R	32	Interruption of IS1R (IS1 family) at 108th position		New ST	IncFIA,IncFIB(AP001918),Col156,IncFII			CTX-M-			OXA-1		
LBR706E2	Aspirat	Belgium	RGNOSIS	Jan-15	R	32	WT		New ST	IncFII(K),IncFIB(K),Col156								
ANR5890	Rectum	Belgium	RGNOSIS	Mar-16	R	64	with IS1X2 (IS1) interruption in promoter		ST405	IncFIB(K),IncR			CTX-M-	SHV-76	TEM-	OXA-1		
LBR4766E	Rectum	Belgium	RGNOSIS	Jan-17	R	64	WT		ST15	IncFIA(HI1), IncR, ColpVC			CTX-M-	SHV-28	TEM-	OXA-1		
LBR4817E	Rectum	Belgium	RGNOSIS	Nov-16	R	64	TA insertion at 124th nt led to extension		ST405	IncFIB (K), IncFIB, IncFII(K)			CTX-M-	SHV-76	TEM-	OXA-1		
LBR4766	Rectum	Belgium	RGNOSIS	Dec-16	R	64	WT		ST15	ColpVC				SHV-28				
LBR4766	Aspirat	Belgium	RGNOSIS	Jan-16	R	64	WT		ST15	ColpVC, IncR, IncFIA (HI1)			CTX-M-	SHV-28	TEM-	OXA-1		
LBR5569	Rectum	Belgium	RGNOSIS	Mar-17	R	64	WT		ST512	IncFII, IncFIB< ColRNAI		KPC-		SHV-11	TEM-	OXA-9		
UZR5497	Anal	Belgium	RGNOSIS	Aug-15	R	>256	C to A at 117th premature stop codon (C44*)		ST11	IncL/M, IncFIB, IncFII, ColRNAI, IncR				SHV-11		OXA-1, OXA-48		
LBR5372	Aspirat	Belgium	RGNOSIS	Feb-17	R	64	WT		ST512	IncFII, ColpVC, IncFIB, ColRNAI		KPC-		SHV-11	TEM-	OXA-9		
IBIVAP4	ETTUBE	Belgium	IBIVAP		R	128	pmrB mutation		ST101	IncFII(K),ColRNAI,IncFIA(HI1),IncR,Col(BS512)			CTX-M-	SHV-1	TEM-	OXA-1,OXA-9		
ITR0132C	Rectum	Italy	RGNOSIS	Oct-14	R	64	WT		ST101	IncL/M(pOXA-48),IncR,IncFIA(HI1),IncFIB(K),IncFII(K)			CTX-M-	SHV-1	TEM-	OXA-1, OXA-9,		

Table 3.2.1 (continued)

ITR0137C	Aspirat	Italy	RGNOSI	Oct-14	R	32	WT		ST101	IncL/M(pOXA-48),IncR,IncFIA(HI1),IncFIB(K),IncFII(K)			CTX-M-	SHV-1	TEM-	OXA-1, OXA-9,		
ITR0169C	Rectum	Italy	RGNOSI	Nov-14	R	64	13 nt segment deletion led elongation of the gene		ST101	IncL/M,IncR,IncFIA(HI1),IncFIB(K),IncFII(K)			CTX-M-	SHV-1	TEM-	OXA-1, OXA-9,		
ITR0176C	Aspitrat	Italy	RGNOSI	Dec-14	R	64	with IS1X2 (IS1) interruption in promoter		ST101	IncR,IncL/M(pOXA-48),IncFIA(H1)			CTX-M-	SHV-1	TEM-	OXA-1, OXA-9,		
ITR0195C	Aspirat	Italy	RGNOSI	Dec-14	R	32	WT		ST307	IncL/M(pOXA-48), IncFII(K), IncFIB(K)			CTX-M-	SHV-28	TEM-	OXA-1, OXA-48		
ITR0205C	Rectum	Italy	RGNOSI	Jan-15	R	32	with IS1X2 (IS1) interruption in promoter		ST101	IncR, IncFIA(H1), IncL/M(pOXA-48)			CTX-M-	SHV-1	TEM-	OXA-1, OXA-9,		
ITR0307C	Rectum	Italy	RGNOSI	May-15	R	128	Interruption of IS1R (IS1 family) at 108th position		ST409	IncFIB(K),IncFII(K),IncL/M(pOXA-48)			CTX-M-	SHV-27	TEM-	OXA-1, OXA-48		
ITR0321C	Rectum	Italy	RGNOSI	May-15	R	64	WT		ST327	IncL/M(pOXA-48), IncFII(K), IncFIB(K)			CTX-M-	SHV-28	TEM-	OXA-1, OXA-48		
ITR0374C	Aspirat	Italy	RGNOSI	Aug-15	R	128	WT		ST101	IncFIA(HI1), IncR,IncL/M(pOXA-48),IncFII(K),IncFIB(K)			CTX-M-	SHV-1	TEM-	OXA-9, OXA-48		
ITR0385C	Aspirat	Italy	RGNOSI	Sep-15	R	64	mutation at 7th base led to stop codon		New-ST	IncFIB(K), IncL/M(pOXA-48), IncR, IncFIA(HI1), IncFII(K)			CTX-M-	SHV-1	TEM-	OXA-1, OXA-48		
ITR0416	Rectum	Italy	RGNOSI	Aug-15	R	64	mutation at 7th baseT to A and also C to T at		ST101	IncFIB(K), IncFIA(H1), IncR, IncL/M(pOXA-48), IncFII(K)			CTX-M-	SHV-1	TEM-	OXA-1, OXA-48		
ITR0439C	Rectum	Italy	RGNOSI	Sep-15	R	256	C to T at 88 th premature stop codon (tQ33*		ST101	IncFII(K),IncFIA(HI1),IncFIB(K),IncR,IncL/M(pOXA-48)			CTX-M-	SHV-1	TEM-	OXA-9, OXA-48		
ITR0441C	Rectum	Italy	RGNOSI	Sep-15	R	128	WT		ST101	IncFII(K),IncFIA(HI1),IncFIB(K),IncR,IncL/M(pOXA-48)			CTX-M-	SHV-1	TEM-	OXA-9, OXA-48		
ITR0452C	Blood	Italy	RGNOSI	Sep-15	R	128	WT		ST101	IncFIB(K),IncL/M(pOXA-48),IncFII(K),IncR,IncFIA(H1),			CTX-M-	SHV-1	TEM-	OXA-1, OXA-48		
BCR0799	Aspirat	Spain	RGNOSI	Sep-16	R	8	Insertion IS1 at 51nt mgrB		ST405	IncFII(K),IncFIB(K),IncL/M(pOXA-48),IncFIB(Mar)			CTX-M-	SHV-76	TEM-	OXA-1,OXA-48		
FE0680CP	Rectum	Spain	RGNOSI	Jul-15	R	64	Interruption of IS1X2 at 51nt mgrB		ST437	IncI1,IncFIB(K),IncL/M(pOXA-48),IncFIB(Mar),IncH1B				SHV-12		OXA-1,OXA-48		
FE0783EC	Rectum	Spain	RGNOSI	Aug-15	R	32	no modfidications		ST437	IncI1,IncFIB(K),IncL/M(pOXA-				SHV-12		OXA-1		
FE1004CP	Rectum	Spain	RGNOSI	Nov-15	R	64	no modfidications		ST437	IncI1,IncFII(K),IncFIB(K),IncR,IncHI2,IncFIB(Mar),IncH1B				SHV-12		OXA-1		
FE1054CP	Rectum	Spain	RGNOSI	Dec-15	R	64	no modfidications		ST437	IncI1,IncFII(K),IncFIB(K),IncL/M(pOXA-				SHV-12		OXA-1,OXA-48		
FE1353S	Rectum	Spain	RGNOSI	Apr-16	R	32	no modfidications		ST147	IncFII(K),ColRNAI,IncHI2,IncHI2,IncFIB(Mar)			CTX-M-	SHV-12				
FE1592S	Rectum	Spain	RGNOSI	Aug-16	R	64	no modfidications		ST147	IncFII(K),IncFIB(K),ColRNAI,IncHI2,IncHI2,IncFIB(Mar)			CTX-M-	SHV-12		OXA-1		
FE1599S	Rectum	Spain	RGNOSI	Aug-16	R	32	no modfidications		ST147	IncFII(K),IncFIB(K),IncHI2,IncHI2,IncFIB(Mar)			CTX-M-	SHV-12		OXA-1		
FE1678	Rectum	Spain	RGNOSI	Sep-16	R	32	no modfidications		ST147	IncFII(K),IncFIB(K),ColRNAI,IncHI2,IncHI2,IncFIB(Mar)			CTX-M-	SHV-12		OXA-1		
FE1730	Rectum	Spain	RGNOSI	Oct-16	R	128	no modfidications		ST147	IncFIB(K),IncL/M(pOXA-48),IncHI2,IncHI2,IncFIB(Mar)			CTX-M-	SHV-12		OXA-1,OXA-48		
FE1827	Aspirat	Spain	RGNOSI	Dec-16	R	16	no modfidications		ST147	IncFII(K),IncFIB(K),IncL/M(pOXA-			CTX-M-	SHV-12		OXA-1,OXA-48		
FE1856	Aspirat	Spain	RGNOSI	Dec-16	R	64	no modfidications		ST147	IncFIB(K),IncL/M(pOXA-48),IncHI2,IncHI2,IncFIB(Mar)			CTX-M-	SHV-12		OXA-1,OXA-48		
FE1921	Rectum	Spain	RGNOSI	Jan-17	R	64	no modfidications		ST147	IncFIB(K),IncL/M(pOXA-48),IncHI2,IncHI2,IncFIB(Mar)			CTX-M-	SHV-11		OXA-1,OXA-48		
FE2054	Aspirat	Spain	RGNOSI	Mar-17	R	8	no modfidications		ST147	IncFII(K),IncFIB(K),IncL/M(pOXA-48),IncHI2,IncHI2,IncFIB(Mar)			CTX-M-	SHV-12		OXA-1,OXA-48		
IS207	NA	Israel	SATUR	NA	R	32	no modification		ST258-II	IncFII(K),IncFIB(K),IncFIB(pQil),ColRNAI		KPC-		SHV-11	TEM-	OXA-9		
ITR0450C	Throat	Italy	RGNOSI	Nov-15	R	32	WT		ST307	IncFII(K),IncFIB(K),ColRNAI,IncL/M(pOXA-48)			CTX-M-	SHV-28	TEM-	OXA-1,OXA-48		
ITR0477C	Rectum	Italy	RGNOSI	Oct-15	R	32	mutation at 7th base premature		ST101	IncFII(K),IncL/M(pOXA-48),IncFIA(HI1),IncR,IncFII(pRSB107)				SHV-1		OXA-48		
ITR0661C	Rectum	Italy	RGNOSI	May-16	R	64	WT		ST101	IncFII(K),IncFIB(K),IncL/M(pOXA-48),IncFIA(HI1),IncR				SHV-1		OXA-48		
ITR0680C	Aspirat	Italy	RGNOSI	Jun-16	R	64	WT		ST15	ColRNAI,ColpVC				SHV-28				
ITR0680C	Throat	Italy	RGNOSI	Jul-16	R	16	WT		ST101	IncFII(K),IncFIB(K),ColRNAI,IncR	VIM-			SHV-1				
PS2445	Rectum	Spain	RGNOSI	Jan-16	R	32	WT		New-ST	IncFII(K),IncFIB(K),Col156			CTX-M-	SHV-1	TEM-	OXA-1		
PS2965EC	Rectum	Spain	RGNOSI	Jun-16	R	16	WT		ST15	IncFII(K),IncFIB(K),IncR,IncN				SHV-28	TEM-	OXA-1		
FE0883CP	Rectum	Spain	RGNOSI	Oct-15	R	64	WT		ST437	IncFIB, IncFII, IncL/M, IncHI1B, IncI1, IncR,				SHV-5		OXA-48, OXA-1		
FE0908CP	Rectum	Spain	RGNOSI	Oct-15	R	32	WT		ST437	IncFIB, IncFII, IncL/M, IncHI1B, IncI1, IncR, Col				SHV-12		OXA-48, OXA-1		
FE0991CP	Rectum	Spain	RGNOSI	Nov-15	R	64	WT		ST437	IncFIB, IncFII, IncL/M, IncHI1B, IncI1, IncR,				SHV-11		OXA-48, OXA-1		
FE1147	Rectum	Spain	RGNOSI	Jan-16	R	128	insertion of ISKpn26 (IS5 family) at 75th position		ST147	IncHI2A, IncFIB, IncHI2, IncFII			CTX-M-	SHV-12		OXA-1		
FE1221	Blood	Spain	RGNOSI	Feb-16	R	256	WT		ST147	IncFIB, IncHI2A, IncFIB, IncHI2, IncFII			CTX-M-	SHV-12		OXA-1		
FE1279	Rectum	Spain	RGNOSI	Mar-16	R	32	WT		ST147	IncFIB, IncHI2A, IncFIB, IncHI2, IncFII (K) IncL/M			CTX-M-			OXA-48, OXA-1		
FE1442	Rectum	Spain	RGNOSI	Jun-16	R	128	WT		ST437	IncFIB, IncL/M, IncHI1B, IncI1				SHV-12		OXA-48		
FE1454	Rectum	Spain	RGNOSI	Jun-16	R	128	WT		ST147	IncFIB, IncHI2A, IncFIB, IncHI2, IncL/M, IncHIB	VIM-			SHV-1		OXA-1		
FE1305	Rectum	Spain	RGNOSI	Mar-16	R	64	WT		ST147	IncFIB, IncHI2A,IncHI2, IncFII (K)			CTX-M-	SHV-12		OXA-1		
FE1649	Rectum	Spain	RGNOSI	Aug-16	R	16	WT		ST147	IncFIB, IncHI2A, IncFIB (K), IncHI2, IncL/M			CTX-M-	SHV-12		OXA-48, OXA-1		
FE1674	Rectum	Spain	RGNOSI	Aug-16	R	32	WT		ST147	IncFIB, IncHI2A, IncFIB, IncHI2, IncFII (K) IncL/M			CTX-M-			OXA-48, OXA-1		
FE1856	Rectum	Spain	RGNOSI	Dec-16	R	64	WT		ST147	IncFII(K),IncFIB(K),IncL/M(pOXA-48),IncHI2,IncHI2,IncFIB(Mar)			CTX-M-	SHV-12		OXA-1, OXA-48		
BCR0801	Rectum	Spain	RGNOSI	Sep-16	R	32	WT		ST405	IncFIB, IncFII (K), IncL/M, IncFIB(K)			CTX-M-	SHV-76	TEM-	OXA-48, OXA-1		
SIN1	Blood	India	INKPN	May-15	R	>4	2nd base mutation T>C premature stop codon		ST231	IncFII(K),IncFIB(pQil),IncFIA,ColKP3			CTX-M-	SHV-1	TEM-	OXA-232		
SIN2	wound	India	INKPN	Jul-15	R	>4	IS905B interruption in promoter		ST231	IncFIB(pQil),IncFIA,IncFII(pRSB107)			CTX-M-	SHV-1	TEM-			
SIN3	wound	India	INKPN	May-15	R	>4	WT		ST2096	IncFII(K),IncFIB(K),IncFIB(Mar),IncH1B,ColKP3			CTX-M-	SHV-28	TEM-	OXA-1,OXA-232		
SIN4	wound	India	INKPN	Nov-15	R	>4	WT		ST14	IncFII(K),IncFIB(K),IncFIB(Mar),IncH1B,ColKP3			CTX-M-	SHV-28		OXA-1		NDM-
SIN5	Blood	India	INKPN	Jan-16	R	>4	PmrB mutation		ST2096	IncFII(K),IncFIB(K),IncFIB(Mar),IncH1B,ColKP3			CTX-M-	SHV-28	TEM-	OXA-1,OXA-232		
SIN6	Blood	India	INKPN	Jan-16	R	>4	WT		ST2096	IncFII(K),IncFIB(K),ColRNAI,IncFIB(Mar),IncH1B,ColKP3			CTX-M-	SHV-28	TEM-	OXA-1,OXA-232		
SIN7	Wound	India	INKPN	Sep-16	R	>4	IS element interruption at promoter		ST307	IncFIB(K),IncHI2,IncHI2,IncFIB(Mar),IncH1B			CTX-M-		TEM-			NDM-
SIN8	Pelvic	India	INKPN	Nov-16	R	>4	70th base C deletion premature stop codon		ST2096	IncFII(K),IncFIB(K),IncFIB(Mar),IncH1B,ColKP3			CTX-M-	SHV-28	TEM-	OXA-1,OXA-232		
UNHP17	Urine	Vietna	VLIR-	Jan-15	R	4	no modification		ST37	IncFII, IncH1B			CTX-M-	SHV-2	TEM-	OXA-9		
UNHP62	Urine	Vietna	VLIR-	Mar-15	R	4	no modification		ST1427	IncFIA(H1), IncFII, IncFIB(K)			CTX-M-			OXA-1		

3.2.3.2 Assembly validation by WGM

WGMs of selected strains (3319S, 3359R and 117) and *in vitro* (3319-10R) strain were compared to the scaffolds of the respective strains. The misassemblies of the scaffolds were identified and corrected. The scaffolds were ordered against the WGMs and pseudochromosomes (data not shown) were generated.

3.2.3.3 *In-vitro* generation of colistin-resistant clones and stability of colistin resistance in *K. pneumoniae*

Colistin-resistant *in vitro* clones with MICs ranging from 64 to 2048 µg/L were generated in three serial passages from two clinical colistin sensitive strains (3319S and 3789S) (Table 3.2.7). High level resistant *in vitro* clones were tested for the resistance stability and it showed the stable persistence of resistance for more than ~6500 generations (Table 3.2.2).

Strain	MIC	MIC (post 140*)	MIC (post 6210*)	MIC (post 6094*)
3319-2R	64	64	64	
3319-4R	32	>64	64	
3319-8R	64	64	64	
3319-10R	64	64	64	
3789-1R	2048	2048	2048	512
3789-2R	2048	2048	2048	512
3789-4R	2048	2048	2048	512
3789-9R	>2048	2048	2048	512

* number of generations

Table 3.2.2 Stability experiment for assessing CL resistance of *in vitro* generated strains.

3.2.3.4 Fitness study by a competition assay

In this experiment, we estimated the biological cost of colistin resistance using an isogenic clinical pair and the *in vitro* generated resistant clones from the same sensitive strain (3319S). Interestingly, the two resistant strains showed different *mgrB* modifications. The *in vivo* resistant strain 3359R showed interruption by IS element in the coding sequence of *mgrB*, and *in vitro* strain, 3319-10R had an interruption at the promoter region. Of note, the relative fitness of the clinical colistin-resistant strain (3359-R) was significantly higher than the colistin sensitive strain (3319S) (1.87 vs 0.54, respective, P = 0.004). While the fitness of 3319S compared was on average 7% higher than the *in vitro* generated colistin-resistant strains (3319-2R, 4R, 8R and 10R) (Table 3.2.3).

Table 3.2.3 The relative fitness of *in vivo* and *in vitro* resistant vs sensitive counterpart pairs were compared.

MIX	Strains	CFU/mL (mix before 6h incubation)	CFU/mL (after 6h incubation)	log 10 (time 0)	log 10 (time end)	log 10(2)	g = log10 (end of mix)-log10 (time0)/log10(2)	relative fitness	strain characteristics
1	3319-S	7.0E+06	1.2E+08	6.8	8.1	0.3	4.1	0.56	In-vivo strains 3319S & 3359R (S > R)
	3359-R	7.6E+06	1.2E+09	6.9	9.1	0.3	7.3	1.78	
2	3319-S	4.8E+06	9.6E+08	6.7	9.0	0.3	7.6	1.12	In-vitro strains (MIC 64) with 3319-S
	3319-2R	7.8E+06	8.8E+08	6.9	8.9	0.3	6.8	0.89	
3	3319-S	4.4E+06	2.1E+09	6.6	9.3	0.3	8.9	1.16	In-vitro strains (MIC 64) with 3319-S
	3319-4R	7.8E+06	1.6E+09	6.9	9.2	0.3	7.7	0.86	
4	3319-S	9.6E+06	2.9E+09	7.0	9.5	0.3	8.2	1.02	In-vitro strains (MIC 64) with 3319-S
	3319-8R	6.0E+06	1.6E+09	6.8	9.2	0.3	8.1	0.98	
5	3319-S	7.6E+06	2.7E+09	6.9	9.4	0.3	8.5	1.07	In-vitro strains (MIC 64) with 3319-S
	3319-10R	7.2E+06	1.7E+09	6.9	9.2	0.3	7.9	0.93	

3.2.3.5 Comparative genome analysis of colistin-resistant determinants

First, we compared genome sequences of colistin sensitive strain (3319S) and *in vitro* generated resistant strain (3319-10R) to identify global changes in the chromosome under colistin pressure. We identified 27 non-synonymous SNPs in 27 protein-coding genes. (Table 3.2.4) Among these, the putative outer membrane lipoprotein (*yaiW:* surface-exposed outer membrane lipoprotein) has been shown to play a role in helping the movement of cationic peptides across the outer membrane (28). Mutation in *yaiW* led to predicted amino acid change from a hydrophobic valine79 to a non-hydrophobic alanine in the helical region. Also, we found mutations in the ethanolamine ammonia lyase heavy chain (*eutB*), known to be involved in cellular amino acid metabolic process and ethanolamine to acetaldehyde and ammonia.

Table 3.2.4 Genes with single nucleotide polymorphisms (SNPs) between 3319S vs *in vitro* strain 3319-10R

3319S	3319-10R	Gene	Amino acid change
A	T	Xanthine-uracil permease	Xanthine-uracil permease:p.Thr239Ser
G	A	UPF0229 protein YeaH	UPF0229 protein YeaH:p.Ile119Val
G	T	Uncharacterized protein YphG_TPR-domain containing	Uncharacterized protein YphG- TPR-domain containing:p.Val682Gly
G	T	Transcriptional regulator-LysR	Transcriptional regulator-LysR family:p.Val276Gly
C	A	Transcription-repair coupling factor	Transcription-repair coupling factor:p.His732Pro
C	T	Signal peptidase I (EC 3.4.21.89)	Signal peptidase I (EC 3.4.21.89):p.Val294Ala
A	C	Shikimate 5-dehydrogenase I gamma (EC 1.1.1.25)	Shikimate 5-dehydrogenase I gamma (EC 1.1.1.25):p.Asn194Lys
A	T	Regulatory protein CII	Regulatory protein CII:p.Glu102Val
G	A	Pyruvate decarboxylase (EC 4.1.1.1); Alpha-keto-acid decarboxylase (EC 4.1.1.-)	Alpha-keto-acid decarboxylase (EC 4.1.1.-):p.Glu516Gly
C	T	putative oxidoreductase protein	Putative oxidoreductase protein:p.Ser138Pro
G	A	Putative outer membrane lipoprotein yaiW	Putative outer membrane lipoprotein-yaiW:p.Val79Ala
C	A	Protein with ParB-like nuclease domain in PFGI-1-like cluster	Protein with ParB-like nuclease domain in PFGI-1-like cluster:p.Glu295Ala
A	C	Predicted transcriptional regulator of the myo-inositol catabolic operon	Predicted transcriptional regulator of the myo-inositol catabolic operon:p.Phe42Leu
G	C	Phosphoenolpyruvate-dihydroxyacetone phosphotransferase operon regulatory protein DhaR	Phosphoenolpyruvate-dihydroxyacetone phosphotransferase operon regulatory protein DhaR:p.Ser13Cys
G	A	NAD-specific glutamate dehydrogenase (EC 1.4.1.2); NADP-specific glutamate dehydrogenase (EC 1.4.1.4)	NADP-specific glutamate dehydrogenase (EC 1.4.1.4):p.Glu129Gly
G	T	MoaE protein (Molybdenum cofactor biosynthesis)	MoaE protein:p.Met86Arg
C	T	LSU ribosomal protein L15p (L27Ae)	LSU ribosomal protein L15p (L27Ae):p.Leu79Pro
G	C	Glycerate-2-kinase (EC 2.7.1.165)	Glycerate-2-kinase (EC 2.7.1.165):p.Ala143Gly
G	T	General secretion pathway protein L	General secretion pathway protein L:p.Val192Gly
G	T	Ethanolamine ammonia-lyase heavy chain (EC 4.3.1.7)- EutB	Ethanolamine ammonia-lyase heavy chain (EC 4.3.1.7)-EutB:p.Val50Gly
C	T	Cytochrome d ubiquinol oxidase subunit II (EC 1.10.3.-)	Cytochrome d ubiquinol oxidase subunit II (EC 1.10.3.-):p.Glu193Gly
G	T	Chlorohydrolase/deaminase family protein	Chlorohydrolase/deaminase family protein:p.Trp270Gly
G	T	Aspartokinase (EC 2.7.2.4) / Homoserine dehydrogenase (EC 1.1.1.3)	Aspartokinase (EC 2.7.2.4) / Homoserine dehydrogenase (EC 1.1.1.3):p.His759Pro
G	A	Alcohol dehydrogenase (EC 1.1.1.1)	Alcohol dehydrogenase (EC 1.1.1.1):p.Ser93Gly
C	A	ADP-L-glycero-D-manno-heptose-6-epimerase (EC 5.1.3.20)	ADP-L-glycero-D-manno-heptose-6-epimerase (EC 5.1.3.20):p.Cys8Gly
C	A	Acryloyl-CoA reductase AcuI/YhdH (EC 1.3.1.84)	Acryloyl-CoA reductase AcuI/YhdH (EC 1.3.1.84):p.Val159Gly
C	A	2-dehydro-3-deoxyphosphogalactonate aldolase (EC 4.1.2.21)	2-dehydro-3-deoxyphosphogalactonate aldolase (EC 4.1.2.21):p.Cys167Gly

All the sequences of colistin-resistant strains were compared against reference sensitive strain (HS11286; NC_016845) to identify known and unknown genetic modifications associated with colistin resistance mechanisms. Interestingly, from the clinical resistant strains (n=100) and *in vitro* resistant clones (n=2), *mgrB* modifications were identified in 70% of the strains. Also, these strains lack other known colistin genetic modifications. Out of the 70% of *mgrB* modified strains, ~25% modifications were due to IS interruption. Majority of the IS elements belong to IS1 and IS5 types (Figure 3.2.6).

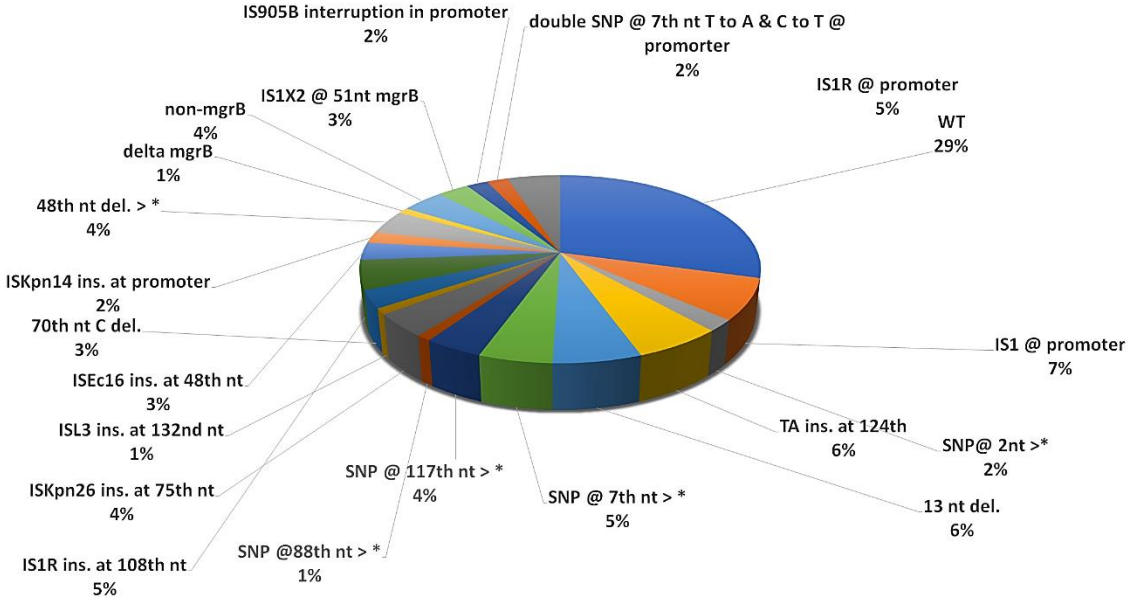


Figure 3.2.6. The diversity of *mgrB* modifications in colistin-resistant strains

Furthermore, the clinical isogenic strains from the patient, who had received colistin therapy were compared to the sequences of *in vitro* generated clone (3319-10R) from the sensitive strain (3319S). The comparative genome analysis of an isogenic clinical pair of 3319S and 3359R revealed that *mgrB* gene of the colistin sensitive strain (3319S) did not have any mutation or disruption of IS elements. Our analysis of these strains revealed that IS element disruption in the strain 3359R is due to one of the largest IS (ISKpn25) elements known to disrupt *mgrB*. ISKpn25 belonged to ISL3, and this IS element was inserted between the 132nd and 133rd base of the *mgrB* gene (Figure 3.2.7A). ISKpn25 is ~8.1 kb long and consists of three ORFs that encode Type 1 Restriction Modifications System (RMS), DNA methyltransferase subunit M, and RMS specificity subunit S and subunit R. Next, the long read scaffolds of the strains (3319S, 3359R and 3319-10R) were screened for plasmid origin scaffolds. The ISKpn25 element was present in the plasmid origin scaffold and in the scaffold harbouring *mgrB* gene.

The ISKpn25 was not found anywhere else in the chromosome. Moreover, blast search indicates that ISKpn25 is associated with plasmids pKpQIL, and it belongs to IncFIB incompatibility typing, which is widely known to harbour many antibiotic resistance genes in different species.

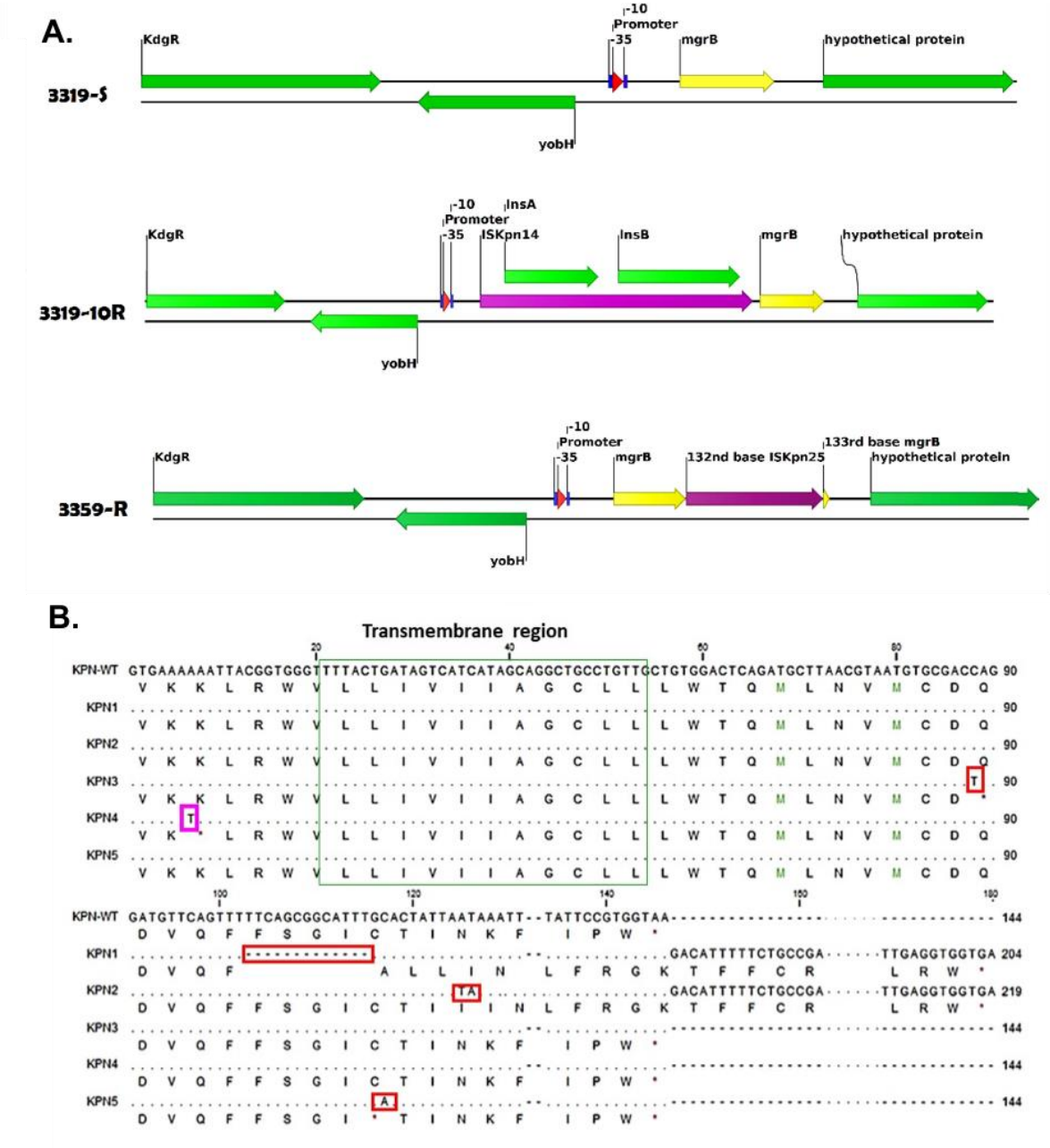


Figure 3.2.7 (A) 3319S *in vivo* colistin sensitive strain with a MIC of 2 µg/mL and shows that *mgrB* is not disrupted. 3319-10R *in vitro* colistin-resistant strain generated from 3319S with an MIC of 64 shows that the promoter region is disrupted by ISKpn14 (IS1). A 3359R colistin-resistant clinical isogenic isolate of the same patient shows that *mgrB* is disrupted by ISKpn25 (ISL3), between 132nd and 133rd base of the gene. (**B**) Novel *mgrB* variants identified in this study. Alignment of wild-type and variant *mgrB* sequences. Novel modifications identified in this study marked in red colour blocks.

Next, we analysed sequences from *in vitro* generated strains 3319-4R and 10R, and found that ISKpn14 had interrupted region between the promoter and the *mgrB* start codon,

with an insertion occurring at the 21st base upstream from the *mgrB* start codon. The 768-bp long ISKpn14 insert harboured two ORFs (*insA* and *insB*) necessary for transposase function (Figure 3.2.7A). Further analysis of scaffolds from these strains suggested that this particular element is not found in the chromosome of the strain 3319S, but in the scaffolds belonging to plasmid pKpQIL. So, we hypothesise that the transposition event might have occurred from the pKpQIL plasmid to the chromosome. The ISKpn14 belongs to the IS1. This family that move by copy and paste mechanisms which was evident as both the pKpQIL and the *mgrB* harboured the ISKpn14 element. From the reference mapping analysis and long read pacbio sequencing, we could determine the pKpQIL plasmid carried two IS elements. Similarly, we did confirm a similar pattern in other *in vitro* clones (3319-4R and 3319-2R) analysed by WGS (data not shown). The size of the plasmid with 113, 640 bp and 109 protein-coding genes and was in line with a previous report of the pKpQIL, (KU874496.1) from Greek strains (21).

We screened other resistant strains for non-*mgrB* modifications and found that only two strains from Belgium (n=1) and Greece (n=1) belong to ST147 with known *pmrB* amino acid change (T157P) and remaining ~ 29 % of *K. pneumoniae* colistin-resistant strains had no known colistin resistance genetic determinants (Figure 3.2.6) (Table 3.2.1).

3.2.3.6 Population analysis of *in vitro* colistin resistant clones identified preferential insertion of ISKpn14 in *mgrB*.

To validate the IS element preference among the resistant strains, we performed population analysis profiling on *in vitro* generated clones from 3319S. The clones selected from different MICs (2, 3, 8, 12, 16, 32 and 64) showed that ISKnp14 interruption in *mgrB* was highly prevalent at every MIC, and this could be because ISKpn14 (780bp) being small might aid an easier mobilisation. However, it also depends on the type of IS elements existing in the genome or population. Also, we identified a low prevalence of ISKpn25 element in *mgrB in vitro* clones (Figure 3.2.8).

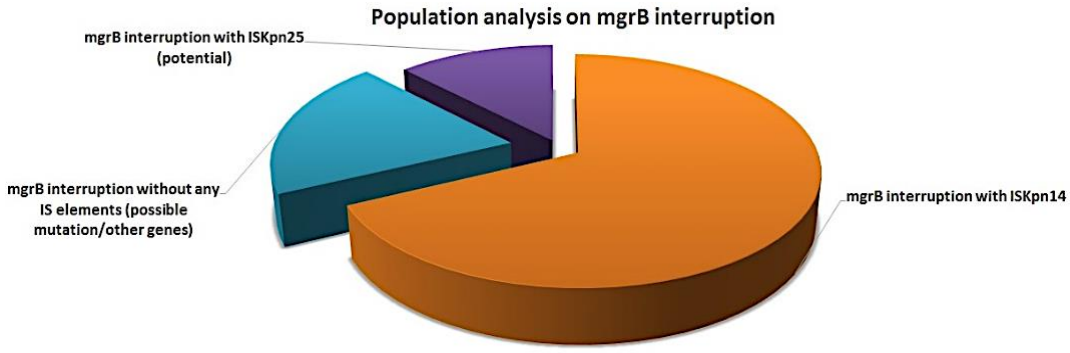


Figure 3.2.8 Population analysis of IS element interruption prevalence in the studied *in vitro* clones.

3.2.3.7 Colistin pressure mediates transposition of IS elements into other genes

From our screening we identified IS element interruptions in other protein-coding genes in the colistin-resistant strains, 3359R and 3319-10R compared to the sensitive counterpart. The promoter region of the translation elongation factor G was interrupted by ISKpn1 (IS3). The pilz domain of *mrKH* was interrupted by ISKpn26 (IS5), and the intergenic region between phosphotransferase system (PTS) fructose specific IIA component and PTS system IIB component was interrupted by IS903B (IS5). The fimbrial protein *staA* gene was disrupted by ISKpn26 (IS5). We also found that a *yqgA*-like gene that is associated with biofilm formation was disrupted byISKpn26 (IS5) in the promoter region. Among these genes, interruptions in pilz domain of *mrkH* and *yqgA*-like genes are interesting and need further research, the ST258 strains 3319S pair does not form a biofilm, whereas *mrkH* and *yqgA*-like genes are involved in biofilm formation. The *mrkH* gene plays a significant role in biofilm formation of *K. pneumoniae* by inducing genes of type 3 fimbriae. The *mrkH* gene has a pilz domain and hence regulated by c-diGMP based signalling (Figure 3.2.9A).

Also, we identified that strains with an IS element (ISKpn26) interruption in the promoter region of *pga* gene cluster by IS5 family. The *pga* gene cluster is shown to be involved in biofilm formation and colony morphology in *K. pneumoniae* suggesting that clinical strains within the host that are under antibiotic pressure can adapt their phenotype (biofilm formation and mucoid or non-mucoid) due to the interruption by IS elements in the genes. However, the screening was done in few sequence types (ST258, ST383 and ST147 and ST101) and it might differ for other sequence types based on their genetic background.

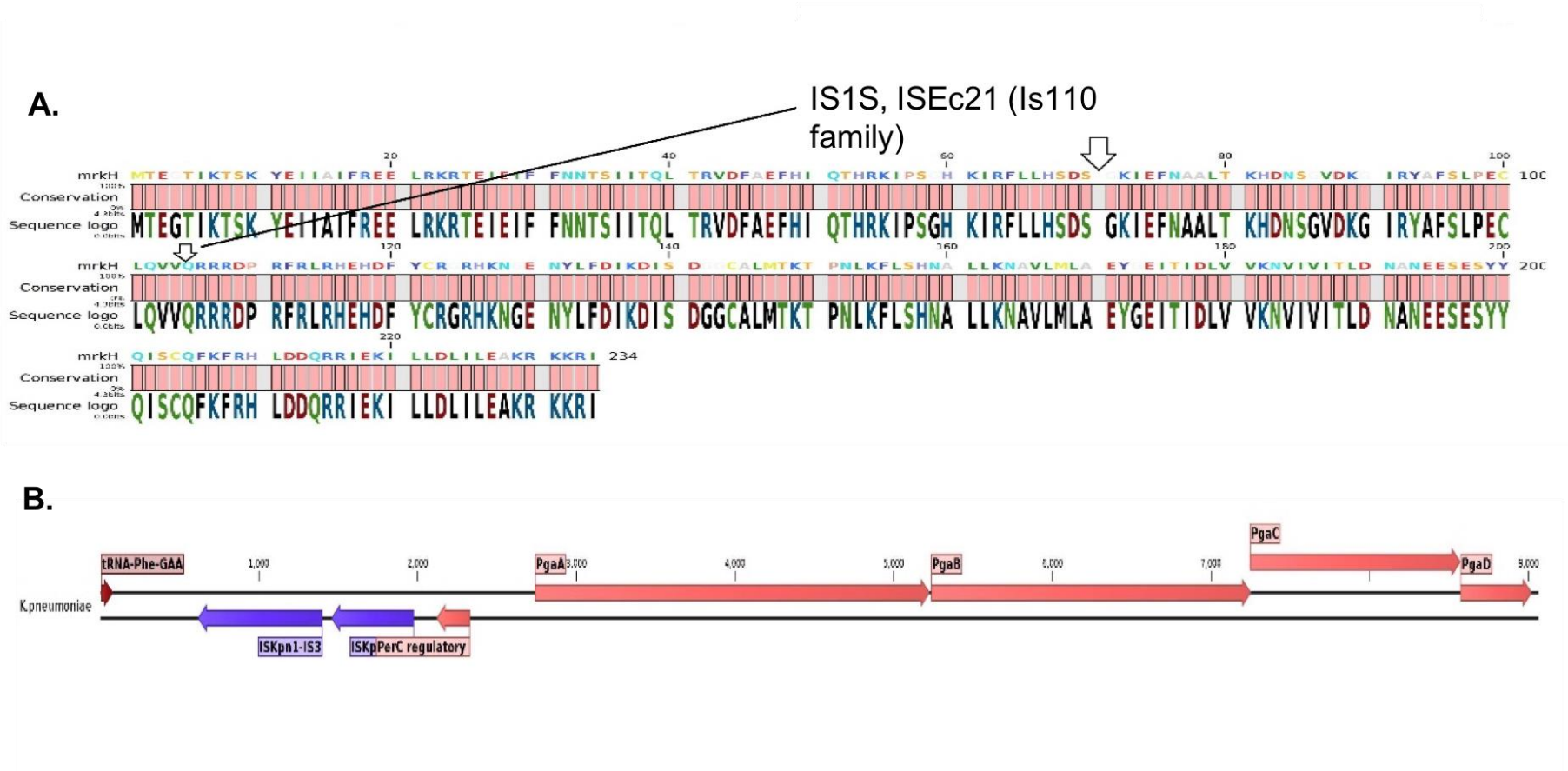


Figure 3.2.9A. The *mrkH* pilZ domain interrupted by IS element between the 70S and 71G (ST258 clone) and 104Q and 105V (ST383) indicated by down arrow. **B.** Biofilm *pgaA* operon also disrupted by IS elements in the promoter region of ST258 clones.

3.2.3.8 Population dynamics of *K. pneumoniae*

3.2.3.8.1 Core genome phylogenetic analysis of colistin sensitive and resistant *K. pneumoniae* strains

The phylogenetic analysis based on core genome SNPs revealed that the different genetic clusters among the studied strains belonged to phylogroup KP1 (Figure 3.2.10). There are a total of 117,942 polymorphic sites identified in the core genomic region compared to HS11286 (ST11), a colistin sensitive strain from public database.

First, we compared the core genome SNPs according to their ST types followed by a comparison of the core genome SNPs in clinical isogenic isolates. The isogenic ST383 clinical pair, 116S and 117R showed an average of 97 core genome SNP differences. Secondly, we compared the core genome SNP differences in the ST258 clinical isogenic pair (3319S-3359R) and identified 420 SNP differences between the sensitive and resistant strain. Interestingly *in vitro* generated clones (3319-4R and 3319-10R) that belonged to ST258 were compared to the sensitive counterpart (3319S) and showed 416 core genome SNPs.

A similar analysis was done for the ST147 clone using isogenic strains (30S and 31R) from a single patient, and core genome SNPs were found to be approximately 1336. Interestingly, ST147 is known to be involved in many outbreaks, has spread globally and is considered to be an emerging epidemic clone.

Analysi of the ST101 strain against HS11286 (ST11) showed average SNP differences of 858 and the strains were isolated from a single hospital. However, two colistin-resistant 'isogenic' strains from a single patient showed SNPs differences of only 45. Interestingly, 4 ST2096 strains from a single Indian hospital isolated from 4 patients over a period of year showed a high average core genome SNP differences of 1920.

From our analysis, base substitution frequency was much higher in the epidemic clones (ST258 and ST147) and lower in upcoming clones (ST383 and ST101).

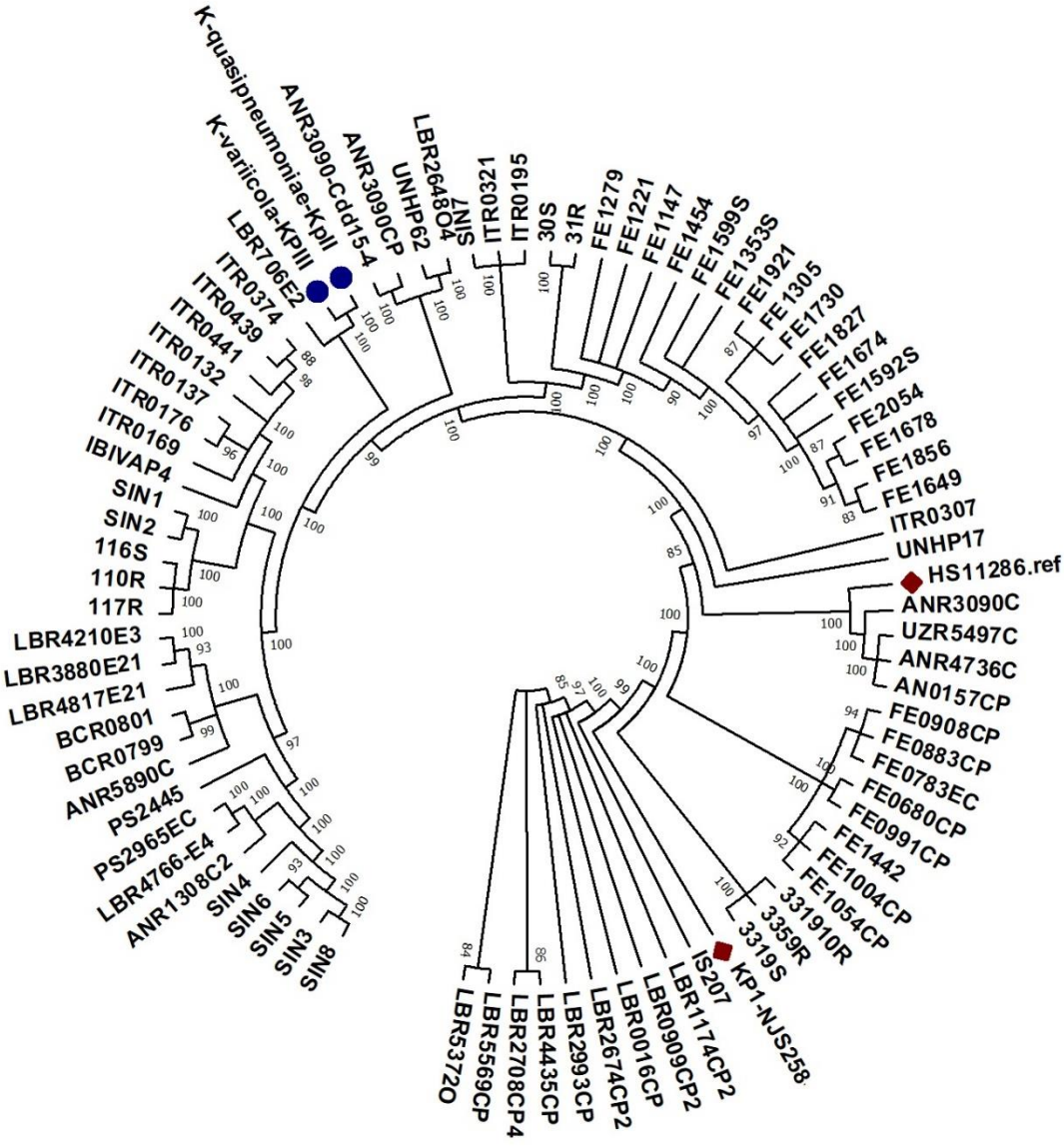


Figure 3.2.10 The core genome SNP based maximum likelihood tree demonstrating the diversity among colistin-resistant *K. pneumoniae* strains in this study**.**

3.2.8.9 Pan-genome analysis

The genomic content of *K. pneumoniae* in the strains analysed is diverse. The core genome consists of 3793 core genes. The accessory genes ranged from 905 to 1765 and depended on the sequence types. The average accessory genes were 1427 and indicated the genomic plasticity of *K. pneumoniae* clones (Figure 3.2.11).

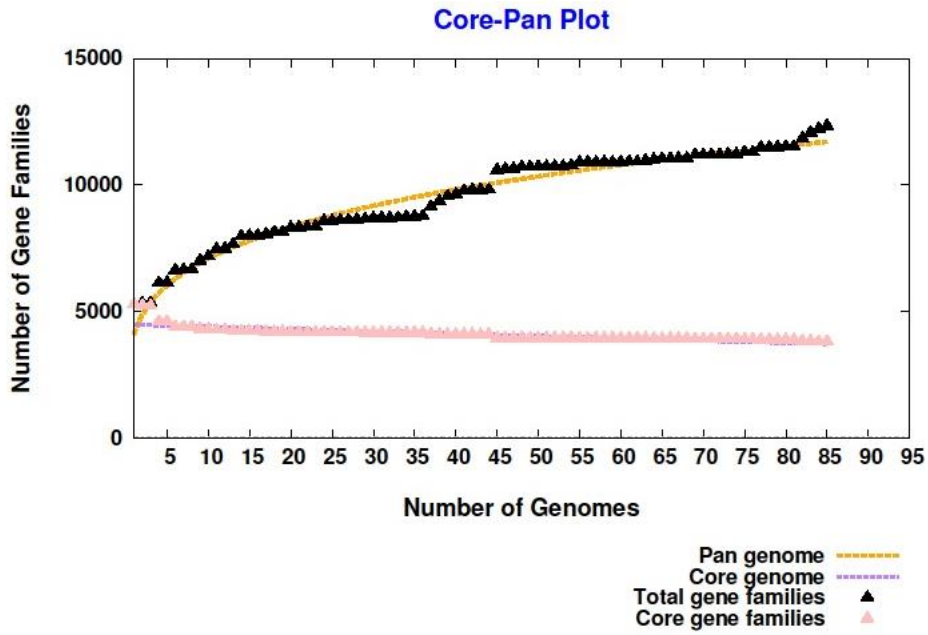


Figure 3.2.11 Core-pan gene diversity profile of the 95 strains

The highest number of accessory genes were from the strains in India that belonged to ST307.

3.2.3.10 Transcriptomic analysis

Transcriptomic analysis of ST383 clones including each of 6 colistin sensitive and resistant strains showed 34 DEGs with a 1.5-fold cut-off (FDR p <0.05). Based on the genetic modifications in the colistin resistant strains *mgrB* gene, transcriptome analysis was divided into four different groups; (i) absence of *mgrB* gene (ii) a single base deletion at 48th 'C' in *mgrB* gene leading to a premature stop codon (iii) interruption between the promoter and start codon of *mgrB* gene caused by IS (ISR1) element (iv) group sensitive and resistant strains with all *mgrB* variations

3.2.3.10.1 Quality of the data

The RNA-sequencing data was processed and checked for quality. The box plot shows that all sequenced samples were comparable, and no variable samples were present. Similar results were obtained from Principal Component Analysis (PCA) plot (Figure 3.2.12A & B).

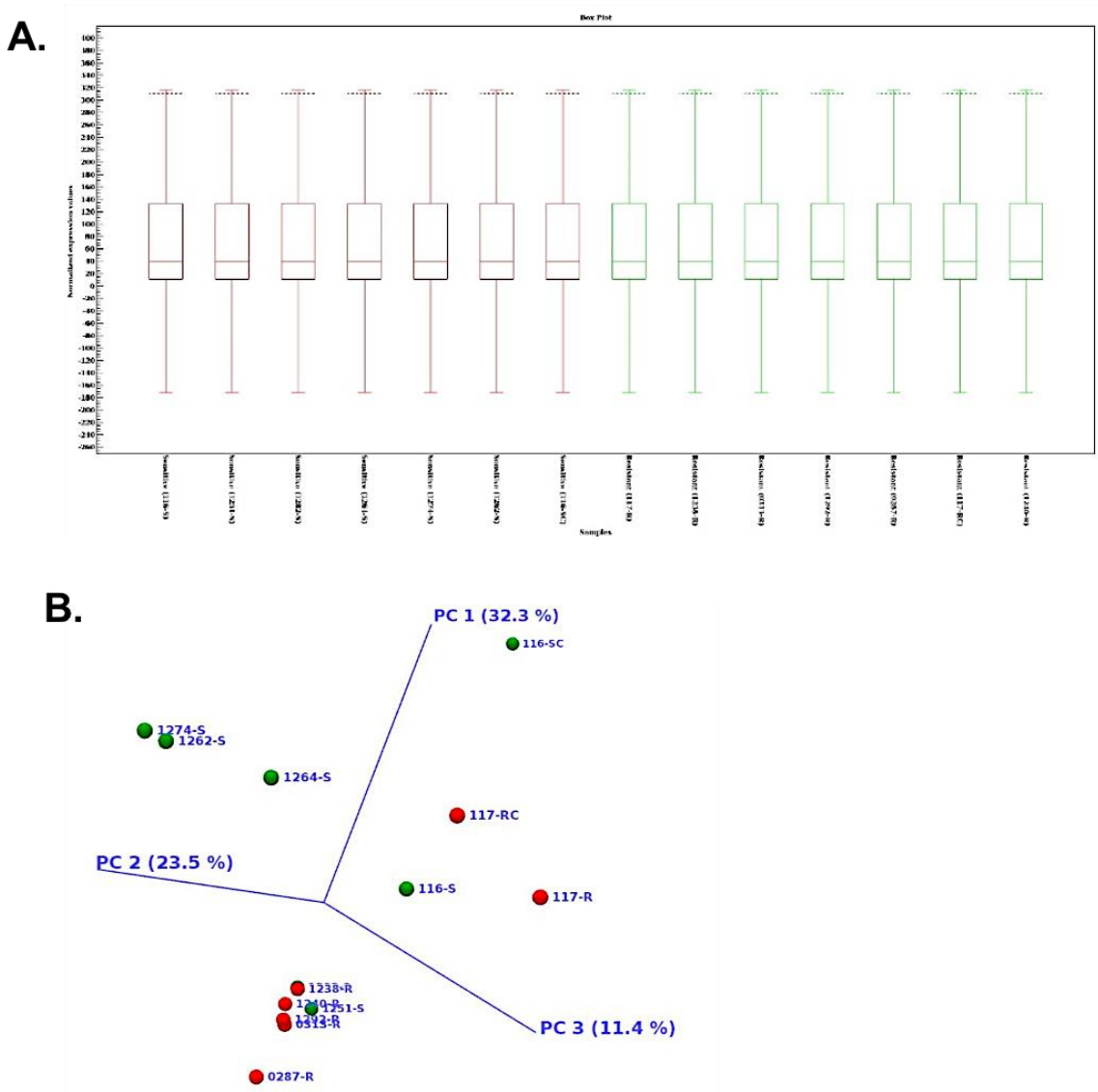


Figure 3.2.12 The box plot of all samples RNA sequencing data after quantile normalisation **B**. A principal component analysis of the samples on the three-dimensional scale spanned by the first, second and third primary component of the covariance matrix. The green colour indicates sensitive and red colour indicates resistant strains. 116SC and 117RC refer to the induced transcriptome of the colistin-resistant phenotype.

The raw sequencing data information is provided in Table 3.2.4, and the mapping efficiency was well above ~85% of the reference used for the transcriptome analysis. The average coverage of all samples was > 180-fold, except one sample 116SC (Table 3.2.5).

Table 3.2.5 The coverage details of RNA sequencing samples

Strain ID	Total Mapped reads	Forward reads	Reversed reads	Average coverage (mapped region)	% GC	% of reference covered
116S	11227133	5304242	5922891	242.99	57.25	87%
117R	9887200	4735210	5151990	217.7	57.25	85%
1238R	19572835	9519005	10053830	313.42	57.25	87%
0313R	20178983	9718401	10460582	323.95	57.25	86%
1251S	21318673	10349168	10969505	340.21	57.25	87%
1292R	17735932	8662023	9073909	287.57	57.25	86%
1282S	6605757	3233120	3372637	109.58	57.25	84%
1264S	19969165	8860361	11108804	323.74	57.25	86%
0287R	19253582	9381422	9872160	313.85	57.25	85%
117RPC	13393718	6206908	7186810	289.56	57.25	87%
1274S	19746867	8000573	11746294	322.3	57.25	85%
1240R	21003261	10115997	10887264	334.71	57.25	87%
1262S	19093331	8315093	10778238	309.9	57.25	85%
116SC	9341539	4860466	4481073	199.07	57.25	88%

First, the sensitive phenotype vs resistant strain (0313R), with no *mgrB* was compared. Interestingly, hypothetical protein: *yebO* (inner membrane protein) was differentially expressed and was ~11-fold lowly expressed compared to sensitive phenotype. The y*ebO* gene was situated downstream immediately after the *mgrB* position (Figure 3.2.7A). It indicated that expression of the *yebO* gene was also controlled by the same regulator and was functionally interlinked with all other inner membrane proteins (Figure 3.2.13, Table 3.2.6)

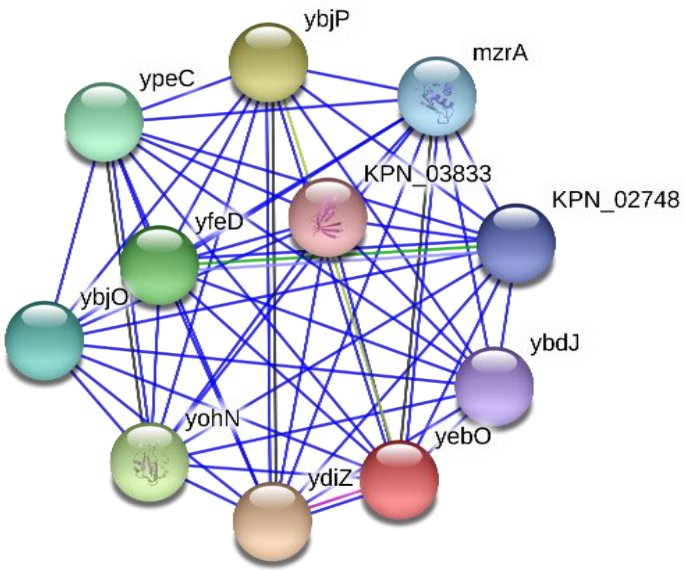


Figure 3.2.13 Protein network of the inner membrane protein *yebO* showed link to functionally linked proteins. Protein network image generated from STRING database https://string-db.org/.

Table 3.2.6 List of differentially expressed genes between sensitive vs Δ *mgrB* strain

S.No	Gene	Log_2 fold change	Fold change	P-value	FDR p-value	Bonferroni
1	FIG00731922: yebO	-11.51	-2909.31	2.25E-013	5.99E-010	1.20E-009
2	Cold shock protein CspC	-11.2	-2360.03	0	0	0
3	Manganese efflux pump, mntP	-8.39	-335.2	4.34E-012	7.69E-009	2.31E-008
4	Pyruvate formate-lyase (EC 2.3.1.54)_2	-5.9	-59.73	8.55E-006	8.36E-003	0.05
5	Maltose-6'-phosphate glucosidase (EC 3.2.1.122)_1	-5.28	-38.9	2.19E-005	0.01	0.12
6	chaperone FimC_1	2.65	6.28	1.72E-005	0.01	0.09
7	COG2197: CheY-like receiver domain and an HTH DNA-binding domain	2.99	7.95	9.35E-005	0.05	0.5
8	type 1 fimbriae protein FimI, unknown function	3.01	8.05	7.18E-007	9.53E-004	3.81E-003
9	Polymyxin resistance protein ArnC, glycosyl transferase (EC 2.4.-.-)	3.22	9.29	8.21E-005	0.04	0.44
10	FIG00731973 putative pile assembly chaperone	3.56	11.79	9.44E-006	8.36E-003	0.05
11	pilZ assembly protein pilZ	3.68	12.79	3.95E-005	0.02	0.21

The second comparison was made between sensitive vs resistant strains (n=2) with *mgrB* modifications of 48th base 'C' deletion that leads to a premature stop codon. In this phenotype, the top most DEGs, cell envelope integrity inner membrane protein (*tolA*), polymyxin resistance protein (*arnC*), glycosyltransferase and malate synthase genes were up-regulated (Figure 3.2.14 & Table 3.2.6).

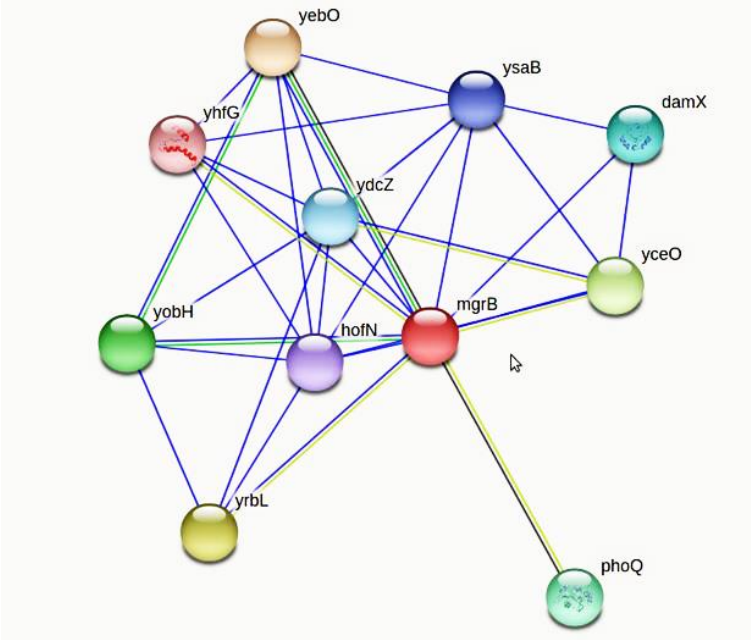


Figure 3.2.14 Protein network of genes linked directly to the *mgrB* gene.

Table 3.2.7 List of differentially expressed genes between sensitive vs resistant strain with *mgrB* truncated protein**.**

S.No	Gene	Log_2 fold change	Fold change	P-value	FDR p-value	Bonferroni
1	Aspartate-semialdehyde dehydrogenase (EC 1.2.1.11)_2	-16.95	-126975.8	3.09E-005	6.31E-003	0.16
2	Cell envelope integrity inner membrane protein TolA	-12.72	-6759.16	0	0	0
3	Ribosomal-protein-serine acetyltransferase (EC 2.3.1.-)	-7.5	-180.55	1.34E-006	4.45E-004	7.13E-003
4	Iron-containing redox enyzme HemeO	-7.24	-151.61	4.85E-009	3.22E-006	2.58E-005
5	Putative amino acid ABC transporter, periplasmic protein	-6.97	-125.14	9.93E-006	2.51E-003	0.05
6	Transcriptional regulator, LysR family_10	-6.68	-102.61	1.41E-007	7.46E-005	7.46E-004
7	Maltose-6'-phosphate glucosidase (EC 3.2.1.122)_1	-6.52	-91.66	8.16E-011	8.67E-008	4.34E-007
8	Pyruvate formate-lyase (EC 2.3.1.54)_2	-6.02	-65.1	2.77E-008	1.64E-005	1.47E-004
9	Acetate kinase (EC 2.7.2.1) @ Propionate kinase (EC 2.7.2.15)	-5	-32.03	9.35E-006	2.48E-003	0.05
10	Threonine catabolic operon transcriptional activator TdcA	-4.79	-27.61	7.05E-006	1.97E-003	0.04
11	hypothetical protein_136	-4.74	-26.76	7.03E-005	0.01	0.37
12	PTS system, maltose and glucose-specific IIC component (EC 2.7.1.69)	-4.7	-26.02	2.91E-009	2.20E-006	1.54E-005
13	Potassium-transporting ATPase A chain (EC 3.6.3.12) (TC 3.A.3.7.1)	-4.66	-25.29	1.32E-006	4.45E-004	7.02E-003
14	L-threonine transporter, anaerobically inducible	-4.54	-23.3	1.58E-005	3.66E-003	0.08
15	L-serine dehydratase, (PLP)-dependent (EC 4.3.1.17)	-4.45	-21.87	1.85E-004	0.03	0.98
16	Ribose ABC transport system, periplasmic ribose-binding protein RbsB (TC 3.A.1.2.1)_3	-3.86	-14.54	1.63E-006	5.09E-004	8.65E-003
17	Potassium-transporting ATPase B chain (EC 3.6.3.12) (TC 3.A.3.7.1)	-3.54	-11.64	4.33E-005	7.93E-003	0.23
18	Transport system permease protein LivH (TC 3.A.1.4.1)_2	-3.42	-10.7	1.19E-005	2.87E-003	0.06
19	Putative sugar phosphotransferase component II B	3.57	11.85	1.11E-004	0.02	0.59
20	Polymyxin resistance protein ArnC, glycosyl transferase (EC 2.4.-.-)	3.65	12.57	9.47E-005	0.02	0.5
21	Isocitrate lyase (EC 4.1.3.1)	3.72	13.15	3.29E-005	6.47E-003	0.17
22	RelB/StbD replicon stabilization protein (antitoxin to RelE/StbE)	3.72	13.21	6.62E-006	1.95E-003	0.04
23	two-component regulatory system wtih UhpB/LuxR	3.91	15.05	7.44E-005	0.01	0.39
24	chaperone FimC_1	3.97	15.62	1.31E-004	0.02	0.69
25	Ribulose-phosphate 3-epimerase (EC 5.1.3.1)_2	4.02	16.23	2.69E-005	5.71E-003	0.14
26	Translation elongation factor G_2	4.1	17.18	2.01E-004	0.03	1
27	FIG00731913: hypothetical protein	4.23	18.77	1.15E-006	4.36E-004	6.10E-003
28	type 1 fimbriae protein FimI, unknown function	4.29	19.58	1.04E-004	0.02	0.55
29	PTS system, fructose-specific IIA component (EC 2.7.1.69)	4.39	21.01	3.85E-005	7.30E-003	0.2
30	Anaerobic glycerol-3-phosphate dehydrogenase subunit C (EC 1.1.5.3)	4.41	21.31	2.02E-005	4.47E-003	0.11
31	hypothetical protein_117	4.44	21.75	1.36E-004	0.02	0.72
32	Anaerobic glycerol-3-phosphate dehydrogenase subunit B (EC 1.1.5.3)	4.75	26.94	3.98E-007	1.76E-004	2.12E-003
33	Malate synthase (EC 2.3.3.9)	4.87	29.25	8.85E-007	3.61E-004	4.70E-003
34	RND multidrug efflux transporter; Acriflavin resistance protein	5.01	32.33	2.87E-009	2.20E-006	1.52E-005
36	RND efflux system, membrane fusion protein CmeA	6.68	102.68	5.81E-013	7.71E-010	3.09E-009
37	Right origin-binding protein_1	9.54	745.19	1.20E-013	2.12E-010	6.37E-010
38	NAD(P)H-flavin oxidoreductase_1	9.71	837.13	0	0	0

Our third comparison was between sensitive vs resistant strains that had an IS element interruption at promoter region, and this interruption was usually identified with majority of resistant *K. pneumoniae.* IS1 and IS5 elements directly impact the *mgrB* expression showing a 6-fold lowered expression compared to *pilZ* domain of *mrkH* and fimbria like proteins that showed high expression (>8 fold) in the resistant strains, suggesting the role of IS elements in controlling the expression of genes (Table 3.2.7).

Table 3.2.8 List of differentially expressed genes between sensitive vs IS element interruption phenotype**.**

S.No	Gene	Log_2 fold change	Fold change	P-value	FDR p-value	Bonferroni
1	mgrB	-6	-63.8	1.60E-007	4.24E-004	8.48E-004
2	FIG00731934: pilZ assembly protein pilZ	4.98	31.65	1.31E-006	2.32E-003	6.95E-003
3	putative fimbrial-like protein_1	8.35	326.46	7.77E-010	4.13E-006	4.13E-006

Lastly, all sensitive strains were combined and treated as a single group and was compared against all the resistant strains, and the topmost DEGs are shown in Figure 3.2.15 and Table 3.2.9. The comparison showed that propanediol gene cluster was lowly-expressed in the resistant phenotype. Interestingly, all the comparisons showed higher expression of pilZ domain, RND multi-drug resistance exporter gene and *arnC*.

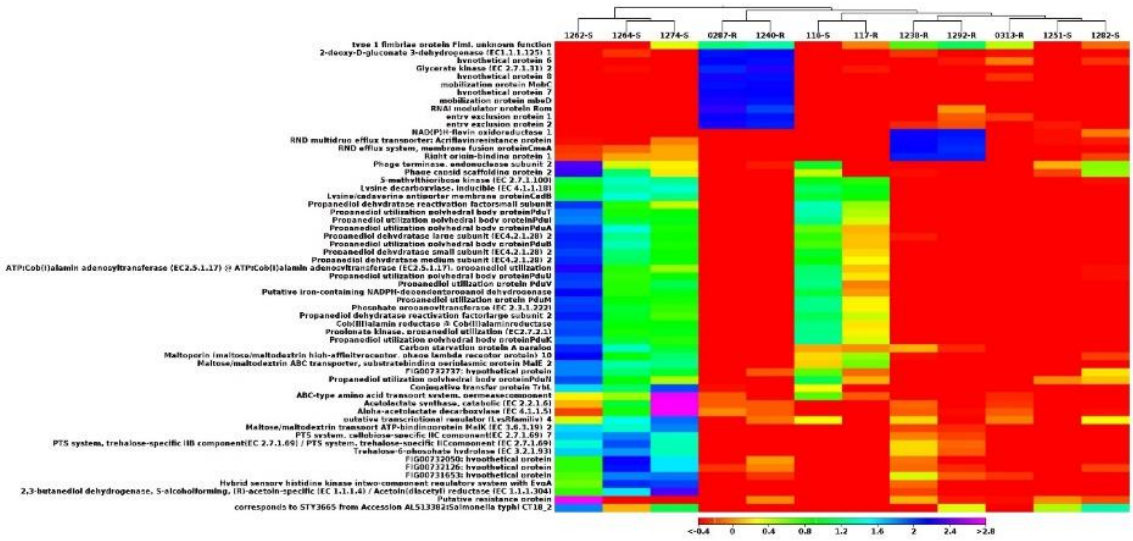


Figure 3.2.15 The heat map shows that differentially expressed genes between colistin sensitive and resistant group of strains.

Table 3.2.9 List of differentially expressed genes between sensitive vs resistant phenotype as a group-based comparison

S.No	Gene	Log_2 fold change	Fold change	P-value	FDR p-value	Bonferroni
1	Propanediol utilization polyhedral body protein PduJ	-3.79	-13.88	1.24E-06	9.41E-04	6.59E-03
2	Propanediol utilization polyhedral body protein PduN	-3.62	-12.33	1.37E-08	3.63E-05	7.25E-05
3	Propanediol dehydratase small subunit (EC 4.2.1.28)_2	-3.34	-10.11	1.25E-05	4.61E-03	0.07
4	Propanediol utilization polyhedral body protein PduK	-3.3	-9.87	3.15E-05	9.84E-03	0.17
5	Propanediol utilization polyhedral body protein PduA	-3.29	-9.78	1.95E-06	1.29E-03	0.01
6	Propanediol utilization polyhedral body protein PduB	-3.14	-8.84	8.67E-06	4.18E-03	0.05
7	Phosphate propanoyltransferase (EC 2.3.1.222)	-3.14	-8.84	1.13E-05	4.61E-03	0.06
8	Propanediol dehydratase large subunit (EC 4.2.1.28)_2	-2.99	-7.93	3.43E-05	0.01	0.18
9	Propanediol utilization protein PduV	-2.97	-7.85	5.68E-06	3.02E-03	0.03
10	Propanediol utilization polyhedral body protein PduT	-2.97	-7.82	5.04E-05	0.01	0.27
11	Propanediol dehydratase medium subunit (EC 4.2.1.28)_2	-2.91	-7.53	8.13E-05	0.02	0.43
12	Propanediol utilization protein PduM	-2.9	-7.47	4.85E-06	2.86E-03	0.03
13	2,3-butanediol dehydrogenase, Acetoin (diacetyl) reductase (EC 1.1.1.304)	-2.83	-7.1	2.08E-04	0.04	1
14	Propanediol dehydratase reactivation factor large subunit_2	-2.8	-6.95	1.41E-04	0.03	0.75
15	Propanediol utilization polyhedral body protein PduU	-2.65	-6.27	6.54E-05	0.02	0.35
16	Cob(III)alamin reductase @ Cob(II)alamin reductase	-2.5	-5.64	3.31E-04	0.05	1
17	Propionate kinase, propanediol utilization (EC 2.7.2.1)	-2.46	-5.52	1.68E-04	0.03	0.89
18	Putative iron-containing NADPH-dependent propanol dehydrogenase	-2.24	-4.73	2.87E-04	0.04	1
19	FIG00732737: hypothetical protein	-2.21	-4.63	1.80E-04	0.03	0.96
20	ATP:Cob(I)alamin adenosyltransferase (EC 2.5.1.17), propanediol utilization	-2.13	-4.37	1.85E-04	0.03	0.98
21	Putative periplasmic substrate-binding transport protein	1.48	2.79	2.43E-04	0.04	1
22	type 1 fimbrae adaptor subunit FimG	1.5	2.82	3.28E-04	0.05	1
23	UDP-4-amino-4-deoxy-L-arabinose-oxoglutarate aminotransferase (EC 2.6.1.87)	1.62	3.08	4.50E-05	0.01	0.24
24	type 1 fimbriae regulatory protein FimB	1.73	3.33	5.72E-05	0.01	0.3
25	type 1 fimbrae adaptor subunit FimF_1	1.83	3.55	3.81E-05	0.01	0.2
26	FIG01049513: hypothetical protein	1.85	3.61	9.44E-05	0.02	0.5
27	type 1 fimbriae anchoring protein FimD_1	1.89	3.71	1.06E-05	4.61E-03	0.06
28	PTS system, mannose-specific IID component (EC 2.7.1.69)_2	1.92	3.78	2.52E-04	0.04	1
29	chaperone FimC_1	2.45	5.45	8.81E-07	7.80E-04	4.68E-03
30	type 1 fimbriae protein FimI, unknown function	2.61	6.1	2.31E-07	2.45E-04	1.23E-03
31	RND multidrug efflux transporter; Acriflavin resistance protein	2.71	6.56	1.30E-05	4.61E-03	0.07
32	RND efflux system, membrane fusion protein CmeA	3.95	15.4	9.92E-08	1.32E-04	5.27E-04
33	Right origin-binding protein_1	5.3	39.38	2.98E-10	1.58E-06	1.58E-06
34	NAD(P)H-flavin oxidoreductase_1	5.51	45.58	6.63E-08	1.17E-04	3.52E-04

3.2.3.10.2 Analysis of colistin induced transcriptome

In order to get a clean background on colistin specific gene expression profile, the sensitive strain 116S (MIC 2 µg/ml) and resistant strain 117R (MIC 32 µg/ml) was induced under 8 µg/ml of colistin concentration (Table 3.2.10 and Table 3.2.11). Both the strains were compared, and the analysis showed a low expression of ethanolamine utilisation cluster in the induced phenotype (116RC and 117RC). The gene cluster is involved in corrinoid adenosylation pathway that is involved in a variety of nucleoside transfers that uses Mg^{2+} as a cofactor. Also, cation efflux system protein, sugar transport inner membrane protein, an outer membrane protein with unknown function were highly expressed in the induced phenotype.

Table 3.2.10 List of differentially expressed genes between sensitive vs induced (116SC)

S.No	Name	Log_2 fold change	Fold change	P-value	FDR p-value	Bonferroni
1	ATP:Cob(I)alamin adenosyltransferase (EC 2.5.1.17) ethanolamine utilization	-4.64	-24.94	2.82E-08	3.65E-06	1.50E-04
2	N-Acetyl-D-glucosamine ABC transport system, permease protein 2	3.03	8.18	6.99E-09	1.28E-06	3.71E-05
3	Cytosine deaminase (EC 3.5.4.1)_2	3.1	8.58	2.71E-08	3.60E-06	1.44E-04
4	Amino acid ABC transporter, permease protein,	3.21	9.25	3.66E-09	8.10E-07	1.94E-05
5	CFA/I fimbrial subunit C usher protein	3.23	9.41	8.40E-10	3.09E-07	4.46E-06
6	[NiFe] hydrogenase metallocenter assembly protein HypF	3.25	9.52	7.33E-08	6.49E-06	3.89E-04
7	Periplasmic hemin-binding protein	3.26	9.58	3.32E-08	3.88E-06	1.77E-04
8	Gamma-glutamyltranspeptidase (EC 2.3.2.2)_2	3.32	10.01	6.81E-08	6.36E-06	3.62E-04
9	Ribokinase (EC 2.7.1.15)_1	3.35	10.22	1.99E-08	2.85E-06	1.06E-04
10	Nitrite reductase [NAD(P)H] large subunit (EC 1.7.1.4)_2	3.35	10.22	7.00E-08	6.36E-06	3.72E-04
11	3-oxosteroid 1-dehydrogenase (EC 1.3.99.4)_2	3.36	10.24	3.08E-08	3.81E-06	1.64E-04
12	Phosphonate ABC transporter ATP-binding protein (TC 3.A.1.9.1)	3.41	10.64	2.96E-09	7.13E-07	1.57E-05
13	Butyryl-CoA dehydrogenase (EC 1.3.8.1)_1	3.45	10.96	5.28E-11	4.68E-08	2.81E-07
14	PhnJ protein	3.46	11.01	5.46E-08	5.48E-06	2.90E-04
15	Methionine ABC transporter ATP-binding protein_1	3.46	11.04	9.62E-09	1.60E-06	5.11E-05
16	transport system permease protein_2	3.51	11.37	1.21E-08	1.89E-06	6.43E-05
17	Taurine-binding periplasmic protein TauA	3.52	11.49	4.85E-09	1.01E-06	2.58E-05
18	Manganese transport protein MntH_2	3.54	11.59	7.07E-08	6.36E-06	3.75E-04
19	oligopeptide/dipeptide ABC transporter, ATPase subunit	3.55	11.71	1.72E-10	9.15E-08	9.15E-07
20	Maltoporin (maltose/maltodextrin high-affinity receptor, phage lambda receptor protein	3.55	11.74	3.39E-09	7.84E-07	1.80E-05
21	Replication protein P_1	3.57	11.84	1.21E-08	1.89E-06	6.42E-05
22	extracellular solute-binding protein, family 3_1	3.59	12	5.88E-08	5.78E-06	3.12E-04
23	Cobalt/zinc/cadmium efflux RND transporter, CzcB family	3.63	12.37	3.71E-08	4.11E-06	1.97E-04
24	Carbonic anhydrase (EC 4.2.1.1)_3	3.64	12.5	4.49E-08	4.67E-06	2.38E-04
25	Formate hydrogenlyase subunit 3	3.66	12.66	9.51E-10	3.16E-07	5.05E-06
26	Phosphonates transport ATP-binding protein PhnK	3.73	13.28	8.31E-09	1.42E-06	4.41E-05
27	Uncharacterized protein ImpJ/VasE_1	3.73	13.3	5.44E-09	1.04E-06	2.89E-05
28	3-carboxyethylcatechol 2,3-dioxygenase (EC 1.13.11.16)	3.74	13.34	4.95E-09	1.01E-06	2.63E-05
29	Outer membrane protein_1	3.79	13.82	5.49E-09	1.04E-06	2.92E-05
30	Methionine ABC transporter permease protein_1	3.87	14.64	3.74E-11	3.98E-08	1.99E-07
31	Cation efflux system protein CusF precursor	3.91	15.01	3.17E-08	3.82E-06	1.68E-04
32	Enterobactin exporter EntS	3.95	15.42	2.04E-09	5.23E-07	1.09E-05
33	General secretion pathway protein K	3.95	15.44	2.07E-09	5.23E-07	1.10E-05
34	L-galactonate dehydrogenase (EC 1.1.1.-)	3.96	15.52	3.88E-08	4.13E-06	2.06E-04
35	3',5'-cyclic-nucleotide phosphodiesterase (EC 3.1.4.17)_2	4.08	16.86	1.16E-09	3.64E-07	6.19E-06
36	Ferric iron ABC transporter, permease protein_2	4.23	18.71	4.14E-10	1.71E-07	2.20E-06
37	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase	4.24	18.91	7.10E-11	5.39E-08	3.77E-07
38	Putative transmembrane sugar transport protein	4.26	19.1	3.35E-10	1.62E-07	1.78E-06
39	3-oxosteroid 1-dehydrogenase (EC 1.3.99.4)_1	4.43	21.59	1.08E-10	6.38E-08	5.74E-07
40	ABC-type sugar transport system, periplasmic component_1	4.61	24.35	8.66E-11	5.75E-08	4.60E-07
41	Oxaloacetate decarboxylase alpha chain (EC 4.1.1.3)	4.67	25.38	5.38E-14	9.53E-11	2.86E-10
42	putative membrane protein_3	4.7	26.08	1.41E-08	2.08E-06	7.49E-05
43	Formate hydrogenlyase subunit 2	4.79	27.76	4.44E-16	2.36E-12	2.36E-12
44	Nitrate ABC transporter, nitrate-binding protein	4.96	31.11	2.22E-15	5.90E-12	1.18E-11
45	Oxaloacetate decarboxylase gamma chain (EC 4.1.1.3)	5.14	35.25	8.72E-10	3.09E-07	4.63E-06
46	Nitrate ABC transporter, permease protein	5.6	48.5	1.69E-09	4.72E-07	8.97E-06

Similarly, we induced phenotype of resistant strain transcriptome shows common genes such as ethanolamine utilisation cluster. As this cluster have a direct role in lipid A modification which is the main target for colistin resistance.

Table 3.2.11. List of differentially expressed genes between resistant (117R) vs induced (117RC) transcriptome

S.No	Gene	Log_2 fold change	Fold change	P-value	FDR p-value	Bonferroni
1	Ethanolamine utilization protein EutQ	-6.52	-91.63	0	0	0
2	Ethanolamine utilization polyhedral-body-like protein EutN	-6.32	-79.94	0	0	0
3	Ethanolamine utilization	-6.24	-75.4	0	0	0
4	Phosphate acetyltransferase (EC 2.3.1.8), ethanolamine utilization-specific	-5.77	-54.68	0	0	0
5	Ethanolamine utilization protein EutP	-5.6	-48.67	0	0	0
6	Acetaldehyde dehydrogenase (EC 1.2.1.10), ethanolamine utilization cluster	-5.27	-38.7	2.81E-13	1.86E-10	1.49E-09
7	Ethanolamine utilization protein EutJ	-5.2	-36.7	0	0	0
8	Ethanolamine utilization polyhedral-body-like protein EutM	-4.95	-30.81	1.74E-14	1.32E-11	9.26E-11
9	Aerobic glycerol-3-phosphate dehydrogenase (EC 1.1.5.3)	-4.93	-30.4	7.47E-10	2.21E-07	3.97E-06
10	Ethanolamine utilization polyhedral-body-like protein EutS	-4.84	-28.56	1.05E-12	5.07E-10	5.58E-09
11	ATP phosphoribosyltransferase (EC 2.4.2.17)	-4.83	-28.4	8.03E-10	2.25E-07	4.27E-06
12	Autoinducer 2 (AI-2) AI-2 binding protein LsrB	-4.56	-23.64	1.62E-08	3.75E-06	8.62E-05
13	Ethanolamine utilization protein EutG	-4.56	-23.63	1.23E-12	5.44E-10	6.52E-09
14	FIG00731653: hypothetical protein	-4.48	-22.35	3.08E-09	8.12E-07	1.63E-05
15	Ethanolamine permease_1	-4.45	-21.84	1.04E-12	5.07E-10	5.54E-09
16	Ethanolamine ammonia-lyase heavy chain (EC 4.3.1.7)_1	-4.32	-20.03	1.61E-11	6.58E-09	8.55E-08
17	Autoinducer 2 (AI-2) LsrC	-4.18	-18.1	3.63E-10	1.21E-07	1.93E-06
18	Ethanolamine ammonia-lyase light chain (EC 4.3.1.7)_1	-4.12	-17.42	1.18E-10	4.17E-08	6.25E-07
19	Autoinducer 2 (AI-2) LsrD	-3.99	-15.87	6.14E-10	1.92E-07	3.26E-06
20	Outer membrane porin OmpF	-3.95	-15.5	7.66E-08	1.57E-05	4.07E-04
21	Ethanolamine utilization protein EutA	-3.76	-13.52	2.49E-11	9.44E-09	1.32E-07
22	Phosphoenolpyruvate-dihydroxyacetone phosphotransferase (EC 2.7.1.121),	-3.67	-12.76	9.25E-08	1.82E-05	4.91E-04
23	Autoinducer 2 (AI-2) kinase LsrK (EC 2.7.1.-)	-3.37	-10.37	7.54E-08	1.57E-05	4.00E-04
24	D-serine dehydratase (EC 4.3.1.18)	3.1	8.58	3.84E-08	8.51E-06	2.04E-04
25	D-serine permease DsdX	3.97	15.69	5.81E-13	3.43E-10	3.08E-09
26	Carbon starvation protein A paralog	4.4	21.11	3.21E-09	8.12E-07	1.70E-05

3.2.3.10.3 Biosignature prediction from transcriptomic data

Based on transcriptomic data, we identified the topmost DEGs that are consistently associated with the colistin-resistant phenotype, i.e., biosignature genes (Table 3.2.12).

Table 3.2.12 Biosignature of colistin resistant *K. pneumoniae*

S. No	Biomarker
1	Glutamate Aspartate periplasmic binding proteinprecursor GltI (TC 3.A.1.3.4)
2	Carbon starvation protein A
3	Formate dehydrogenase O beta subunit (EC1.2.1.2)
4	Inner membrane protein YqjK

The DEGs were mapped on to the reference metabolic pathway of *K. pneumoniae*. We had identified amino acid metabolisms such as tryptophan and histidine, TCA cycle, propionate metabolism, lipopolysaccharide biosynthesis and purine metabolic pathways associated with colistin resistance mechanisms. A detailed functional based study is required to elucidate the complete colistin resistance pathway in *K. pneumoniae* (Figure 3.2.16).

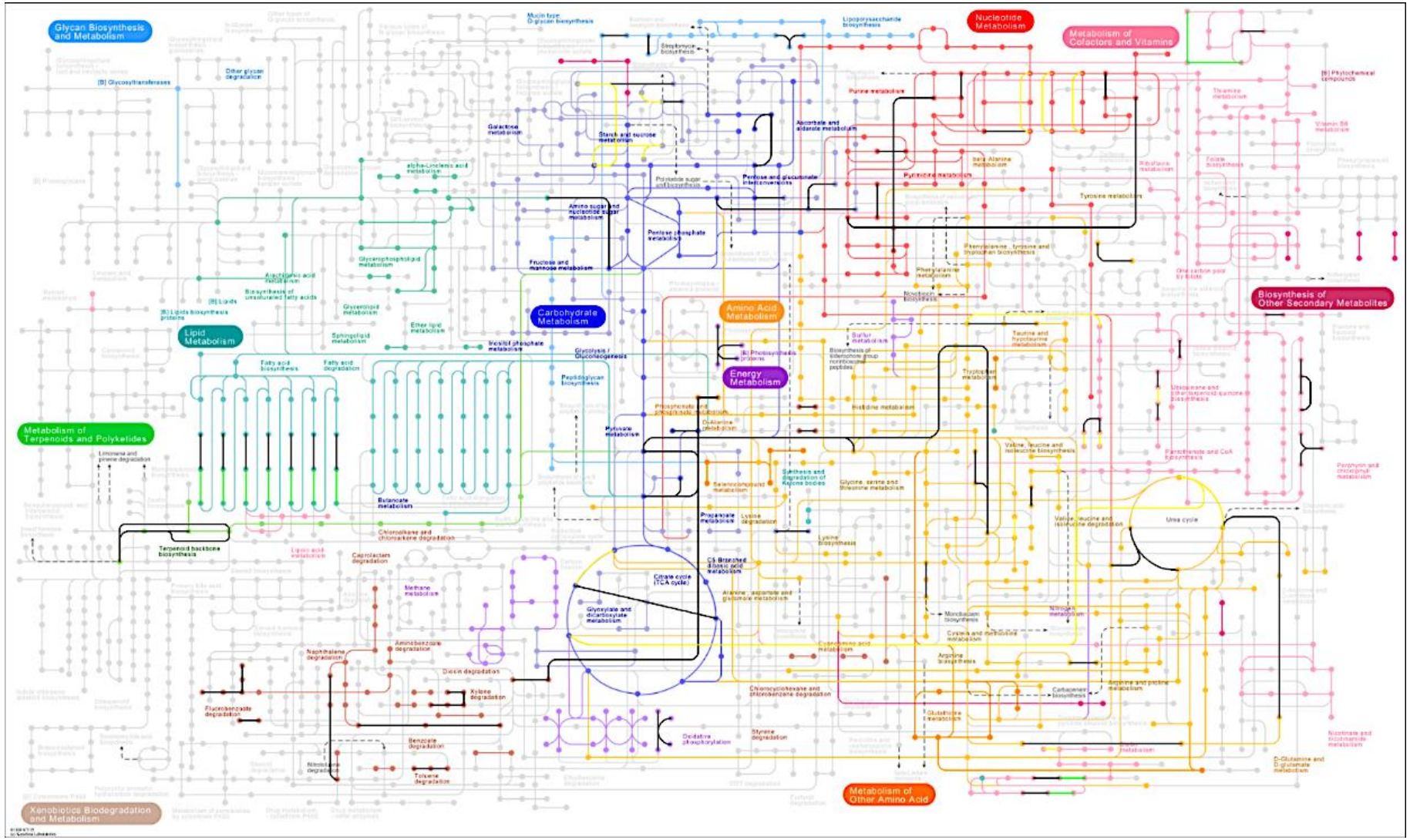


Figure 3.2.16 *K. pneumonia* metabolic pathway shows that all differentially expressed genes from every comparison marked in thick black line.

3.2.4 Discussion

Previous studies have shown that disruption in *mgrB* is the leading cause of colistin resistance in *K. pneumoniae* (7, 29, 30), and *mgrB* modifications are predominantly due to IS elements followed by small deletions and SNPs. Our study showed a similar pattern with high prevalence of modifications of *mgrB* due to IS element interruption. Taking the current knowledge further and with the help of long read technology, our *in vitro* and *in vivo* data provide strong evidence that transposition of IS elements harboured on the pKpQiL plasmid to *mgrB* was triggered by colistin selection pressure. Interestingly, MICs of both *in vitro*-generated and the clinical *mgrB*-disrupted strains were similar (32 and 64 μg/ml), indicating that the first response to colistin pressure is modification or disruption of *mgrB* through mobilization of IS elements.

Of note, both colistin-resistant strains tested have different families of inserted IS elements (the *in vitro* 3319-10R had IS1, and the clinical 3359-R had ISL3). The ST258 3359-R strain was isolated from a patient after 30 days of colistin treatment and showed the disruption of *mgrB* by the ISKpn25 element at the 132nd base of the gene. This is one of the largest (~ 8.1 Kb) element disruption in *K. pneumoniae in vivo* and it is also reported in ST512, a single allelic variant of ST258. Interestingly, the IS1 family element transposition frequency is known to be higher compared to that of the ISL3 element (32) and could also be influenced by potential stress-related pressure.

We also found some novel *mgrB* modifications and disruption by new IS elements that highlight the complexity of the mechanisms as well as the possibility of novel changes in the same target. The MIC values in these strains were also remarkably different and even MICs in the same range did not correlate to similar modifications, which might be related to differences in the genomic background of the strains.

Finally, our transcriptomic approach identified a robust biosignature of the colistin resistant *K. pneumoniae* that needs to be further validated in a larger collection of strains. The induced transcriptome in the presence of colistin enabled us to delineate a unique list of genes that are involved in the immediate response to colistin pressure. These included toxin/antitoxin genes, carbon starvation protein and other stress-related genes.

Author contribution: BBX was involved in experimental work, data analysis, interpretation and drafting manuscript.

3.2.5 References

1. WHO. The evolving threat of antimicrobial resistance - Options for action. 2012.
2. WHO. Antimicrobial resistance: global report on surveillance 2014. Report No.: 978 92 4 156474 8
3. Li J, Nation RL, Turnidge JD, Milne RW, Coulthard K, Rayner CR, et al. Colistin: the re-emerging antibiotic for multidrug-resistant Gram-negative bacterial infections. Lancet Infect Dis. 2006;6(9):589-601.
4. Walsh TR, Toleman MA. The emergence of pan-resistant Gram-negative pathogens merits a rapid global political response. J Antimicrob Chemother. 2012;67(1):1-3.
5. Falagas ME, Kasiakou SK. Colistin: the revival of polymyxins for the management of multidrug-resistant gram-negative bacterial infections. Clin Infect Dis. 2005;40(9):1333-41.
6. Li J, Nation RL, Milne RW, Turnidge JD, Coulthard K. Evaluation of colistin as an agent against multi-resistant in Gram-negative bacteria. International Journal of Antimicrobial Agents. 2005;25(1):11-25.
7. Olaitan AO, Morand S, Rolain JM. Mechanisms of polymyxin resistance: acquired and intrinsic resistance in bacteria. Frontiers in microbiology. 2014;5:643.
8. Cannatelli A, Giani T, D'Andrea MM, Di Pilato V, Arena F, Conte V, et al. MgrB Inactivation Is a Common Mechanism of Colistin Resistance in KPC-Producing Klebsiella pneumoniae of Clinical Origin. Antimicrobial Agents and Chemotherapy. 2014;58(10):5696-703.
9. Srinivasan VB, Rajamohan G. KpnEF, a new member of the Klebsiella pneumoniae cell envelope stress response regulon, is an SMR-type efflux pump involved in broad-spectrum antimicrobial resistance. Antimicrob Agents Chemother. 2013;57(9):4449-62.
10. Liu Y-Y, Wang Y, Walsh TR, Yi L-X, Zhang R, Spencer J, et al. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. The Lancet Infectious Diseases. 2015;16(2):161-8.
11. Xavier BB, Lammens C, Ruhal R, Kumar-Singh S, Butaye P, Goossens H, et al. Identification of a novel plasmid-mediated colistin-resistance gene, mcr-2, in Escherichia coli, Belgium, June 2016. Eurosurveillance. 2016;21(27):30280.
12. Dafopoulou K, Xavier BB, Hotterbeekx A, Janssens L, Lammens C, Dé E, et al. Colistin-Resistant Acinetobacter baumannii Clinical Strains with Deficient Biofilm Formation. Antimicrobial Agents and Chemotherapy. 2016;60(3):1892-5.
13. Huang TW, Lam I, Chang HY, Tsai SF, Palsson BO, Charusanti P. Capsule deletion via a lambda-Red knockout system perturbs biofilm formation and fimbriae expression in Klebsiella pneumoniae MGH 78578. BMC Res Notes. 2014;7:13.
14. Wang X, Dubey AK, Suzuki K, Baker CS, Babitzke P, Romeo T. CsrA post-transcriptionally represses pgaABCD, responsible for synthesis of a biofilm polysaccharide adhesin of Escherichia coli. Molecular Microbiology. 2005;56(6):1648-63.
15. Schumacher MA, Zeng W. Structures of the activator of K. pneumonia biofilm formation, MrkH, indicates PilZ domains involved in c-di-GMP and DNA binding. Proceedings of the National Academy of Sciences of the United States of America. 2016;113(36):10067-72.

16. Fong SS, Joyce AR, Palsson BO. Parallel adaptive evolution cultures of Escherichia coli lead to convergent growth phenotypes with different gene expression states. Genome research. 2005;15(10):1365-72.
17. Liesbet Van H, Samuel C, Christine L, Niel H, Herman G, Surbhi M-K. Antimicrobial Drug Use and Macrolide-Resistant Streptococcus pyogenes, Belgium. Emerging Infectious Disease journal. 2012;18(9):1515.
18. Rozen DE, McGee L, Levin BR, Klugman KP. Fitness Costs of Fluoroquinolone Resistance in Streptococcus pneumoniae. Antimicrobial Agents and Chemotherapy. 2007;51(2):412-6.
19. Xavier BB, Vervoort J, Stewardson A, Adriaenssens N, Coenen S, Harbarth S, et al. Complete Genome Sequences of Nitrofurantoin-Sensitive and -Resistant Escherichia coli ST540 and ST2747 Strains. Genome Announcements. 2014;2(2).
20. Sabirova JS, Xavier BB, Coppens J, Zarkotou O, Lammens C, Janssens L, et al. Whole-genome typing and characterization of blaVIM19-harbouring ST383 Klebsiella pneumoniae by PFGE, whole-genome mapping and WGS. Journal of Antimicrobial Chemotherapy. 2016;71(6):1501-9.
21. Chaudhari NM, Gupta VK, Dutta C. BPGA- an ultra-fast pan-genome analysis pipeline. 2016;6:24373.
22. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics. 2015;31(16):2745-7.
23. Xavier BB, Sabirova J, Pieter M, Hernalsteens J-P, de Greve H, Goossens H, et al. Employing whole genome mapping for optimal de novo assembly of bacterial genomes. BMC Research Notes. 2014;7(1):1-4.
24. Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, et al. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in Klebsiella pneumoniae, an urgent threat to public health. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(27):E3574-E81.
25. 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 Biology. 2014;15(11):524.
26. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution. 2016.
27. Conesa A, Götz S. Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. International Journal of Plant Genomics. 2008;2008:619832.
28. Arnold MFF, Caro-Hernandez P, Tan K, Runti G, Wehmeier S, Scocchi M, et al. Enteric YaiW Is a Surface-Exposed Outer Membrane Lipoprotein That Affects Sensitivity to an Antimicrobial Peptide. Journal of Bacteriology. 2014;196(2):436-44.
29. Cheng Y-H, Lin T-L, Pan Y-J, Wang Y-P, Lin Y-T, Wang J-T. Colistin Resistance Mechanisms in Klebsiella pneumoniae Strains from Taiwan. Antimicrobial Agents and Chemotherapy. 2015;59(5):2909-13.
30. Olaitan AO, Diene SM, Kempf M, Berrazeg M, Bakour S, Gupta SK, et al. Worldwide emergence of colistin resistance in Klebsiella pneumoniae from healthy humans and patients in Lao PDR, Thailand, Israel, Nigeria and France owing to inactivation of the PhoP/PhoQ regulator mgrB: an epidemiological and molecular study. International Journal of

Antimicrobial Agents. 2014;44(6):500-7.
31. Chen L, Mathema B, Pitout JDD, DeLeo FR, Kreiswirth BN. Epidemic Klebsiella pneumoniae ST258 Is a Hybrid Strain. mBio. 2014;5(3).
32. Sousa A, Bourgard C, Wahl LM, Gordo I. Rates of transposition in Escherichia coli. Biol Letters. 2013;9(6)

3.3 COMPLETE SEQUENCE OF AN INCFII PLASMID HARBORING THE COLISTIN-RESISTANCE GENE MCR-1 ISOLATED FROM BELGIAN PIG FARMS

Xavier BB[1], Lammens C[1], Butaye P[2,3,4], Goossens H[1], Malhotra-Kumar S[1].

[1]Laboratory of Medical Microbiology, Vaccine & Infectious Disease Institute, University of Antwerp, Antwerp, Belgium. [2]Faculty of Veterinary Medicine, Ghent University, Ghent, Belgium [3]CODA-CERVA, Brussels, Belgium [4]Ross University School of Veterinary Medicine, ST Kitts and Nevis, West Indies.

3.3.1 Letter of Communication

The monumental increase in antibiotic resistance among important bacterial pathogens, driven by inappropriate and appropriate use of ineffective drugs is currently recognized as one of the most pressing threats to human health by the World Health Organization.(WHO 2014) In particular, the last decade has seen a significant rise in infections caused by multidrug- and extremely drug-resistant (MDR and XDR) Gram-negative pathogens, such as *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Acinetobacter baumannii*. (Nordmann and Poirel 2014; Walsh and Toleman 2012) Antibiotics of the polymyxin group such as colistin are sometimes the only drugs to which these bacteria show susceptibility and therefore, reports of emergence of a plasmid-mediated *mcr-1* encoded mechanism of resistance to colistin have been especially alarming.(Liu, et al. 2015b) We have previously screened rectal swabs collected during 2011-12 in Belgium from post-weaning diarrhea in pigs and found a 13.2% (7/53) prevalence.(Malhotra-Kumar, et al. 2016c) In this study, we carried out a detailed analysis of the 7 *mcr-1* harbouring strains (colistin resistance defined by Sensititre > 4 mg/L) and also sequenced the plasmid harbouring the gene in one porcine strain. Briefly, colistin MICs were first reconfirmed using macro broth dilution method, and were found to be 8 and 16 mg/L for all 7 strains. Screening for ESBLs by PCR,(Monstein, et al. 2007) and by phenotypic tests (double disc diffusion and E-tests) did not identify any strain as an ESBL producer. MLST,(Wirth, et al. 2006) showed that the 7 strains belonged to different STs; ST10 (n=2), ST100 (n=2), ST90 (n=1) and two strains demonstrated novel ST types. Of these, one (allelic profile: 10-27-5-10-172-9-2) was a single allele variant of ST100 (allelic profile: 10-27-5-10-12-9-2) with a SNP in the malate dehydrogenase-encoding *mdh* at position 172, and the other (allelic profile: 10-11-4-8-266-8-2) was as a single allele variant of ST10 (allelic profile: 10-11-4-8-8-8-2) also showing differences in *mdh* at position 266. Geoburst analysis showed that these two novel STs might have evolved from ST100 and ST10, respectively.

Of these, *E. coli* ST10 have been previously associated with human infections, (Leverstein-van Hall, et al. 2011) and KP81, one of the two ST10 strains, was selected for plasmid sequencing. Plasmid DNA was extracted (PureLink HiPure Plasmid Miniprep Kit, Invitrogen, Carlsbad, CA, USA) followed by Nextera XT sample and library preparation and sequencing (MiSeq, v2, 2x150bp, Illumina Inc.,). The plasmid sequence was *de novo* assembled (SPAdes v.3.6.2), annotated (BASys) and compared with the published *mcr-1* gene carrying

plasmid pHNSHP45 (Accession No: KP347127) using CLC Genomics Workbench v7.5.1 (clcbio, Denmark). Incompatibility type, pMLST and insertion sequences were defined using Plasmid finder, pMLST database (https://cge.cbs.dtu.dk/services/) and IS Finder tools (https://www-is.biotoul.fr/).(Carattoli, et al. 2014) Integron types were identified and annotated using blast search against INTEGRALL database (http://integrall.bio.ua.pt/) and RAC: (Repository of Antibiotic Resistance Cassettes http://rac.aihi.mq.edu.au/rac/). Plasmid pKP81-BE (91,049 bp, Figure 1A) encoded 45 ORFs, showed a GC content of 44.9 % and belonged to IncFII incompatibility type and the pMLST type was F2: A-B. The serotype was O50/O2:H32. The plasmid backbone showed only 4% similarity to that of pHNSHP45 (Accession No: KP347127), (Liu, et al. 2015b) which included the *mcr-1* gene and its adjoining insertion sequence. Also, pKP81-BE showed 58.52% similarity to plasmid pKH457-3-BE (accession no. KU353730), an IncP type plasmid carrying *mcr-1* isolated from *E. coli* from neonatal diarrhoea in a calf in the Wallonia region of Belgium.(Malhotra-Kumar, et al. 2016c) In contrast to pKH457-3-BE, the *mcr-1* harbouring ISA*pI1* mobile element on pKP81-BE was not truncated and included the transposase-encoding *tnpA* gene. The pKP81-BE backbone was 99% similar to *Salmonella enterica* subsp. *enterica* serovar Typhimurium strain plasmid pHK0653 (accession no. KT334335) with 78% query coverage. In addition to *mcr-1*, pKP81-BE harboured a *sul3*-containing class 1 integron, *In640*, previously shown to be associated with *Salmonella* and *E. coli* of animal origin.(Antunes, et al. 2007) This integron showed the presence of resistance genes encoding resistance to trimethoprim (*dfrA12*), aminoglycosides (*aadA1a* and *aadA2*), sulphonamides (*sul3*), and phenicols (*cmlA1*) (Figure 1B). Other insertion elements identified were Tn*As1*, Tn*As2*, Tn*As3*, IS*Pa38*, IS*Pa40* (Tn*3*), IS*26*, IS*15DIV*, IS*15DII*, IS*15* (IS*6*).

Interestingly, both plasmids, porcine and bovine, isolated from the northern and southern parts of Belgium were present in *E. coli* belonging to ST10. This ST type although commonly encountered as an antimicrobial-susceptible, low-virulence human intestinal colonizer, also has been associated with human infections and ESBL production.(Leverstein-van Hall, et al. 2011) However, none of the *mcr-1* harbouring *E. coli* in this study were ESBL producers based on phenotypic testing. Further screening of all scaffolds derived from the plasmid sequencing identified an IncH12 plasmid along with the *mcr-1* harbouring IncFII plasmid.

IncFII type plasmids have been important in spreading resistance genes such as $bla_{\text{NDM-1}}$ and $bla_{\text{CTX-M-15}}$ globally (Chen, et al. 2014b) so we chose publicly available IncFII plasmid

sequences from *E. coli* for a comparative analysis. We selected pGUE-NDM (accession no. NC_019089) $bla_{\text{NDM-1}}$ harbouring plasmid from *E. coli* ST131, the ST type reported to be carrying *mcr-1* gene in Europe and China(Hasman, et al. 2015; Liu, et al. 2015b) and pXZ (accession no. JF927996) from China and pHK23a (accession no.JQ432559) also a pig isolate from China. However, comparative analysis showed only <10% similarly restricted to genetic regions associated with plasmid replication and transfer (data not shown). The sequence of the plasmid pKP81-BE was deposited in Genbank: (accession no: KU994859).

To our knowledge, this is the first report of the association of *mcr-1* with an IncFII plasmid. The fact that we identified two different genetic associations of *mcr-1* within a small geographic region (Belgium) further underscores the high promiscuity of the mobile element harbouring *mcr-1*.

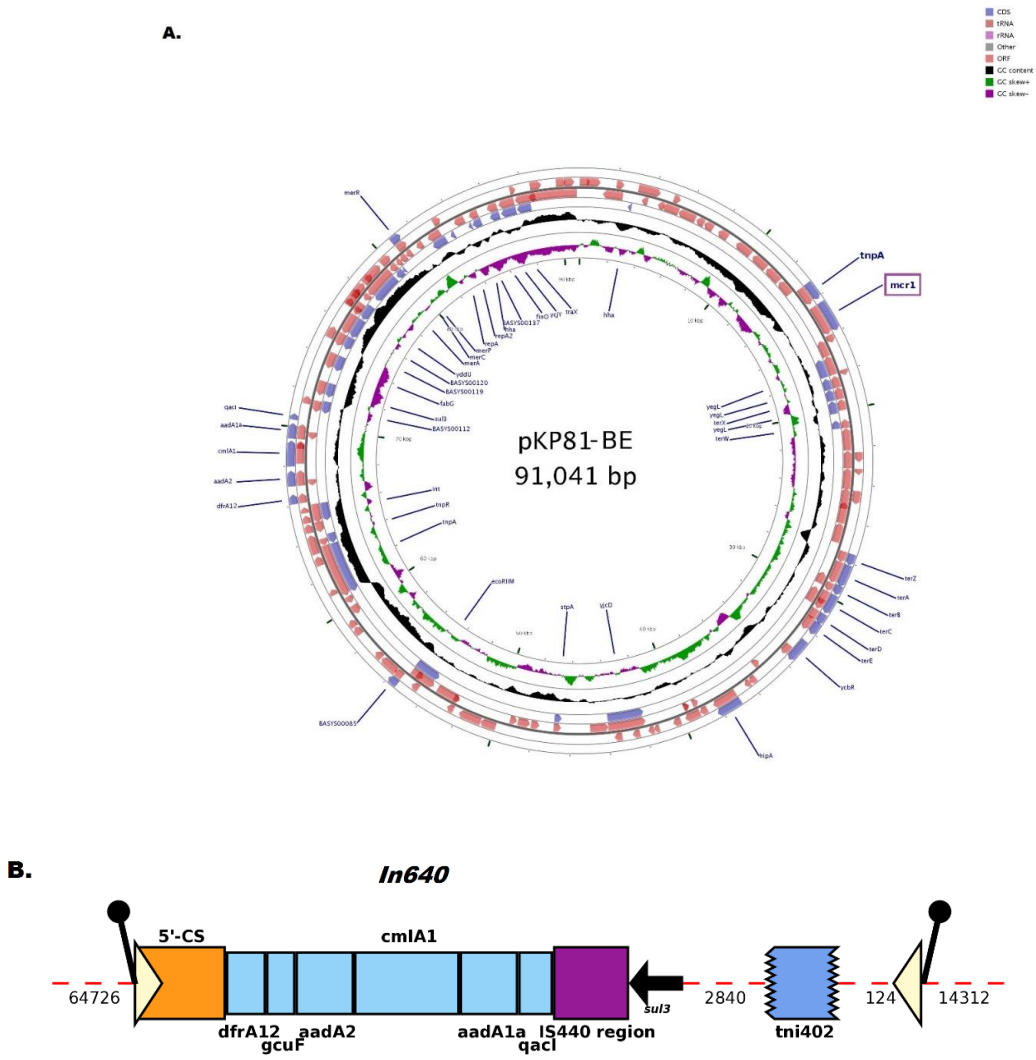


Figure 3.2.1 (A). Genetic organization and structure of *mcr-1* harbouring pKP81-BE. Image was generated using http://stothard.afns.ualberta.ca/cgview_server/ (B). Resistance genes harboured on the *In640* integron.

Author contribution: BBX was involved in experimental work, data analysis, interpretation and drafting manuscript.

3.3.2 References

1. WHO.Antimicrobial resistance: global report on surveillance 2014.http://www.who.int/drugresistance/documents/surveillancereport/en/
2. Walsh TR, Toleman MA. The emergence of pan-resistant Gram-negative pathogens merits a rapid global political response. *J Antimicrob Chemother* 2012; **67**: 1-3.
3. Nordmann P, Poirel L. The difficult-to-control spread of carbapenemase producers among Enterobacteriaceae worldwide. *Clinical Microbiology and Infection* 2014; **20**: 821-30.
4. Liu Y-Y, Wang Y, Walsh TR et al. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *The Lancet Infectious Diseases* 2015; **16**: 161-8.
5. Malhotra-Kumar S, Xavier BB, Das AJ et al. Colistin resistance gene mcr-1 harboured on a multidrug resistant plasmid. *The Lancet Infectious Diseases* 2016; **16**: 283-4.
6. Monstein HJ, ÖStholm-Balkhed Å, Nilsson MV et al. Multiplex PCR amplification assay for the detection of blaSHV, blaTEM and blaCTX-M genes in Enterobacteriaceae. *APMIS* 2007; **115**: 1400-8.
7. Wirth T, Falush D, Lan R et al. Sex and virulence in Escherichia coli: an evolutionary perspective. *Molecular Microbiology* 2006; **60**: 1136-51.
8. Leverstein-van Hall MA, Dierikx CM, Stuart JC et al. Dutch patients, retail chicken meat and poultry share the same ESBL genes, plasmids and strains. *Clinical Microbiology and Infection* 2011; **17**: 873-80.
9. Carattoli A, Zankari E, García-Fernández A et al. In Silico Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. *Antimicrobial Agents and Chemotherapy* 2014; **58**: 3895-903.
10. Antunes P, Machado J, Peixe L. Dissemination of sul3-Containing Elements Linked to Class 1 Integrons with an Unusual 3′ Conserved Sequence Region among Salmonella Isolates. *Antimicrobial Agents and Chemotherapy* 2007; **51**: 1545-8.
11. Chen X, He L, Li Y et al. Complete sequence of a F2:A-:B- plasmid pHN3A11 carrying rmtB and qepA, and its dissemination in China. *Veterinary Microbiology* 2014; **174**: 267-71.
12. Hasman H, Hammerum AM, Hansen F et al. Detection of mcr-1 encoding plasmid-mediated colistin-resistant Escherichia coli isolates from human bloodstream infection and imported chicken meat, Denmark 2015. *Euro surveill.* 2015; **20**(49):pii=30085. DOI: http://dx.doi.org/10.2807/1560-7917.ES.2015.20.49.30085.

3.4 COLISTIN RESISTANCE CONFERRING MCR-1 ISOLATED FROM BELGIAN BOVINE AND PIG FARMS IS HARBORED ON A NOVEL MULTI-DRUG RESISTANT PLASMID

Malhotra-Kumar S[1], **Xavier BB[1]**, Das AJ[1], Lammens C[1], Butaye P[2,3,4], Goossens H[1].

[1] Laboratory of Medical Microbiology, Vaccine & Infectious Disease Institute, University of Antwerp, Antwerp, Belgium. [2] Faculty of Veterinary Medicine, Ghent University, Ghent, Belgium [3]CODA-CERVA, Brussels, Belgium [4]Ross University School of Veterinary Medicine, ST Kitts and Nevis, West Indies.

3.4.1 Letter of Communication

In the November 18, 2015 issue, Liu *et al* reported, for the first time, plasmid-mediated colistin resistance in *Escherichia coli* isolated from animals, food and patients in China. (Liu, et al. 2015c) These data bring to the fore an as yet unknown facet of colistin resistance and yet again highlight the impact of antibiotic use in animal farming on human health.(Callens, et al. 2012; Hammerum, et al. 2014) Here, we screened a selection of 105 colistin-resistant *E. coli* strains (sensititre MICs of colistin ≥ 4 mg/L) isolated during 2011-12 from a passive surveillance of *E. coli* diarrhea in calves (n=52) and piglets (n=53) from the Walloon and Flanders regions of Belgium, respectively. All strains were screened for the presence of *mcr-1* using PCR and Sanger sequencing. The gene was detected in 13/105 (12.4%) *E. coli* (macrobroth dilution MICs of colistin 4 and 8 mg/L), of which 6 (11.5%) and 7 (13.2%) were isolated from calves and piglets, respectively. The *mcr-1* allele showed 100% sequence similarity to the Chinese allele.(Liu, et al. 2015c) Plasmid sequencing (MiSeq, Illumina) from one bovine strain isolated pKH-457-3-BE that showed an IncP backbone and a size of 79,798 bp (Figure 3.3.1). Blast comparison with pHNSHP45 showed 100% similarity only in a short, 2604 bp region that included *mcr-1* (1626 bp) and a truncated IS*ApII* mobile element that did not include the transposase-encoding *tnpA* gene. pKH-457-3-BE showed 99% similarity (73% query coverage) to plasmid pHXY0908 (acc. no. KM877269) found in *Salmonella enterica subsp. enterica serovar typhimurium* isolated from chicken stool in China. In contrast to pHNSHP45, pKH-457-3-BE harbored multiple resistance-encoding genes to trimethoprim (*dfrA1*), tetracycline (*tetA*), aminoglycoside (*aadA1*, *aph (6)-Id or strA*, and *aph(3'')-Ib/strB*) and sulphonamide (*sul1*) antibiotics. Phenotypic testing showed absence of extended-spectrum beta-lactamase and carbapenemase production in all *mcr-1* positive strains.

We demonstrate here a marked presence of *mcr-1* in animal pathogenic bacteria in Europe, an indication that this is already a truly global phenomenon. That *mcr-1* was present in *E. coli* circulating in Belgian farm animals during 2011-2012 and was harbored on a different plasmid backbone than the one isolated from pigs in China (IncI2) or from imported chicken meat in Denmark (IncX4),(Hasman, et al. 2015; Liu, et al. 2015c) indicates a high promiscuity of this gene guided by the adjoining mobile element. Of note, majority of the *mcr-1* positive *E. coli* isolated here were enterotoxigenic (ETEC) and verocytotoxic (VTEC) strains that affect animals but do not cause pathology in humans. It will be of utmost importance to assess transfer frequencies and sustainability of the *mcr-1* harboring genetic elements in human-associated *E. coli* and other Gram-negative pathogens to fully delineate the impact of these findings. Finally,

also noteworthy are the 92 colistin-resistant *E. coli* wherein *mcr-1* was not detected and might potentially harbor other (transferable) colistin-resistant mechanism(s) or even novel *mcr-1* alleles. These findings highlight the importance of a more proactive 'one health' screening approach to identify novel mechanisms and vectors of antibiotic resistance.

Author contribution: BBX was involved in experimental work, data analysis, interpretation and drafting manuscript.

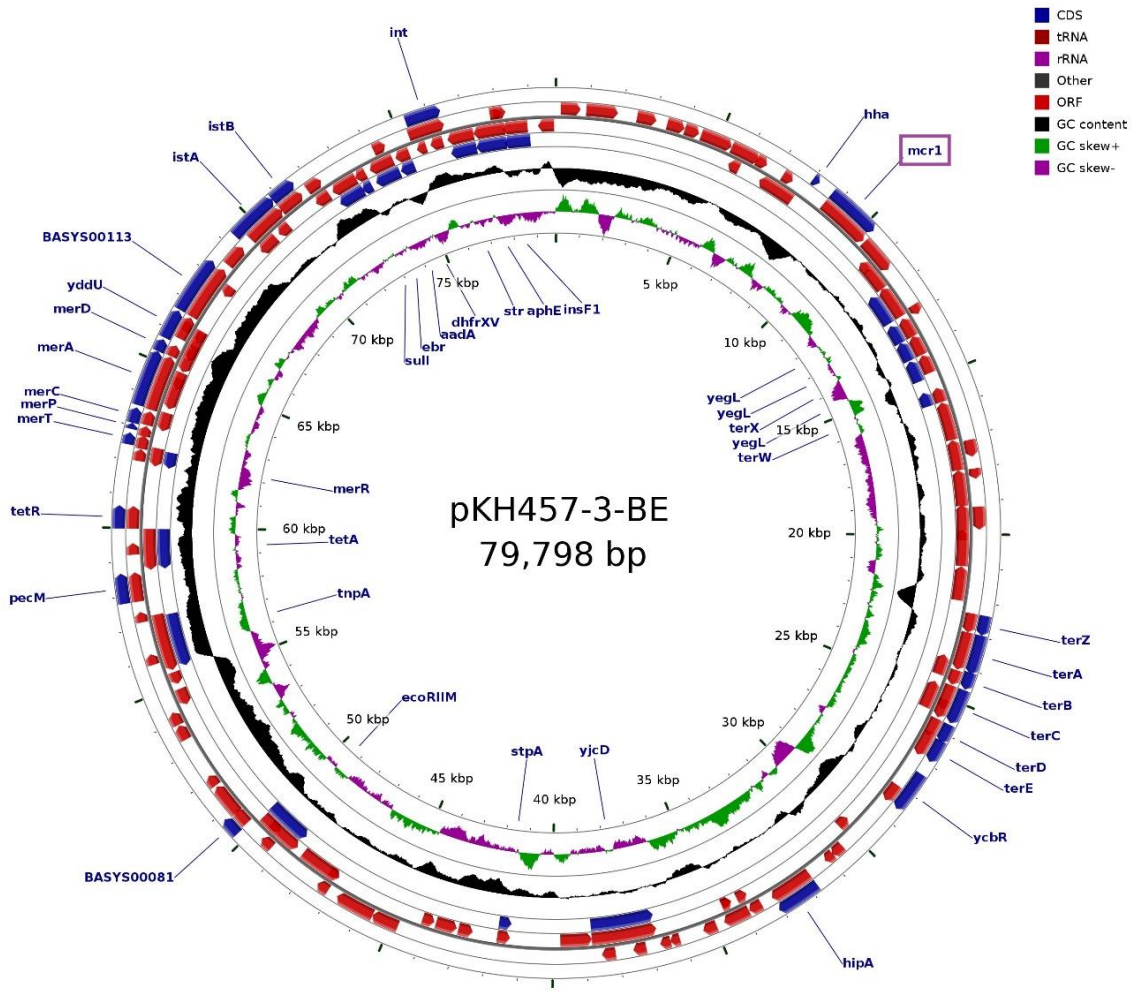


Figure 3.3.1 Genetic organization and structure of *mcr-1* harboring plasmid pKH-457-3-BE. Figure was generated using http://stothard.afns.ualberta.ca/cgview_server/

3.4.2 References

1. Liu Y-Y, Wang Y, Walsh TR, et al. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *The Lancet Infectious Diseases* 2015.
2. Hammerum AM, Larsen J, Andersen VD, et al. Characterization of extended-spectrum β-lactamase (ESBL)-producing Escherichia coli obtained from Danish pigs, pig farmers and their families from farms with high or no consumption of third- or fourth-generation cephalosporins. *Journal of Antimicrobial Chemotherapy* 2014; **69**(10): 2650-7.
3. Callens B, Persoons D, Maes D, et al. Prophylactic and metaphylactic antimicrobial use in Belgian fattening pig herds. *Preventive Veterinary Medicine* 2012; **106**(1): 53-62.
4. Hasman H, Hammerum AM, Hansen F, et al. Detection of mcr-1 encoding plasmid-mediated colistin-resistant Escherichia coli isolates from human bloodstream infection and imported chicken meat, Denmark 2015. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* 2015; **20**(49).

3.5 COLISTIN-RESISTANT *E. COLI* HARBORING MCR-1 ISOLATED FROM FOOD ANIMALS IN HANOI, VIETNAM

Malhotra-Kumar S[1], **Xavier BB[1]**, Das AJ[1], Lammens C[1], Butaye P[2,3,4], Goossens H[1].

[1]Laboratory of Medical Microbiology, Vaccine & Infectious Disease Institute, University of Antwerp, Antwerp, Belgium. [2]Faculty of Veterinary Medicine, Ghent University, Ghent, [3]Belgium CODA-CERVA, Brussels, Belgium. [4]Ross University School of Veterinary Medicine, ST Kitts and Nevis, West Indies.

3.5.1 Letter of Communication

In the November 18, 2015 issue, Liu *et al* reported, for the first time, plasmid-mediated colistin resistance in *Escherichia coli* isolated from animals, food and patients in China.(Liu, et al. 2015c) Here, we screened 24 extended-spectrum beta-lactamase (ESBL, $bla_{\text{CTX-M}}$) harbouring *E. coli* isolated during 2014-15 from rectal swabs taken from chickens on 2 farms (n = 11) in the Van Lam district of the Hung Yen province, and from a pig farm (n = 7) and a pig slaughterhouse (n = 4) located in the Hoai Duc region of the Hanoi province, Vietnam. All strains were screened for the presence of *mcr-1* using PCR and Sanger sequencing. The *mcr-1* gene was detected in 9/24 *E. coli* strains (37.5%), of which 6 were isolated from the rectal swabs of pigs on the farm, one rectal swab from a pig about to be slaughtered, and 2 strains from swabs collected from the lairage area of the slaughterhouse. All 9 isolates exhibited phenotypic colistin resistance (macrobroth dilution MICs 4 and 8 mg/L). The *mcr-1* sequence showed 100% sequence similarity to the gene reported in China.(Liu, et al. 2015c) Plasmid sequencing (MiSeq, Illumina) from one *mcr-1* positive strain isolated multiple genes encoding resistance to trimethoprim (*dfrA12*), tetracycline (*tetA*), aminoglycoside (*aadA3, aph(3')-IA*), phenicol (*cmlA1*), quinolone (*qnrS1, oqxA*), lincosamide (*lnu(F)*) and sulphonamide (*sul2, sul3*), and beta-lactam (extended-spectrum beta-lactamase $bla_{\text{CTX-M55}}$) antibiotics. Blast comparison with the plasmid, pHNSHP45, showed 100% similarity in a 3677 bp region that included *mcr-1* (1626 bp) and a complete IS*ApI1* mobile element including the transposase-encoding *tnpA* gene. Incompatibility typing using PlasmidFinder (Carattoli, et al. 2014) identified IncFII, IncF1A(H1), and IncF1B(K), and IncX1 replicons, however, none of these could be directly linked to the *mcr-1* carrying contig. Of note, the sequence coverage of the *mcr-1* harboring contig was ~100-fold lower than the adjoining regions indicating carriage either on a low-copy plasmid or a chromosomal origin of the gene. Next, we screened 112 ESBL-harboring *E. coli* isolated during 2014-2015 from urines of out- and in-patients with symptomatic urinary tract infections from the National Pediatric, 103 General Military and Cuba hospitals in Hanoi. *mcr-1* was not detected in any of the urinary isolates.

Our findings highlight a high prevalence of the *mcr-1* gene among ESBL-harboring *E. coli* isolated from rectal screenings of pigs in Vietnam; however, we did not find any evidence of transfer or emergence of the gene among ESBL-harboring pathogenic *E. coli* of human origin. These data add to recent studies showing a global emergence of the *mcr-1* harboring mobile genetic element linked to different plasmids,(Hasman, et al. 2015; Malhotra-Kumar S

Submitted) underscoring the importance of active surveillance in both animal and human populations.

Author contribution: BBX was involved in experimental work, data analysis, interpretation and drafting manuscript.

3.5.2 References

1. Liu Y-Y, Wang Y, Walsh TR, et al. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *The Lancet Infectious Diseases* 2015.
2. Carattoli A, Zankari E, García-Fernández A, et al. In Silico Detection and Typing of Plasmids using PlasmidFinder and Plasmid Multilocus Sequence Typing. *Antimicrobial Agents and Chemotherapy* 2014; **58**(7): 3895-903.
3. Hasman H, Hammerum AM, Hansen F, et al. Detection of mcr-1 encoding plasmid-mediated colistin-resistant Escherichia coli isolates from human bloodstream infection and imported chicken meat, Denmark 2015. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* 2015; **20**(49).
4. Malhotra-Kumar S, Xavier BB, Das AJ, Lammens C, Butaye P, Goossens H. Colistin resistance conferring mcr-1 isolated from Belgian bovine and pig farms is harbored on a novel multi-drug resistant plasmid. *The Lancet Infectious Diseases* Submitted.

3.6 IDENTIFICATION OF *MCR-1* GENE IN *ESCHERICHIA COLI* ST88 FROM HUMAN ISOLATE HARBORED IN INCHI2 PLASMID

Xavier BB[1], Bonten M[2], Goossens H[1], Malhotra-Kumar S[1]

[1]Laboratory of Medical Microbiology, Vaccine & Infectious Disease Institute, Universiteit Antwerpen, Antwerp, Belgium; [2]Department of Medical Microbiology and Epidemiology of Infectious Diseases, Julius Center for Health Sciences & Primary Care, UMC Utrecht, The Netherlands.

3.6.1 Letter of Communication

Colistin is one of the last resort to treat extremely and Multi-drug-resistant (MDR and XDR) Gram-negative pathogens, such as *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Acinetobacter baumannii* causing infections (Li, et al. 2006a; Walsh and Toleman 2012) Emergence of colistin resistance are very rapid due to the over use and longer period as mono therapy, so it became very important to tackle the resistance is urgent need (Falagas and Kasiakou 2005; Li, et al. 2005; Walsh and Toleman 2012).Recent identifications of a plasmid-mediated colistin resistance (*mcr*) has attracted global attention. Because, the spread of *mcr*- variants across the world is quite evident. But, plasmid mediate colistin resistance is very limited in humans. (Liu, et al. 2015b).

The strain is subjected to plasmid DNA Isolation by PureLink HiPure Plasmid Miniprep Kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol. Plasmid DNA concentrations were measured by NanoDrop and Qubit. The isolated plasmid DNA were sequenced using Illumina (2 x 250 bp) was performed using Nextera XT sample preparation kit (Miseq). The plasmid sequences were *de novo* assembled using SPAdes (v3.8.1). Assembled contigs were annotated using https://www.basys.ca/ and insertion elements (IS) elements were identified using IS finder https://www-is.biotoul.fr/. Simultaneously, we have also performed reference-based mapping using CLC Genomics workbench 9.5.3 (clcbio, Qiagen) wild type *mcr-1* harboring plasmid pHNSHP45 (accession numberKP347127) used as reference template with more relaxed parameters (overlapping length of 0.5 and similarity of 0.8) to study the mutations or loss of fragments with *mcr-1* and the genetic backbone of the sequenced strains. The plasmid incompatibility type identified using Plasmid finder. The plasmid map was generated using http://stothard.afns.ualberta.ca/cgview_server/.

E. coli isolate was recovered as a surveillance rectal swab from 83-year-old elderly women at the University Hospital Ghent, Ghent Belgium on April 2015. Briefly, colistin MICs were confirmed and were found to be 4 mg/L. Screening for ESBLs by PCR(Monstein, et al. 2007), and by phenotypic tests (double disc diffusion and E-tests) did not identify any strain as an ESBL producer. Highly multi-drug resistant isolate (HMRO) Strain is identified as ST88. Plasmid pUZR4484-BE (193,049 bp, Figure 3.6.1**)** encoded 251 ORFs, showed a GC content of 44.9 % and belonged to IncHI2 incompatibility type. The plasmid backbone showed only 4% similarity to that of pHNSHP45 (Accession No: KP347127) (Liu, et al. 2015b), which included the *mcr-1* gene. It lacks the ISA*pI1* transposon. Also, pUZR4484-BE showed 58.52% similarity to plasmid pKH457-3-BE (accession no. KU353730), an IncP type plasmid carrying *mcr-1* isolated from *E. coli* from neonatal diarrhoea in a calf in the Wallonia region of

Belgium (Malhotra-Kumar, et al. 2016c). In contrast to pKH457-3-BE, the *mcr-1* harbouring ISA*pI1* mobile element on pKP81-BE was not truncated and included the transposase-encoding *tnpA* gene. Diversity of the *mcr* gene associated with IS elements is not new phenomenon. However, the worrisome here is the presence of *mcr-1* gene without IS element in human isolate suggest that it might acquire through horizontal transfer of plasmid either from gut flora and community.

The UZR4484 strain also have other resistance genes encoding resistance to trimethoprim (*dfrA1*), aminoglycosides (*aadA1* and *aacA4*), sulphonamides (*sul1*), and beta-lactamase gene (*blaSHV-44*). To our knowledge, this is the first report of the mcr-*1* from human with an IncHI2 plasmid. The fact that we identified strain harboring ST88 which is another emerging clone seen in animals and human.

Author contribution: BBX was involved in experimental work, data analysis, interpretation and drafting manuscript.

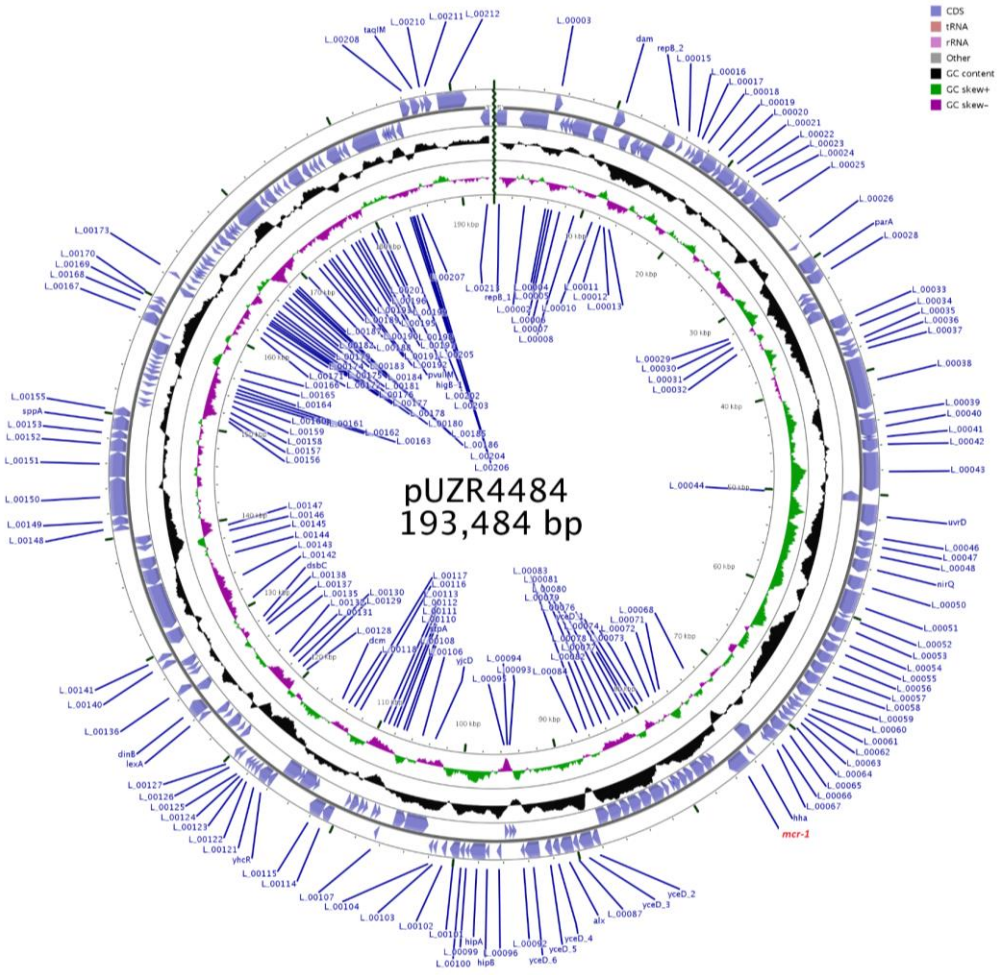


Figure 3.6.1 Genetic organization and structure of *mcr-1* harbouring plasmid pUZR4484-BE.

3.6.2 References

1 Li J, Nation RL, Turnidge JD et al. Colistin: the re-emerging antibiotic for multidrug-resistant Gram-negative bacterial infections. *Lancet Infect Dis* 2006; **6**: 589-601.

2 Walsh TR, Toleman MA. The emergence of pan-resistant Gram-negative pathogens merits a rapid global political response. *J Antimicrob Chemother* 2012; **67**: 1-3.

3 Falagas ME, Kasiakou SK. Colistin: the revival of polymyxins for the management of multidrug-resistant gram-negative bacterial infections. *Clin Infect Dis* 2005; **40**: 1333-41.

4 Li J, Nation RL, Milne RW et al. Evaluation of colistin as an agent against multi-resistant in Gram-negative bacteria. *International Journal of Antimicrobial Agents* 2005; **25**: 11-25.

5 Liu Y-Y, Wang Y, Walsh TR et al. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *The Lancet Infectious Diseases* 2015; **16**: 161-8.

6 Aziz RK, Bartels D, Best AA et al. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 2008; **9**: 1-15.

7 Monstein HJ, ÖStholm-Balkhed Å, Nilsson MV et al. Multiplex PCR amplification assay for the detection of blaSHV, blaTEM and blaCTX-M genes in Enterobacteriaceae. *APMIS* 2007; **115**: 1400-8.

8 Wirth T, Falush D, Lan R et al. Sex and virulence in Escherichia coli: an evolutionary perspective. *Molecular Microbiology* 2006; **60**: 1136-51.

9 Malhotra-Kumar S, Xavier BB, Das AJ et al. Colistin resistance gene mcr-1 harboured on a multidrug resistant plasmid. *The Lancet Infectious Diseases* 2016; **16**: 283-4.

10 Hudson CM, Bent ZW, Meagher RJ et al. Resistance Determinants and Mobile Genetic Elements of an NDM-1-Encoding <italic>Klebsiella pneumoniae</italic> Strain. *PLoS ONE* 2014; **9**: e99209.

11 Leverstein-van Hall MA, Dierikx CM, Stuart JC et al. Dutch patients, retail chicken meat and poultry share the same ESBL genes, plasmids and strains. *Clinical Microbiology and Infection* 2011; **17**: 873-80.

12 Chen X, He L, Li Y et al. Complete sequence of a F2:A-:B- plasmid pHN3A11 carrying rmtB and qepA, and its dissemination in China. *Veterinary Microbiology* 2014; **174**: 267-71.

13 Hasman H, Hammerum AM, Hansen F et al. Detection of mcr-1 encoding plasmid-mediated colistin-resistant Escherichia coli isolates from human bloodstream infection and imported chicken meat, Denmark 2015. *Euro Surveill* 2015; **20**

CHAPTER IV

NEXT GENERATION SEQUENCING TECHNIQUES (WGS AND METAGENOMICS) SERVE AS SOPHISTICATED TOOL IN HUMAN MEDICINE

Chapter IV is further divided into the following sub-chapters:

4.1. **Xavier BB**, Gesuele R, Lammens C, Abdessalam C, Goossens H, Schrenzel J, Harbarth S, Malhotra-Kumar S. 2018. Evidence of an *in vivo* transfer of a *bla*$_{\text{CTX-M14}}$-harboring plasmid under antibiotic pressure, Geneva, Switzerland. (*In revision*)

4.2 **Xavier BB,** Mazlom H, El Hadidi M, Mysara M, Brems P, Lammens C, Huson D, Malhotra-Kumar S, Goossens H. 2018. Faster restoration of the major microbial taxa in the human gut with fecal microbial transplantation (FMT) after treatment with an antibiotic (Moxifloxacin) (*In preparation*)

4.1 EVIDENCE OF AN IN VIVO TRANSFER OF A BLACTX-M14-HARBORING PLASMID UNDER ANTIBIOTIC PRESSURE, GENEVA, SWITZERLAND

Basil Britto Xavier[a], Renzi Gesuele[b], Christine Lammens[a], Cherkaoui Abdessalam[b], Herman Goossens[a], Jacques Schrenzel[b], Stephan Harbarth[b], Surbhi Malhotra-Kumar[a]

[a]Laboratory of Medical Microbiology, Vaccine & Infectious Disease Institute, Universiteit Antwerpen, Antwerp, Belgium; [b]Infection Control Program, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland.

Abstract

Horizontal transfer of plasmid-mediated antibiotic resistance is frequently observed under antibiotic pressure *in vitro*, however, *in vivo* evidence remains rare. We present here evidence of transfer of a $bla_{\text{CTX-M-14}}$ harbouring plasmid from a gut-colonizing *E. coli* to the hypervirulent *K. pneumoniae* ST23 in a patient who had received cefuroxime therapy.

4.1.1 Letter of communication

The spread of antimicrobial resistance presents a grave threat to public health and undermines our ability to respond to bacterial infections. Particularly, transfer of antimicrobial resistance via plasmid exchange is of concern as it enables unrelated (pathogenic) bacteria to acquire resistance. While in vitro experiments provide ample proof of such an event, there are very few reports documenting horizontal transfer of plasmid-mediated antibiotic resistance under antibiotic pressure in a patient (Subbiah, et al. 2011). We present here evidence of transfer of a $bla_{\text{CTX-M-14}}$ harbouring plasmid under cefuroxime therapy.

A 80-year old woman of Italian origin was admitted to the Geneva University Hospitals (HUG), Geneva, Switzerland on 04/10/2017. She was initially hospitalized and operated in February and March 2017 for lateral and anterior lumbar interbody fusions, followed by a surgical revision due to mechanical problems in early April 2017. During those 3 interventions, she received single-shot perioperative prophylaxis with cefuroxime (1.5g IV). One week after her last surgery, she developed signs and symptoms of deep postoperative wound infection at the site of spondylodesis and received empiric IV meropenem therapy (5g per 24h) through continuous perfusion. An extended spectrum beta-lactamase (ESBL)-producing, but carbapenem-sensitive *K. pneumoniae* was isolated from the deep surgical site and blood cultures on 04/10/2017. During the following days, another ESBL-producing *K. pneumoniae* strain was isolated from intra-operative swabs of her spondylodesis. On 04/13/2017, routine rectal swab screening on ESBL chromogenic media (bioMerieux) yielded an ESBL-harboring *E. coli*. On further antimicrobial susceptibility testing, the *E. coli* strain was found to be resistant to amoxicillin, co-amoxiclav, piperacillin, cefuroxime, ceftazidime, ceftriaxone, cefepime and aztreonam and was sensitive to carbapenems, fluoroquinolones, aminoglycosides, fosfomycin and cotrimoxazole. All *K. pneumoniae* isolates from various clinical samples exhibited the same antimicrobial susceptibility profile as the gut colonizing *E. coli* strain, except for fosfomycin resistance. The patient reported to have visited Sicily (Italy) and Moscow (Russia) in 2016, but without any medical contact.

Plasmid DNA from the ESBL-harbouring *E. coli* and one *K. pneumoniae* strain was isolated (PureLink HiPure Plasmid Miniprep Kit, Invitrogen, Carlsbad, USA), measured (NanoDrop and Qubit), and sequenced (Illumina-MiSeq, 2x250 bp, Nextera XT sample preparation kit). Plasmid sequence analysis was done using BacPipe, a bacterial whole genome

sequencing pipeline recently developed by us (Xavier et al, under revision). Briefly, sequences were *de novo* assembled using SPAdes, annotated using Prokka, and insertion elements (IS) elements were identified using IS finder https://www-is.biotoul.fr/. The plasmid incompatibility type was identified using PlasmidFinder, and pMLST was determined. The plasmid map was generated using CLC Genomics Workbench v9.5.3 (Qiagen). Simultaneously, we also performed reference-based mapping (CLC Genomics workbench) using published sequences of IncFII (Accession no HM355591) and IncHIB (Accession No: CP014011.1) as templates with more relaxed parameters (overlapping length of 0.5 and similarity of 0.8) to study mutations or loss of fragments in our plasmids. The variant calling of IncFII plasmids between *E. coli* and *K. pneumoniae* showed no differences, indicating that both harboured the same plasmid with the following characteristics: pMLST F2:A-:B- and incompatibility type IncFII, 70,256 bp size and 81 protein coding regions (Figure 4.1.1). This plasmid harbored $bla_{\text{CTX-M-14}}$ embedded in *ISEcp1* (IS1380 superfamily) (Figure 4.1,1). MLST showed that *E. coli* belonged to ST1 and *K. pneumoniae* to ST23, a hypervirulent clone of the K1 serotype primarily found in the Asia-Pacific region (Shon, et al. 2013). Also, both *E. coli* and *K. pneumoniae* harbored other plasmids and resistance genes: *E coli* showed presence of $bla_{\text{TEM-1C}}$, *strA*, *strB* (both streptomycin resistance encoding genes) and of IncB/O/K/Z and Col156 plasmids. Similarly, *K. pneumoniae* harboured an IncHIB plasmid with 228,916 bp size and 211 protein-coding sequences. This is a 'virulence' plasmid, responsible for the hypervirulent phenotype of this ST type, and carries a host of virulence gene clusters including the aerobactin siderophore biosynthesis cluster and receptor proteins, enterobactin siderophore receptor protein, yersiniabactin cluster and salmochelin gene (Holt, et al. 2015; Schlüter, et al. 2014). Also, the plasmid harbours two copies of the *rmpA* gene that is known to confer hypermucoidy (Struve, et al. 2015). Interestingly, we also found $bla_{\text{SHV-11}}$ integrated in the chromosome. Previous *in vitro* selection experiments have shown that the beta-lactamase $bla_{\text{SHV-11}}$ can easily mutate to an ESBL under cephalosporin pressure and hence its presence signals a predisposition or priming of a *K. pneumoniae* strain to develop resistance to extended-spectrum beta-lactams (Hammond, et al. 2008). However, our data indicates that in the human environment, an ESBL-harbouring plasmid might be preferentially acquired under cephalosporin (cefuroxime) pressure despite the presence of $bla_{\text{SHV-11}}$ in *K. pneumoniae,* which remained unmodified.

Taken together, these findings reiterate the importance of screening patients for presence of ESBLs in the colonizing flora irrespective of travel history of the patient. The hyper virulent ST23 clone is a prevalent cause of liver abscesses and meningitis in Asia and treatment of choice is a broad-spectrum cephalosporin (Ku, et al. 2017). ESBL genes have been rarely detected in

this well-known, community-acquired, hyper-virulent and invasive clone. Infections due to ST23 remain as yet rare in Europe (Gundestrup, et al. 2014). While conjugation events and transfer of ESBL harbouring plasmids occur with a high frequency between *E. coli* and *K. pneumoniae in vitro* and in the mouse gut under cephalosporin pressure (Schjørring, et al. 2008), here we report evidence of a potential *in vivo* transfer of an ESBL-harbouring plasmid under cephalosporin pressure especially into a hyper-virulent *K. pneumoniae* ST23 that would further complicate treatment options for severe infections caused by this clone.

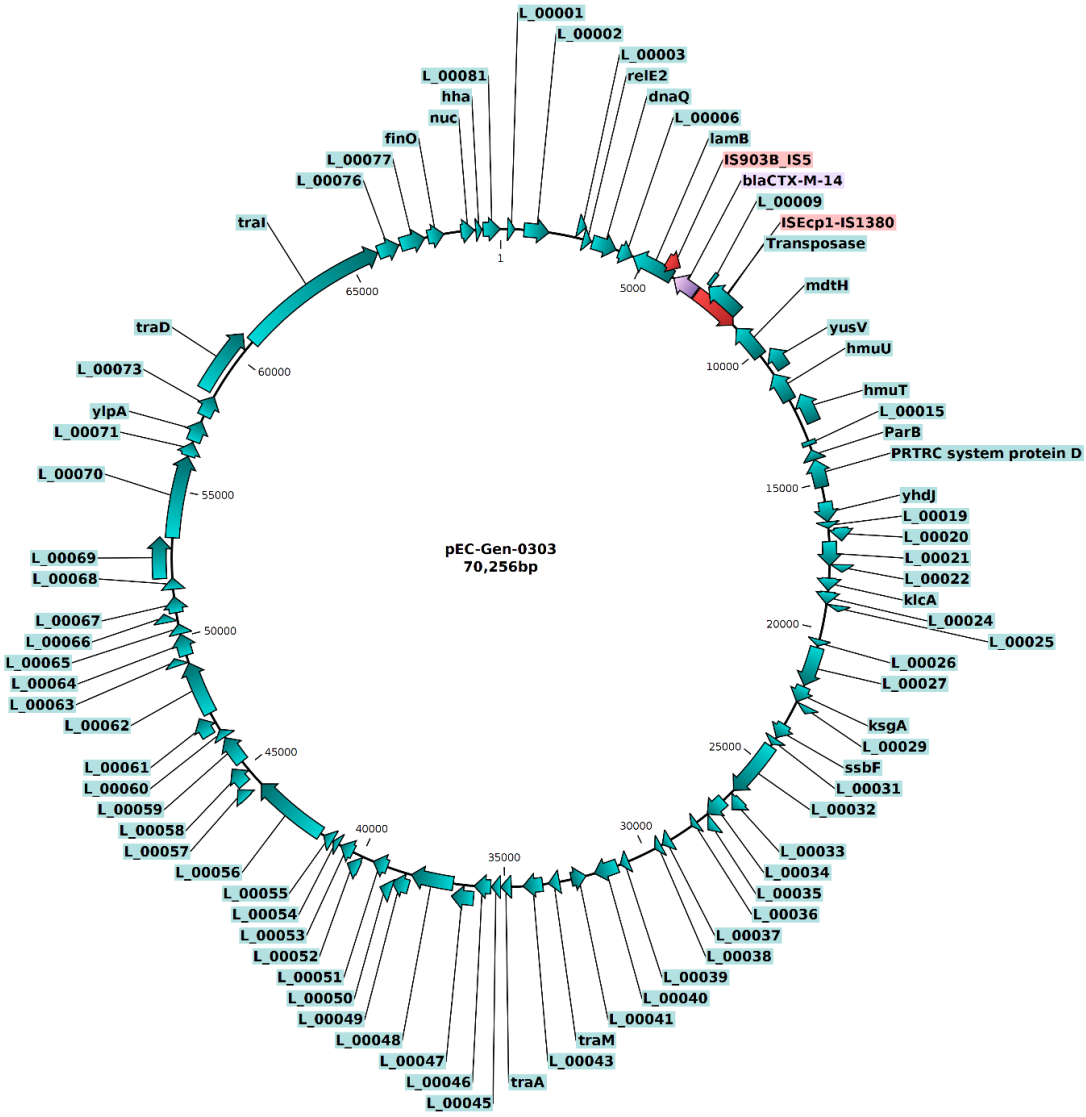


Figure.4.1.1 Genetic map of the *bla*$_{\text{CTX-M14}}$ harbouring IncFII plasmid isolated from both *E. coli* and *K. pneumoniae.*

Author contribution: BBX was involved in experimental work, data analysis, interpretation and drafting manuscript.

4.1.2 References

1. Subbiah M, Top EM, Shah DH, Call DR. Selection Pressure Required for Long-Term Persistence of blaCMY-2-Positive IncA/C Plasmids. Applied and Environmental Microbiology. 2011 July 1, 2011;77(13):4486-93.
2. Shon AS, Bajwa RPS, Russo TA. Hypervirulent (hypermucoviscous) Klebsiella pneumoniae: A new and dangerous breed. Virulence. 2013 01/09 09/21/received 10/29/revised 10/30/accepted;4(2):107-18.
3. Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, et al. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in Klebsiella pneumoniae, an urgent threat to public health. Proceedings of the National Academy of Sciences. 2015 July 7, 2015;112(27):E3574-E81.
4. Schlüter A, Nordmann P, Bonnin RA, Millemann Y, Eikmeyer FG, Wibberg D, et al. IncH-Type Plasmid Harboring bla(CTX-M-15), bla(DHA-1), and qnrB4 Genes Recovered from Animal Isolates. Antimicrobial Agents and Chemotherapy. 2014 03/03/received 04/13/accepted;58(7):3768-73.
5. Struve C, Roe CC, Stegger M, Stahlhut SG, Hansen DS, Engelthaler DM, et al. Mapping the Evolution of Hypervirulent Klebsiella pneumoniae. mBio. 2015 September 1, 2015;6(4).
6. Hammond DS, Harris T, Bell J, Turnidge J, Giffard PM. Selection of SHV Extended-Spectrum-β-Lactamase-Dependent Cefotaxime and Ceftazidime Resistance in Klebsiella pneumoniae Requires a Plasmid-Borne blaSHV Gene. Antimicrobial Agents and Chemotherapy. 2008 February 1, 2008;52(2):441-5.
7. Ku Y-H, Chuang Y-C, Chen C-C, Lee M-F, Yang Y-C, Tang H-J, et al. Klebsiella pneumoniae Isolates from Meningitis: Epidemiology, Virulence and Antibiotic Resistance. Scientific Reports. 2017 07/2602/01/received 06/20/accepted;7:6634.
8. Gundestrup S, Struve C, Stahlhut SG, Hansen DS. First Case of Liver Abscess in Scandinavia Due to the International Hypervirulent Klebsiella Pneumoniae Clone ST23. The Open Microbiology Journal. 2014 03/07 11/24/received 12/18/revised 12/18/accepted;8:22-4.
9. Schjørring S, Struve C, Krogfelt KA. Transfer of antimicrobial resistance plasmids from Klebsiella pneumoniae to Escherichia coli in the mouse intestine. Journal of Antimicrobial Chemotherapy. 2008 08/13 04/25/received 06/11/rev-request 07/03/revised 07/11/accepted;62(5):1086-93.

4.2 FASTER RESTORATION OF THE MAJOR MICROBIAL TAXA IN THE HUMAN GUT WITH FAECAL MICROBIAL TRANSPLANTATION (FMT) AFTER TREATMENT WITH AN ANTIBIOTIC (MOXIFLOXACIN)

Basil Britto Xavier[1], Hadi Mazlom[3], Mohamed El Hadidi[4], Mohamed Mysara[1], Christine Lammens[1], Pieter Brems, Daniel Huson[4], Surbhi Malhotra-Kumar[1], Herman Goossens[1,3].

[1]Laboratory of Medical Microbiology, Vaccine & Infectious Disease Institute, Universiteit Antwerpen, Antwerp, Belgium; [3]University Hospital (UZA) Edegem, Antwerp, Belgium; [4]Department of bioinformatics, University of Tubingen, Tubingen, Germany.

Abstract

Use of broad-spectrum antibiotics has increased over recent years. However, little is known about their long-term impact on the microbial diversity of faecal flora. At the same time, FMT has become an increasingly popular treatment for *Clostridium difficile*-associated diarrhea, often the result of BS-AB treatment. We investigated the microbiota of three healthy male volunteers before and after one-week therapy with moxifloxacin. Intervention was uncontrolled as both case, and the 2 controls (MT, GJB) were in the home-setting with no dietary restrictions. The case was administered FMT (two capsules) for two days after 24-48hrs of completion of moxifloxacin course. FMT Stomach Acid Resistant Gel capsules (DRcaps™) were made from stools collected from case pre-antibiotic therapy. Fecal samples (n=58; 52+6) underwent quantitative culture on EMB and blood agar with and without moxifloxacin. Genomic DNA was prepared (QIAamp DNA Stool Mini Kit, Qiagen), sample and library preparation were done using Nextera XT, and 16S rDNA (V3-V4 region, MiSeq, Illumina) and shotgun-sequencing (HiSeq2500 Illumina, output ≈80 million reads/sample) were performed. Non-human reads were mapped against NCBI-NR database (DIAMOND; identity threshold of 97%; maximum e-value of 10e-5). For taxonomic classification, MEGAN 6 was used to compare normalized samples and create stacked bar plots and network analysis. In all three volunteers, samples taken during antibiotic treatment clustered distinctly from the baseline and the later post-antibiotic samples. Both 16S and shot-gut metagenomics revealed decreased microbial diversity during moxifloxacin treatment in all three volunteers compared to their baseline samples. However, microbial taxonomic changes observed during treatment were unique to each volunteer and dependent on the baseline taxa composition. Post-antibiotic, the microbiome of all three volunteers shifted towards the baseline and was closest to the baseline samples at

the last sampling point (day 23-31), especially for the case who had been administered FMT. We show here that moxifloxacin use alters the microbial diversity, although changes in taxa were unique and dependent on the individual's baseline pre-antibiotic microbiome. FMT facilitated an earlier recovery of the major taxa.

4.2.1 Introduction

The increased life expectancy of human life is one of the significant outcomes from the discovery of antibiotics. It helped humans to overcome life threating infections caused by bacteria. However, we never had realised intake of antibiotic would impact our healthy gut flora in long term. Since the beginning of a genomic era, studies shown the role of the gut microbiome and human health interlinked and it is protecting against colonising pathogens in the gut (25, 26). Antibiotics are not only target the pathogens and it also disturbs ecosystem of human microbiome which comprise of cells, genes, and metabolites from gut flora, and viruses and eukaryotes that inhabit in the host (25). Naturally, gut microbiome act as a reservoir for antimicrobial resistance genes (27, 28) but it depends on the host and environment. The use of broad-spectrum antibiotics such as Moxifloxacin has continued to increase over recent years. Moxifloxacin is a fourth-generation fluoroquinolone widely used against the respiratory and anaerobic pathogens (29, 30), and only a few studies analysed the impact of healthy flora but not on the global impact of microflora (31). Highly diverse microbiome indicator of healthy and lower inflammation (Lloyd-Price, et al. 2016; Menni, et al. 2017) however, this scenario is not same for all body sites. Gut microbiome greatly disturbed by antibiotics and the disturbed microbiome recovery to the original composition also quite varied in time(Lloyd-Price, et al. 2016). The broad spectrum antibiotic target non-specific and good bacteria as well. Restoring the and enhancing the disturbed microbiome by probiotics and faecal microbiome transplant (FMT) favorite techniques (Langdon, et al. 2016). At the same time, faecal transplantation has become an increasingly mostly popular treatment for Clostridium Difficile Associated Diarrhoea (CDAD) often as a result of treatment with broad-spectrum antibiotics. Interestingly only few studies shown with donor faecal transplants for this purpose but never with own faecal transplant. So to know the own FMT impact we utilised the state of art technique metagenomics to study the real impact and how FMT helped in recovery and this proof-of-concept study was also to see the effects of broad spectrum antibiotic (moxifloxacin) on gut microbiota, the duration of this disruption and if faecal transplantation with own faecal transplant was effective or not to recover original microflora.

4.2.2 Materials and Methods

The case and controls (MT and GJB) were identified and screened for standard medical screening including a physical examination and some questions about their medical history and allergies were recorded.

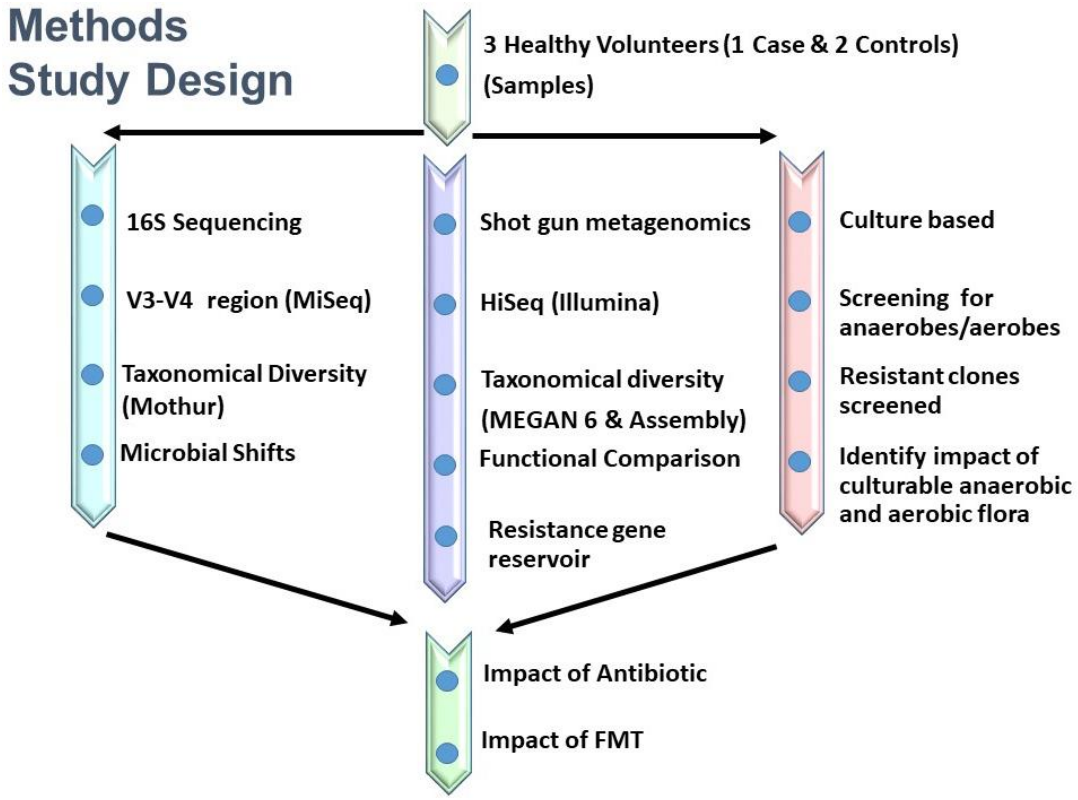


Figure 4.2.1 Overview of study design

For the case, his sample were prepared in oral capsules. Protocol for capsule preparation were modified from another research study (32). Briefly, a faecal suspension was generated in normal saline without preservatives using a blender. Materials were sequentially sieved to remove solid materials. The final slurry was concentrated by centrifugation and resuspended in saline at one-tenth the volume of the initial sample with 10% glycerol added as a bacterial cryoprotectant. Faecal matter solution was then pipetted in Stomach Acid Resistant Gel capsules (DRcaps™) http://www.capsugel.com/ihc/drcaps. Briefly, Capsules were frozen at -80°C. One to two hours before administration, they are transferred to −20°C, then transported to the volunteer (case) on dry ice. The inoculum was prepared to a total 50 grams of sample dependent on capsule size these were administered on two consecutive days (e.g. 20x 1gram capsules for two days) The case took moxifloxacin (400mg) once daily orally for one week

(**Figure 4.2.2**). The case took FMT capsules for two consecutive days after 6 days course of antibiotic completed. The capsules are containing his own faecal material collected before antibiotic exposure for two consecutive days (Figure 4.2.2). The controls (MT and GJB) took moxifloxacin without the faecal-pills. Interventions in this study were associated with risks and side-effects, and the local Ethical committee (UZA) approved the study prior to the start.

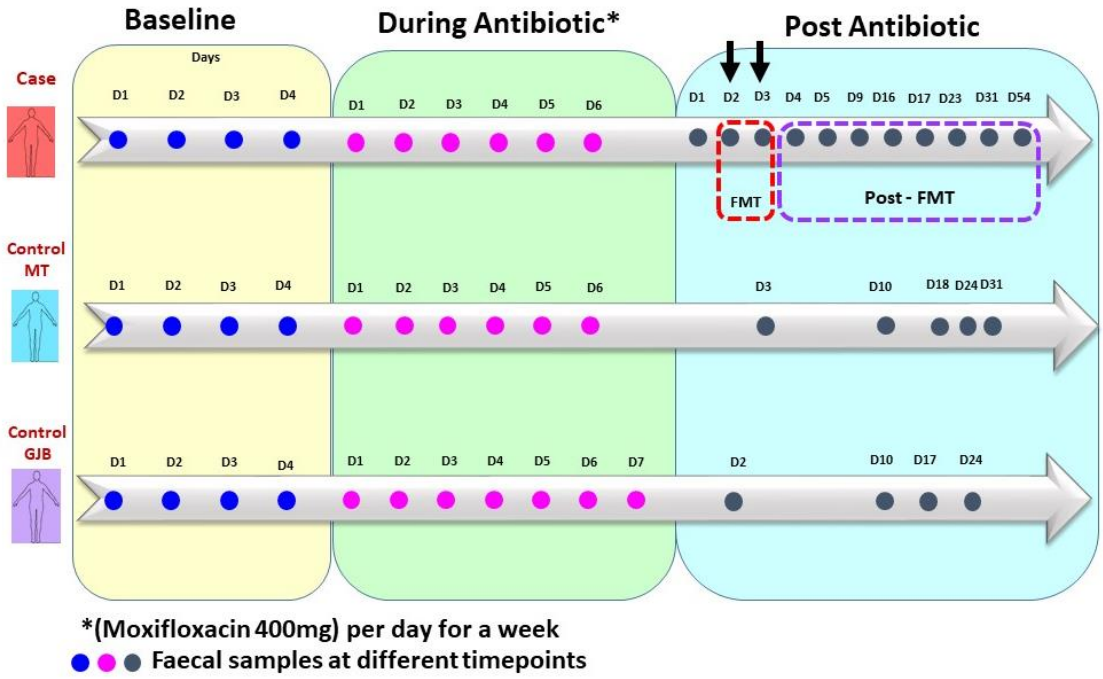


Figure 4.2.2. Sampling timeline of all three individuals

A culture-based screening of stool samples was done on selective media. Faecal samples were collected from case, MT and GJB at the different time points (Figure 4.2.2) Serial dilutions were performed and 100µl of stool suspension plated on blood agar (BA) plates with Eddy Jet (Program 7=log 100µl). Plates were incubated in the anaerobic and aerobic incubator respectively for 24-48h. As part of a study investigated the emergence of antibiotic-resistant Enterobacteriaceae in the faecal flora. Faecal samples (n=58 up to day 122) were collected from all three volunteers and underwent culture on eosin methylene blue (EMB).

The genomic DNA preparation was done using QIAamp DNA Stool Mini Kit (Qiagen) for metagenomics. For shotgun metagenomics sequencing, sample and library prepared using Nextera XT kit, sequenced using HiSeq2500 V4 high output mode 2 x 100 bp (Illumina Inc.,) with output of approximately. 80 million reads per sample the quality of the reads Q30 value:

92.98% and simultaneously 16S library prepared and V3-V4 region was selected, and amplicon sequencing was done using MiSeq.

For 16S rRNA analysis by checking the quality of fastq files using Fastqc. Secondly, sequencing errors were removed using an IPED algorithm - dedicated to denoise MiSeq amplicon sequencing data (33). Chimera was detected using the de novo mode of CATCh algorithm (34). Creation of the operational taxonomic units (OTUs) was performed using UPARSE (35) with default parameters. For beta diversity, the distances between the samples were calculated which was later represented using heat map visualisation (via mothur command) and) plots (via mothur nmds command). For each nmds (non-metric multidimensional scaling) plot, the OTUs that shows statistically significant correlation with the coordinates were indicated ($p < 0.05$, correlation >70%).

Similarily, shot-gun meta-genome sequences were processed, demultiplexed using Illumina CASAVA (1.8.2), and adapters trimming with a skewer (v0.1.116) (36) For better result interpretation, the read depth was enhanced by concatenating paired reads of each sample into a single FASTQ file to allow detection of antibiotic resistance genes. Since the overall quality of reads is high (> 35), no trimming or filtering was applied to contamination with host DNA host and microbial reads were sequenced alongside. We used SASS (Version 0.3.2) to remove reads matching a human sequence using an in-house developed aligner based on Smith-Waterman aligner. The human genome assembly, GRCh37, was downloaded from NCBI and used for SASS read mapping.

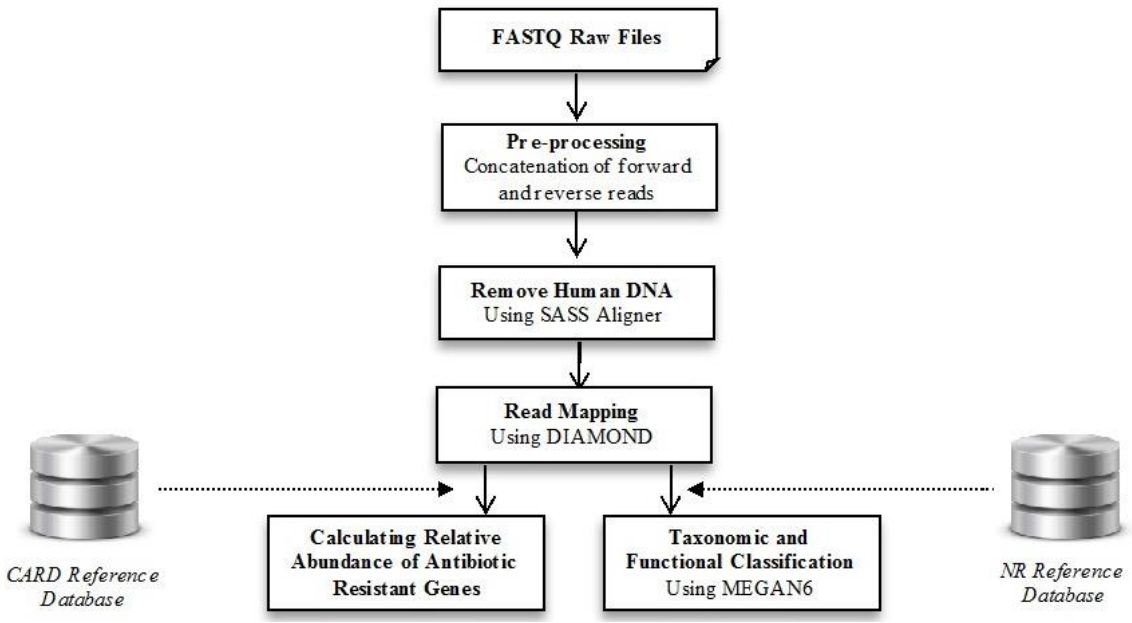


Figure 4.2.3 Flow diagram of read processing and detecting relative abundance of resistance genes in the samples analysed.

Reads were mapped against NCBI NR database (accessed 31 December 2014) using DIAMOND in blastx mode, with an identity threshold of 97% and maximum e-value of 10e-5 for taxonomic classification. DAA files generated by DIAMOND were converted into RMA files to be opened by MEGAN6. For taxonomical classification, MEGAN 6 Ultimate edition was used to compare normalised samples, create stacked bar plots and calculate PCoA plots. For calculations of the relative abundance of antibiotic resistance genes, reads were mapped against CARD database using DIAMOND in blastx mode, with an identity threshold of 97%. The union of all detected antibiotic resistance genes were defined from all samples for each of case, MT and GJB. Using self-scripts, the number of reads matched to each CARD antibiotic resistance genes and the relative abundance of reads were calculated (using ωc method) as described (37). For taxonomical classification, MEGAN 6 Ultimate edition was used to compare normalised samples, create stacked bar plots, and calculate PCoA plots and network distance tree (38). Data shall be analysed using statistical software SPSS.

4.2.3 Results

4.2.3.1 Culture results

From the culture data, lots of variations in baseline samples for all three volunteers were observed. The anaerobic flora was mainly targeted by moxifloxacin compared to the aerobic flora. Control MT samples showed a high-level resistant flora compared to other volunteers. Also, samples from the culture data during the antibiotic intake period showed selective advantage for growth of molds such as *Candida.*

4.2.3.2 16S rRNA & shot gun sequencing results

Upon analysing the alpha diversity, a significant reduction in diversity was reported with the introduction of the antibiotic (up to 50% of the baseline diversity), this was regained with the samples further away from the end of the antibiotic period. For the case, we report an instant increase of diversity in the samples post-FMT (sample 2 & 3 post-antibiotic).

When assessing the beta diversity among the samples, we report from the heat map the following remarks: a) The ampoule samples are closest to the case baseline and the following samples after antibiotic compared to the samples during and just after the antibiotic. b) The heat map illustrates the gradual increase in the distances between Baseline and during antibiotic samples. c) This was reverse gradually post-antibiotic (where the distances between the baseline samples became smaller with samples on late day post-antibiotic). From the analysis, we can

see that the different samples shift towards the similar community as a statistically significant reduction of OTUs belonging to *Bifidobacteriaceae, Ruminococcaceae, Streptococcaceae, Peptostreptococcaceae Peptostreptococcaceae* and *Lachnospiraceae, Rikenellaceae, Coriobacteriaceae and Porphyromonadaceae*. Additionally, the case samples post-antibiotic has a faster recovery to the original baseline community compared to the other controls, as the distance between the baseline samples and the day 5 in the case is smaller than the other two controls.

Upon clustering each sample into metacommunities, a various number of clusters evolved for each. In both controls, the baseline samples and the first two samples during moxifloxacin treatment were clustered together (i.e. similar community). This has changed with the third day of antibiotic treatment, where a shift towards one and two dysbiotic communities occurred for MT and GJB control respectively. The microbial composition (on average over the samples within each cluster) revealed an elevated level of OTUs affiliated to *Blautia, Roseburia,* and *Akkermansia* genus in case of MT control, while in GJB together with *Blautia* and *Roseburia*, various OTUs classified as *Bacteroides* were elevated. On the other hand, reduction in OTUs classified as *Lachnospiraceae, Faecalibacterium* and *Blautia*. In both scenarios, the microbial communities shift back the initial composition at day 9 sample post-antibiotic. While for the case, the baseline samples were clustered into two groups (differing mainly in abundances of various OTUs belonging to *Prevotella*, *Roseburia* and *Faecalibacterium* genera). With start of antibiotic, two distinct clusters of microbial communities were reported, with an elevated level of *Ruminococcus, Faecalibacterium*, and *Prevotella*). After the addition of FMT, an elevated level of OTUs classified as *Akkermansia*, *Phascolarctobacterium*, *Faecalibacterium* and *Ruminococcaceae*, the was latterly normalised to the initial two clusters of the baseline starting from day 7 sample (Figure 4.2.3 and 4.2.4). It is essential, however, to emphasise that during the case antibiotic treatment, we report for the third sample an attempt of the microbial community to resolve the dysbiosis raised by the use of moxifloxacin.

4.2.3.3 Baseline diversity of healthy volunteers

In order to know the baseline composition of all three healthy volunteers we had first analysed baseline faecal samples sequences, and compared, that indicated that all three healthy volunteers have unique microbial composition. As previously reported, the microbial composition was correlated well with dietary habits. All three individuals had *Bacteroides*, as a common genus. The secondary predominant genus was different i.e. case samples showing

predominant *Prevotella, Alistipes, Roseburia* ii. Contro-GJB: *Faecalibacterium, Clostridium, Ruminococcus*, iii. Control-MT: *Dialister, Ruminococcus.* (Figure 4.2.4)

In case presence of sulfate-reducing bacteria (SRB) such as *Desulfovibrio piger* was impacted due to the antibiotic treatment and post-antibiotic this bacterium is not recovered and its higher abundance associated with gastrointestinal diseases (inflammatory bowel diseases) (39). As SRB can colonise the approximately 50% of human guts and this bacterium fitness depends on the diet of the individual (40). Other gram-negative anaerobic bacteria *Butyricimonas virosa* was impacted by antibiotic treatment almost 35% reduced another anaerobic gram-positive rod which was found in normal healthy individual faeces Eubacterium rectangle during the treatment period (41). Another commensal bacterium *Roseburia inulinivorans* is an anaerobic carbohydrate utilising bacterium found in the human colon was impacted during the antibiotic period and but it came back quickly this might be might be associated with dietary conditions (42). *Clostridium leptum* group is one of the dominant group presents in the faecal of microbiota humans and its consist of *Faecalibacterium prausnitzii* and certain species of Eubacterium and Ruminococcus and all members of this group playing a role as butyrate producers (43). Also, other anaerobes including *Alistipes* and *Prevotella* were significantly more abundant in the gut microbiota of healthy subjects.

In Control-MT *Adlercreutzia equolifaciens* wholly lost after the antibiotic treatment which play an essential role in converting isoflavones to equol in human intestinal (44).The human faeces predominant phage crAssphage hosted by Bacteroides is completely wiped out by control-MT after the antibiotic treatment at the same time crAssphage increased in after treatment (45) but the same phage appeared on the post-antibiotic samples of Case, suggesting that antibiotic selection pressure have a role in the spread. In control-GJB, the *E. coli* disappeared during the antibiotic and not recovered until the post-antibiotic day 31.

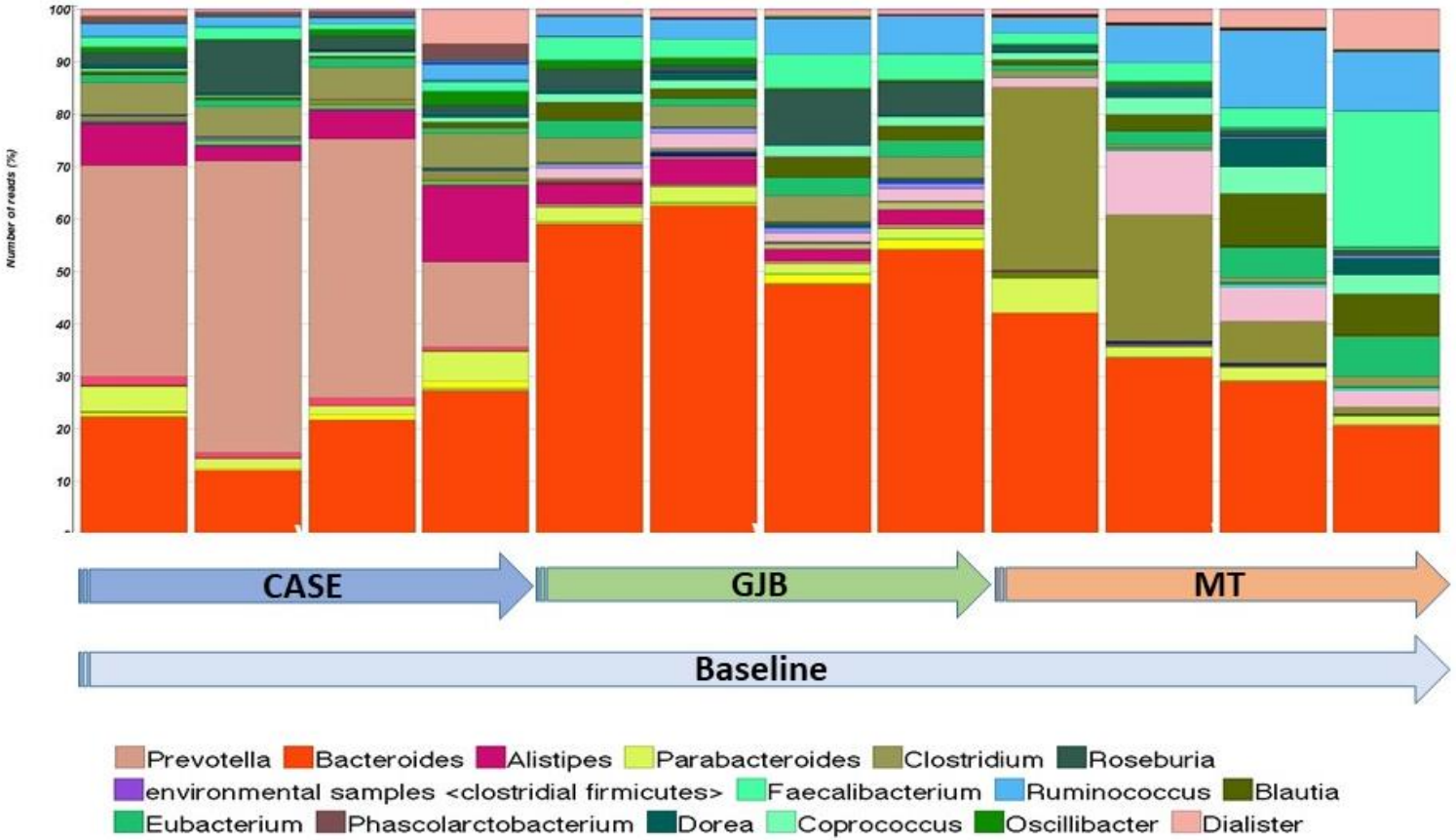


Figure 4.2.4. Baseline diversity of all three healthy volunteers

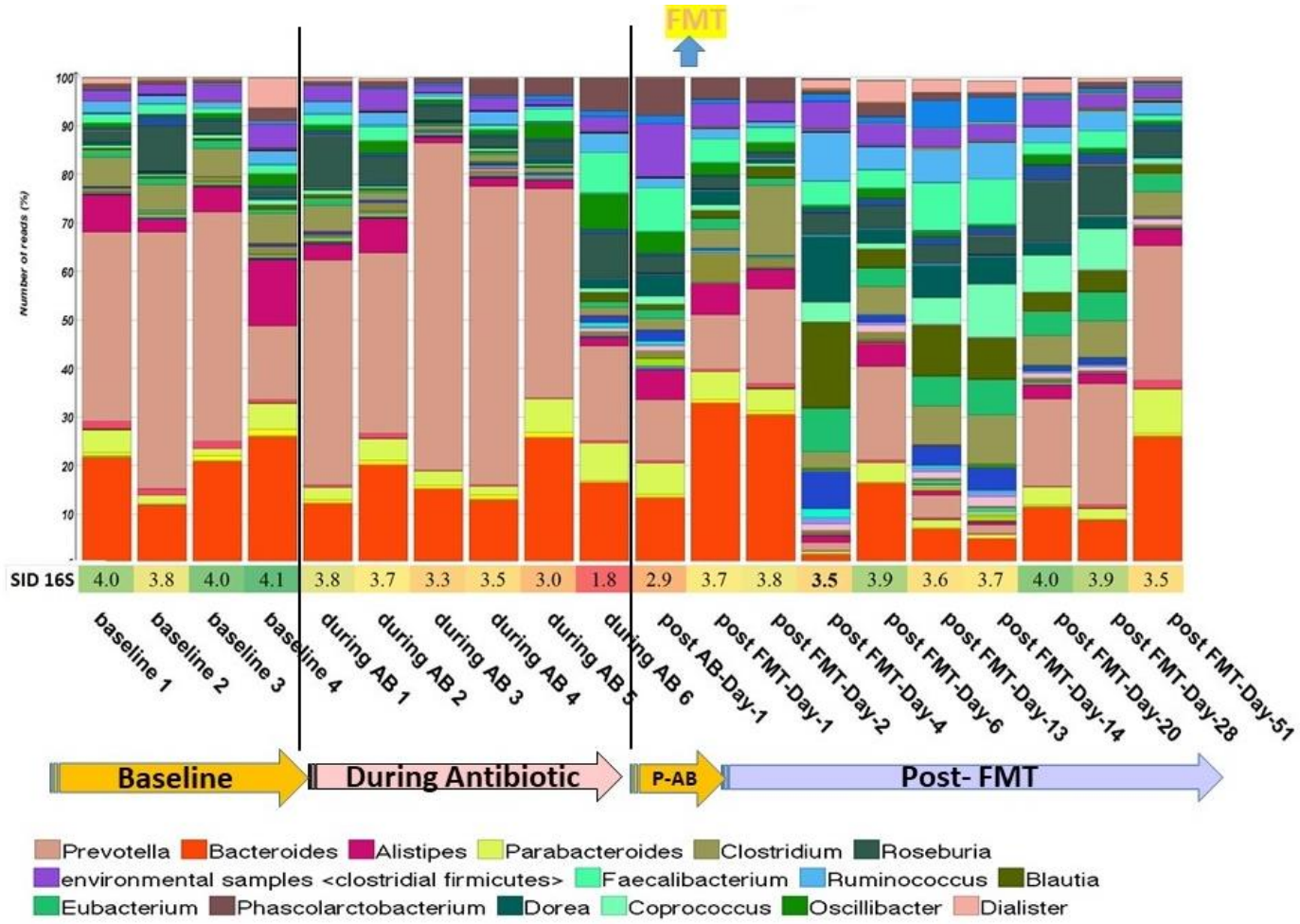


Figure 4.2.5. Case taxonomical diversity of bacterial composition of pre-during and post-antibiotic intake.

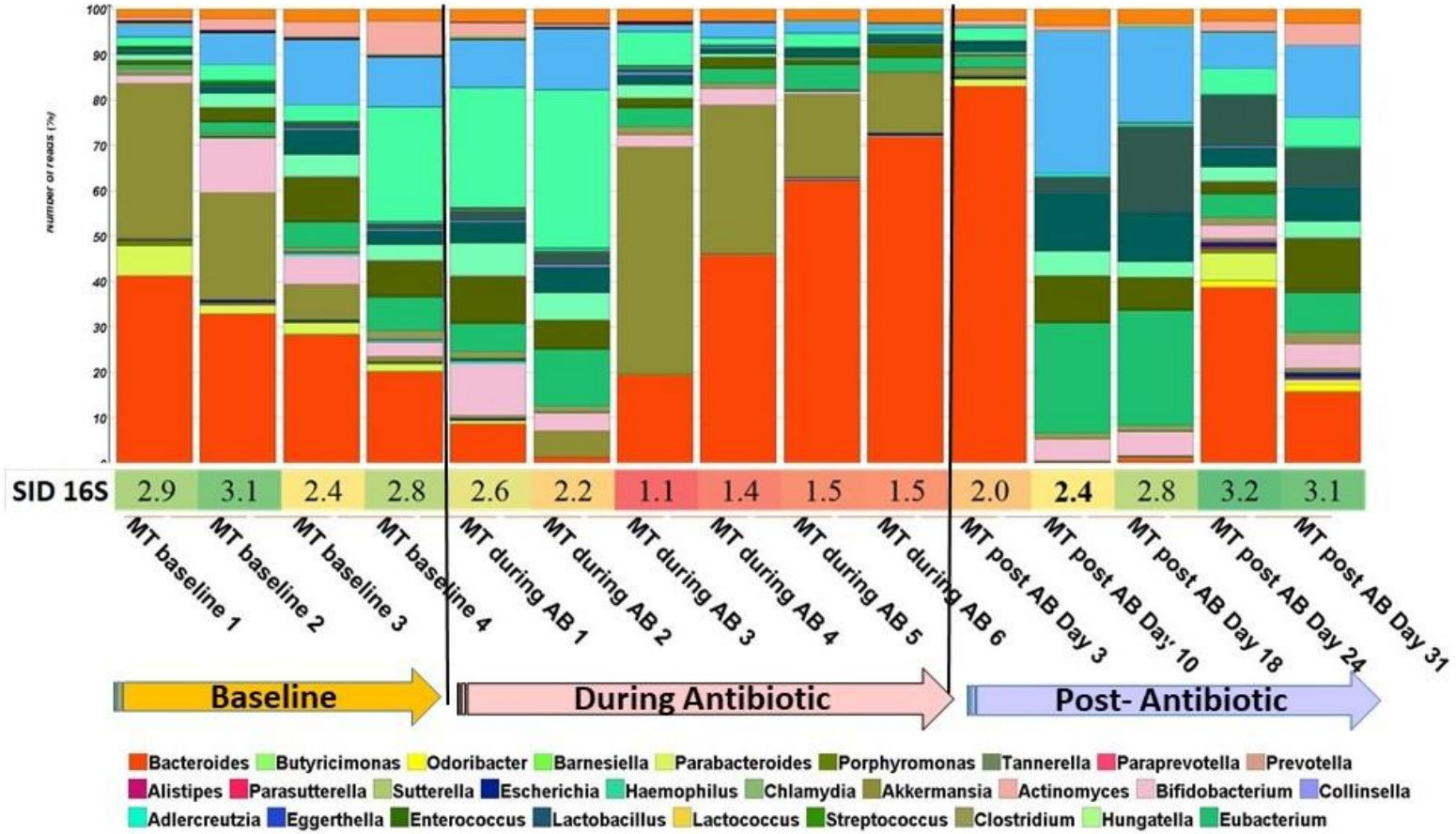


Figure 4.2.6. Control MT-Taxonomical diversity of bacterial composition of pre-during and post-antibiotic intake

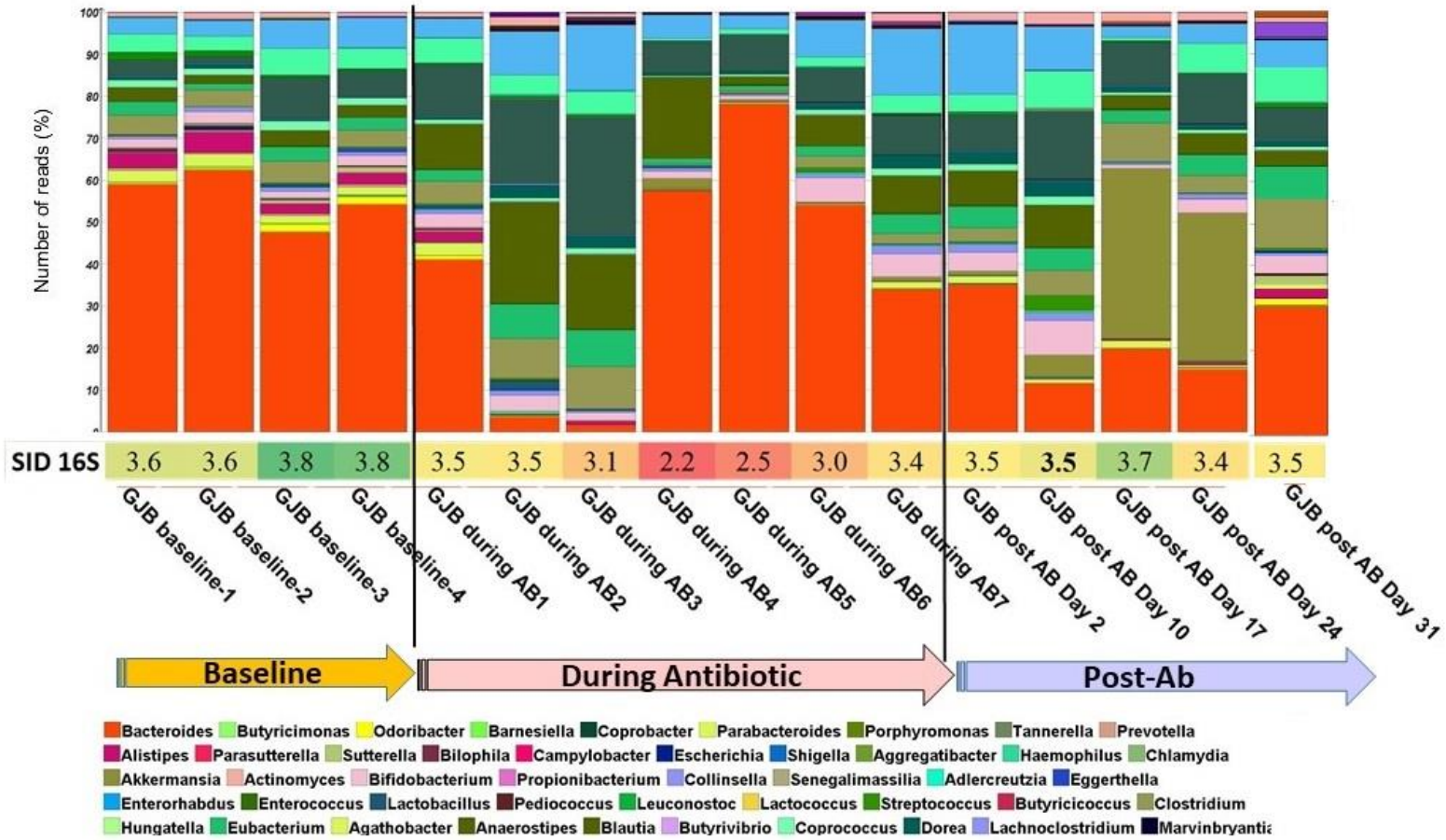


Figure 4.2.7 Control GJB-Taxonomical diversity of bacterial composition of pre-during and post-antibiotic intake

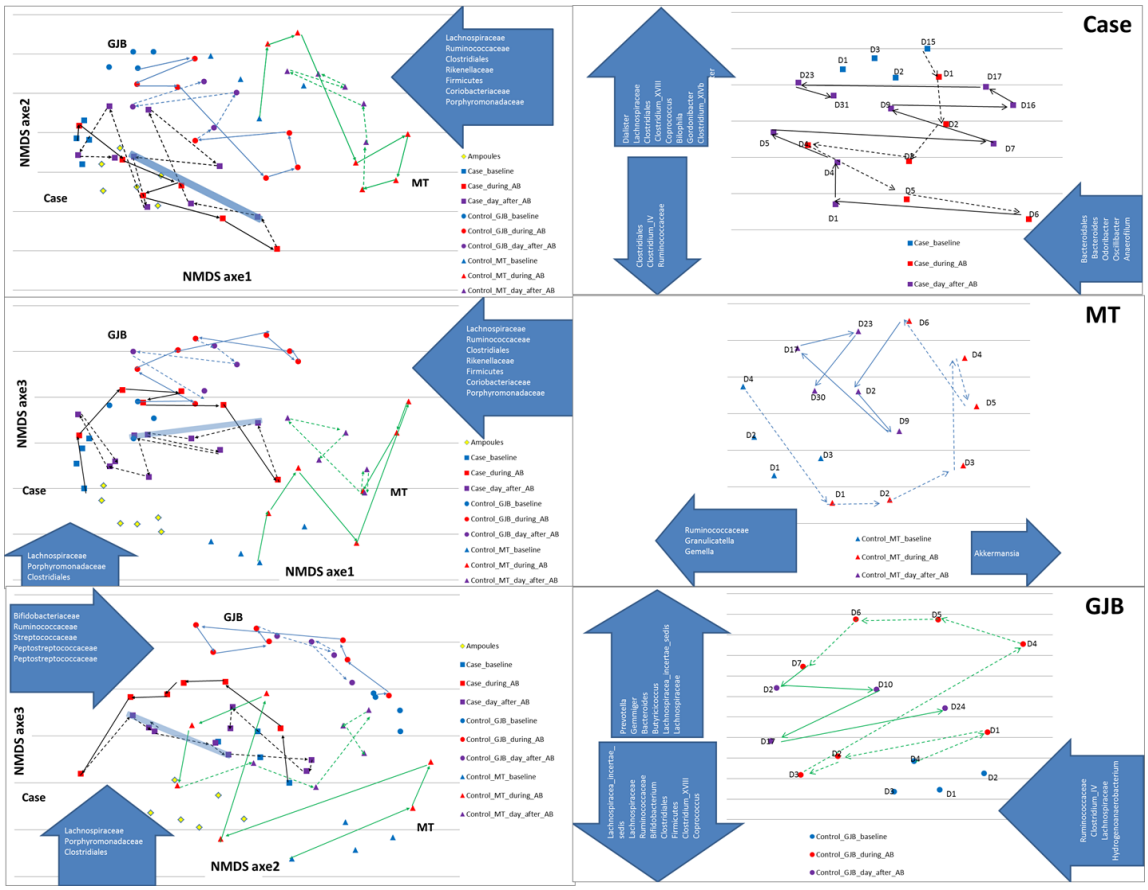


Figure 4.2.8 The left side of the figure shows a 3D nmds plot of the samples of all three-individual combined, as following the left top (axe1 vs axe2), left middle (axe1 vs axe3), and left bottom (axe2 vs axe3). While the right side of the figure illustrates the needs 2D for all samples within each as follows: right top (case), right middle (MT) and right bottom (GJB). The highlighted case samples illustrate the samples equivalent to those of the other two controls (time-wise).

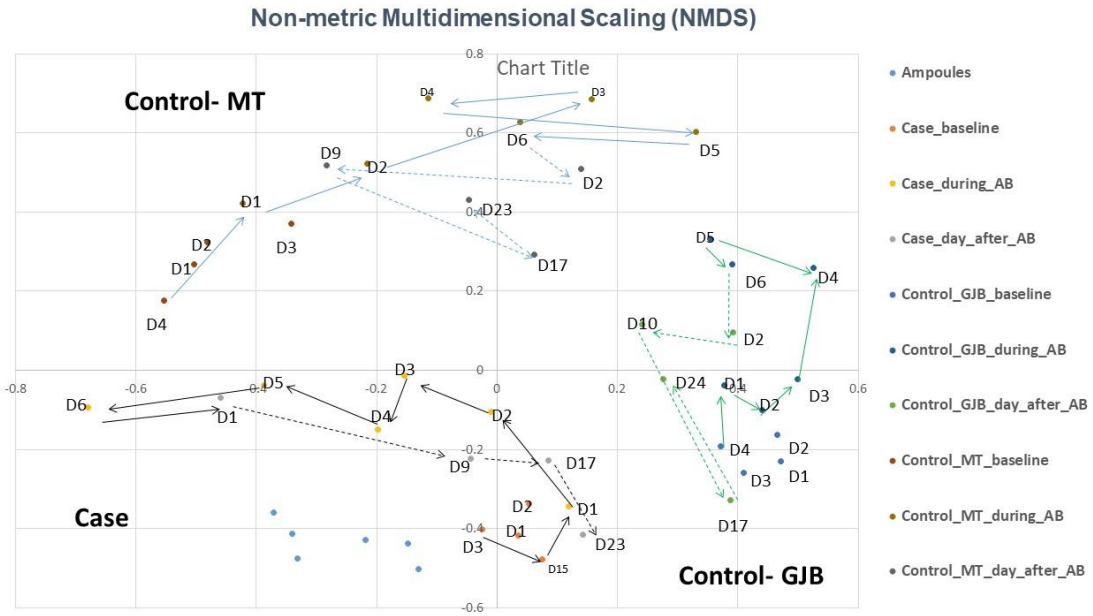
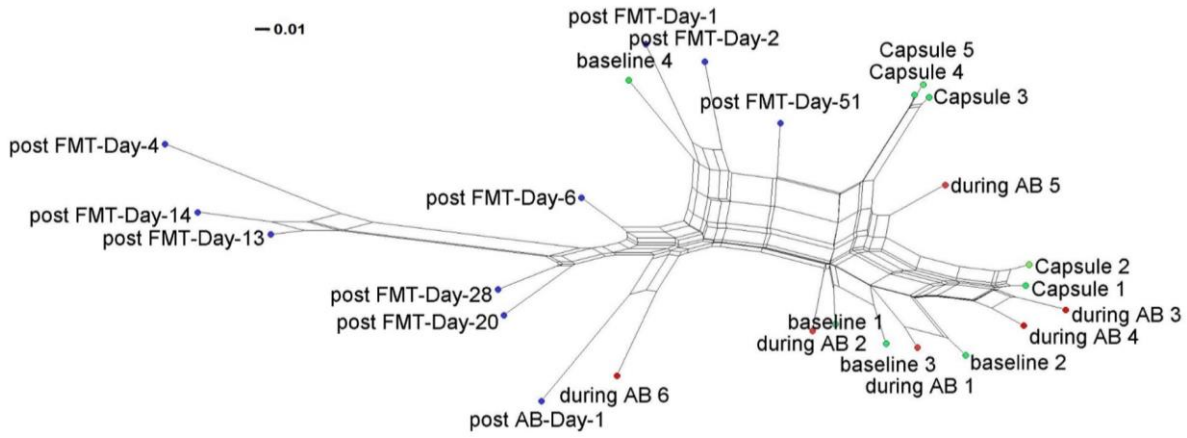


Figure 4.2.9 Non-metric Multidimensional Scaling (NMDS) of smaples

Table 4.2.1 The reservoir genes of all healthy individuals.

Baseline	Beta-lactams	Aminoglycoside	Tetracycline	Macrolide
Case	cfxA, cfxA6	aadE, aph3-III, strA, strB,	tet32, tet40, tet44, tetQ, tetW	ermF, ermB, ermG, lnuC
Control MT	blaOXA-347, blaTEM-1B, cfxA5	aadE, aph(3')-III, ant(6)-Ia, strB	tet40, tetM, tetO, tetQ, tetW	ermF, ermB, ermG, lnuC
Control GJB	cfxA	aadE	tet32, tetO, tetW, tetX	ermF, ermB,
During-Antibiotic				
Case	cfxA, cfxA6	aadE, aph(3')-III, aph(2')-Ib, aac(6')-Im, strA, strB	tet32, tet44, tetM, tetQ, tetW	ermF, ermB, ermG, lnuC
Control MT	blaOXA-347, blaTEM-1B, cfxA5	aadE, aph(3')-III, ant(6)-Ia , strA, strB	tet40, tetO, tetQ, tetM, tetW	ermF, ermB, ermG, lnuC
Control GJB	cfxA	aadE	tet32, tetO, tetW, tetX	ermF, ermB,
Post-Antibiotic				
Case	cfxA, cfxA6	aadE,aph(3')-III, strA, strB	tet32, tet44, tetM, tetQ, tetW, tetX	ermF, ermB, ermG,
Control MT	blaOXA-347, blaTEM-1B, cfxA5,	aadE, aph(3')-III, ant(6)-Ia, strA, strB	tet40, tetM, tetO, tetQ, tetW	ermF
Control GJB	cfxA	aadE	tet32, tetO, tetW, tetX	ermF



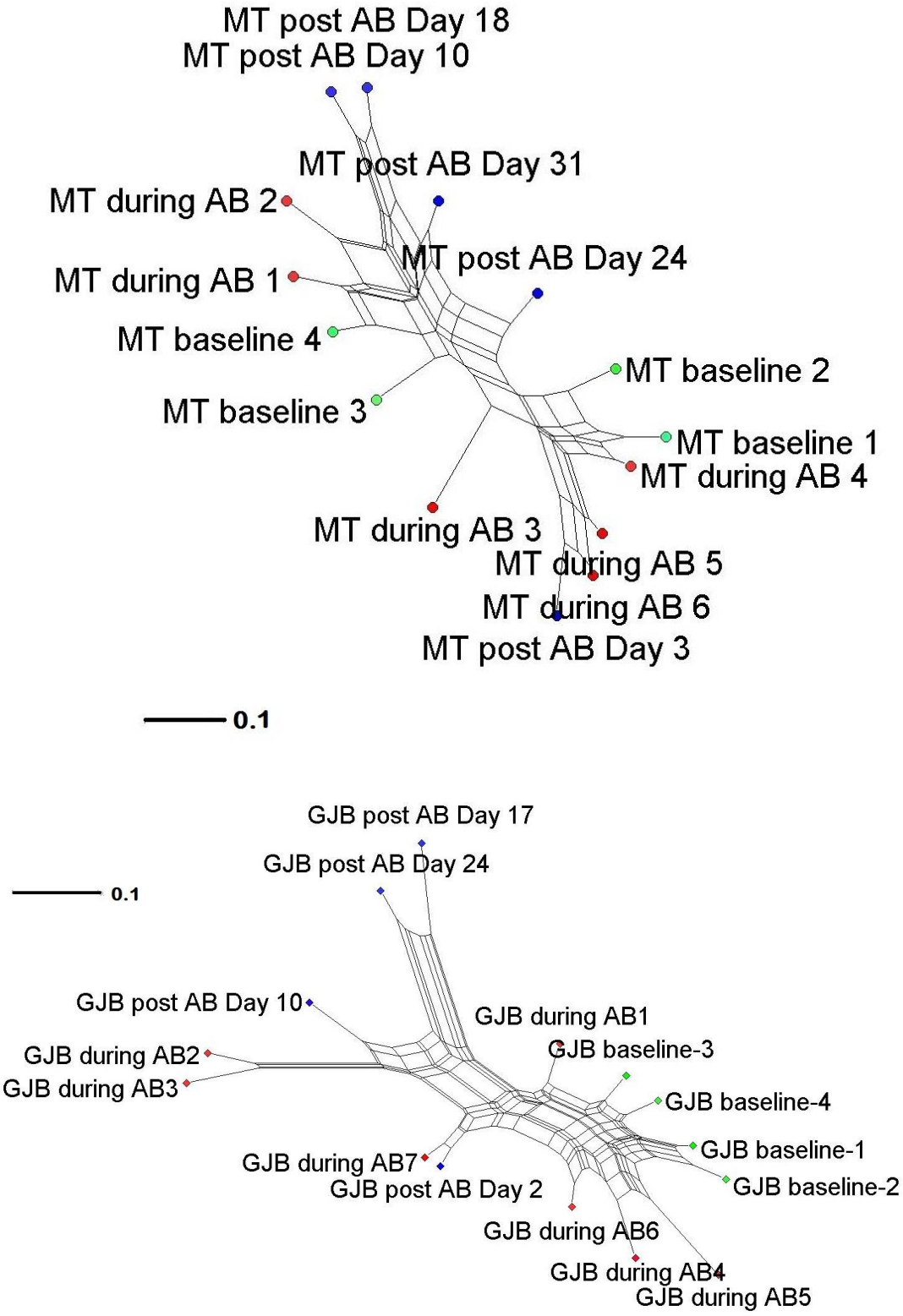


Figure 4.2.10 Network analysis of all three volunteers' samples collected from pre during and post antibiotic intake

4.2.4 Conclusions

The diversity dropped to 50% with moxifloxacin, which regained afterwards (**Figure 4.2.5, 4.2.6, 4.2.7 and 4.2.9**). The effect of moxifloxacin on the gut original microbial composition was reversed after ten days' post-antibiotic intake (**Figure 4.2.4**). The microbial composition was normally recovered with around three weeks' post-antibiotic intake (**Figure 4.2.4**). FMT has been found to show a faster recovery to the first same community (see **Figure 4.2.5 & 4.2.9**). Very importantly, *Parabacteriodes* is beneficial bacteria as it will help in digesting high rich fibers and protecting our guts from inflammation have lost in all three individuals during their antibiotic intake period. In conclusion, the study on small sample size showed that Moxifloxacin treatment did not seem to have any permanent impact on the gastrointestinal flora on long-term instead it had minimal impact on specific aerobic and anaerobic flora in the short term. The functional analysis identified in all healthy individual's gut also act as a reservoir for resistance genes **(Table 4.2.1)**.

Author contribution: BBX was involved in experimental work, data analysis, interpretation and drafting manuscript.

4.2.5 References

1 Jiang **15**, 182-182, doi:10.1186/1471-2105-15-182 (2014)., 182-182, doi:10.1186/1471-2105-15-182 (2014)., 182-182, doi:10.1186/1471-2105-15-182 (2014).

2 Loubinoux, J., Bronowicki, J.-P., Pereira, I. A. C., Mougenel, J.-L. & Le Faou, A. E. Sulfate-reducing bacteria in human feces and their association with inflammatory bowel diseases. *FEMS Microbiology Ecology* **40**, 107-112, doi:10.1111/j.1574-6941. 2002.tb00942.x (2002).

3 Rey, F. E. *et al.* Metabolic niche of a prominent sulfate-reducing human gut bacterium. *Proceedings of the National Academy of Sciences* **110**, 13582-13587, doi:10.1073/pnas.1312524110 (2013).

4 Sghir, A. *et al.* Quantification of Bacterial Groups within Human Fecal Flora by Oligonucleotide Probe Hybridization. *Applied and Environmental Microbiology* **66**, 2263-2266 (2000).

5 Scott, K. P., Martin, J. C., Campbell, G., Mayer, C.-D. & Flint, H. J. Whole-Genome Transcription Profiling Reveals Genes Up-Regulated by Growth on Fucose in the Human Gut Bacterium "Roseburia inulinivorans". *Journal of Bacteriology* **188**, 4340-4349, doi:10.1128/jb.00137-06 (2006).

6 Kabeerdoss, J., Sankaran, V., Pugazhendhi, S. & Ramakrishna, B. S. Clostridium leptum group bacteria abundance and diversity in the fecal microbiota of patients with inflammatory bowel disease: a case–control study in India. *BMC Gastroenterology* **13**, 20-20, doi:10.1186/1471-230X-13-20 (2013).

7 Yuan, J. P., Wang, J. H. & Liu, X. Metabolism of dietary soy isoflavones to equol by human intestinal microflora - implications for health. *Mol Nutr Food Res* **51**, 765-781, doi:10.1002/mnfr.200600262 (2007).

8 Dutilh, B. E. *et al.* A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat Commun* **5**, doi:Artn 44980.1038/Ncomms5498 (2014).

9 Loubinoux, J., Bronowicki, J.-P., Pereira, I. A. C., Mougenel, J.-L. & Le Faou, A. E. Sulfate-reducing bacteria in human feces and their association with inflammatory bowel diseases. *FEMS Microbiology Ecology* **40**, 107-112, doi:10.1111/j.1574-6941. 2002.tb00942.x (2002).

10 Rey, F. E. *et al.* Metabolic niche of a prominent sulfate-reducing human gut bacterium. *Proceedings of the National Academy of Sciences* **110**, 13582-13587, doi:10.1073/pnas.1312524110 (2013).

11 Sghir, A. *et al.* Quantification of Bacterial Groups within Human Fecal Flora by Oligonucleotide Probe Hybridization. *Applied and Environmental Microbiology* **66**, 2263-2266 (2000).

12 Scott, K. P., Martin, J. C., Campbell, G., Mayer, C.-D. & Flint, H. J. Whole-Genome Transcription Profiling Reveals Genes Up-Regulated by Growth on Fucose in the Human Gut Bacterium "Roseburia inulinivorans". *Journal of Bacteriology* **188**, 4340-4349, doi:10.1128/jb.00137-06 (2006).

13 Kabeerdoss, J., Sankaran, V., Pugazhendhi, S. & Ramakrishna, B. S. Clostridium leptum group bacteria abundance and diversity in the fecal microbiota of patients with inflammatory bowel disease: a case–control study in India. *BMC Gastroenterology* **13**, 20-20, doi:10.1186/1471-230X-13-20 (2013).
14 Yuan, J. P., Wang, J. H. & Liu, X. Metabolism of dietary soy isoflavones to equol by human intestinal microflora - implications for health. *Mol Nutr Food Res* **51**, 765-781, doi:10.1002/mnfr.200600262 (2007).
15 Dutilh, B. E. *et al.* A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat Commun* **5**, doi: Artn 4498 10.1038/Ncomms5498 (2014).

CHAPTER V

SUMMARY AND PERSPECTIVES

5.1 Summary

Antibiotic resistance has developed into a global pandemic threat to human health. The inherent adaptability of bacteria coupled with the ever-increasing diversity in environments, attributable for a large part due to rampant and diverse antibiotic use, has made it imperative to more closely understand the dynamics of multi-drug resistant bacteria as a first step in controlling antibiotic resistance. The introduction of next-generation sequencing technology (NGS) in microbiology has been the most revolutionary change since the introduction of PCR. NGS offers many ways to predict and possibly offer exceptional solutions to prevent antibiotic resistance. Thus, the primary goal of this thesis was to demonstrate the use of NGS technologies to study antibiotic resistance in important human pathogens.

In the introductory chapter (**Chapter 1**), we have summarised the global burden of antibiotic resistance and the molecular methods currently used for its detection. Then we focused on the specific mechanisms underlying the development of colistin resistance in *Enterobacteriaceae* and the use of NGS, its technology and advantages for detecting antibiotic resistance. This chapter ends with the general and specific aims of the thesis.

One of the major hurdles in the application of NGS for detection of antibiotic resistance (AR) has been the lack of comprehensive and up-to-date resistance gene databases. In **Chapter 2**, we review the existing antibiotic resistance gene databases as well as their limitations. We then introduce BacPipe, a whole genome 'automated' analysis pipeline developed by us. Lastly, we discuss the utility of whole genome mapping to study bacterial genome stability and its importance as a tool to guide second-generation sequencing data analysis.

Our review of the current state of antibiotic resistance gene databases that are freely available found these to be of substantial value (**Subchapter 2.1)**. Their limitations were related to the frequency of updates, and functionality and comprehensiveness of the resources. Broader coverage, consistent gene nomenclature, and centralised (unifying soil, human and other microbiome/resistome data) up-to-date records are crucial for tracking and identifying novel mutations and genes conferring AR. As efforts to coordinate pathogen genomics at the international level emerge, we hope our study serves as a platform to bring more attention to sustaining and extending AR gene data resources as a critical component of our response to tackling multi-drug resistant bacteria (MDR) and its spread.

Subchapter 2.2 describes an automated whole genome analysis pipeline-BacPipe, targeted to clinical microbiologists and clinicians, as it requires little or no bioinformatics

skills. We have demonstrated its robustness in handling diverse genomes of clinically essential pathogens characterised by different sizes, GC content, and presence of repeat regions that are challenging for downstream data analysis. Along with being comprehensive and modular, BacPipe has the advantage of being computationally low-resource, and the pipeline functionality does not require an internet connection or high-end computers. A simple graphical interface makes it very user-friendly; a user can easily direct the software to input and output folder locations, specify the tools to be included in the analysis and adjust the database/parameters from a drop-down list or buttons. BacPipe can run with raw reads from various sequencing platforms and start or pick-up the analysis from any step throughout the workflow giving tremendous flexibility. Using raw data or contigs from prior publications that have used many different methods to delineate hospital or community-based outbreaks or pathogen transmission, we have demonstrated that the collection of tools in BacPipe could reproduce the entire analyses as a 'one-stop' platform within a few hours. We believe this adequately automated pipeline will overcome one of the primary barriers to analysing and interpreting whole genome sequencing (WGS) data, facilitating applications for routine patient care in hospitals, public health and infection control monitoring. Future development of BacPipe entails the expansion of tools to enable identification of prophages, IS (insertion sequence) elements, CRISPR-Cas elements and, depending upon their open-access status, whole-/core-/pan-genome MLST schemes.

Nonetheless, BacPipe is limited by the quality of the currently available bioinformatics tools. Although, multiple bioinformatics tools are available for data analysis, we encountered many limitations. For instance, de *novo* sequence assembly tools that we utilised, Velvet and SPAdes, resulted in misassembled contigs primarily due to the multiple ribosomal DNA copies being assembled into a single contig. Transposons and plasmids were other problematic regions for assembly tools probably due to the genetically similar regions in these mobile elements. A solution to overcome these issues is to perform long read sequencing and hybrid assembly, which is of course associated with higher sequencing costs. A cheaper alternative could be using short-read sequencing and the relevant assembler, sorting the reads into specific groups and running the independent mini assemblies and merging them into the final scaffolds. Another very important problem is artificial introduction of bases in the scaffolds or contigs by the assembler. This makes variant calling unreliable and negatively impacts all downstream processes such as pathogen transmission predictions or clonal diversity and phylogeny analyses. This shortcoming can be overcome by performing reference mapping of the raw reads

against *de novo* scaffolds of the same strain with maximum identity and by extracting the consensus scaffolds.

Thus, current assembly tools need to be improved to give an overall assembly quality assessment score for every assembly produced, for e.g., >70-80% scaffolds are reliable and free from misassemblies. Another interesting observation we made during our analysis was that *de novo* assemblies generated using *de Bruijn* graphs were always better compared to the overlapping consensus method utilised by for instance by the CLC Genomics workbench (clcbio, Denmark).

Other limitations observed were with genome annotation tools. Firstly, there were no consensus of gene predictions between the tools Prokka, prodigal and others. Secondly, the tool either took a considerable amount of time and space to produce reasonable functional information (Prokka) or it simply produced basic predictions of open reading frames with less or no functional information's (prodigal). Thirdly, we noticed that different tools annotated the same gene differently and this was observed even with core genes. This problem needs to be addressed at a global level to enable development of a common annotation database or platform that can be followed or adopted by all tools. Most of the annotation tools lacked the possibility of small RNA predictions that are rapidly gaining importance in terms of regulators of structural genes.

In **subchapter 2.3,** we describe the application of Whole Genome Mapping (WGM) to track and characterise multi-drug resistant pathogens as part of a European Public Health initiative (EUPHi). WGM is a fluorescence-based high-resolution restriction mapping technique that provides an essential link between PFGE and WGS. It has been used for strain typing during an outbreak and can identify genome-wide variations. WGM also serves as an important assembly validation tool for WGS-assembled contigs, primarily arising from short read sequencing technology such as Illumina. Compared to PFGE, it generates approximately 500 bands and orders them as in the chromosome (*in situ*). We show its utility to study the genomic stability of reference *S. aureus* strains where WGM correctly detected a large genomic deletion in one of the twelve strains. Subsequently, WGS of the altered genome and inspection of the region identified the exact size, genomic location and the nature of this genomic deletion. WGM can thus be employed to rapidly assess the genomic stability of strains over time during storage or after inter-laboratory transport followed by identification of genomic changes by WGS.

Subchapter 2.4 describes the use of WGM for the optimal *de novo* assembly of bacterial genomes. Genome assembly based on *de Bruijn* graphs potentially yields mis-

assemblies when assembly statistics such as length of the contigs and N50 are considered. However, WGM acts as a powerful tool to identify such mis-assemblies and provides the optimal *k mer* sizes that can be used to produce optimally assembled genomes from short read sequencing technologies. Despite the extra cost, the scientific need for error-free and complete genomes makes WGM an indispensable technique during the process of genome assembly and its validation.

The use of WGM to track and detect an outbreak due to multi-drug resistant *K. pneumoniae* in a Greek hospital is described in **Subchapter 2.5.** The strains from the outbreak were characterised using PFGE, WGM, and WGS. This study demonstrated the application of WGM for studying the epidemiology of hospital-associated *K. pneumoniae*. Additionally, a combination of WGM and sequencing provided novel insights on the genomic features of the multi-drug resistant ST383 *K. pneumoniae* clone that is rapidly gaining importance due to its prevalence and clinical significance in Europe.

Subchapter 2.6 details the molecular typing tools that have been used to compare *A. baumannii* strains isolated from hospitals from Greece and India. We have characterised the clonality of *A. baumannii* strains isolated from different hospitals in Greece and one hospital in India. The strains were compared with different typing methods such as PFGE and MLST. The results were compared to high-resolution WGM, and the strains were discriminated at different levels. Comparison of the high-resolution typing technique WGM to other typing methods (MLST and PFGE) showed its superior discriminatory power for the analysis of rapidly evolving hospital-associated *A. baumannii* global clones. The high diversity among the Indian strains studied here was noteworthy.

While WGM has proven to be an essential tool at various levels, high costs for generating whole-genome maps, the lack of high throughput and the dependence of result interpretation and delineation of genetic changes on WGS makes this technology non-competitive. Nonetheless, considering the advantages afforded by WGM such as little or no bioinformatics knowledge required, inter-lab reproducibility, less labor-intensive than similar methods and shorter time to interpret results compared to WGS makes this technology a worthy candidate to undergo further development. If technological advancements can make WGM high throughput and also reduce the cost associated with whole genome map preparations, this technique would play a pivotal role in updating the clinical microbiology laboratoy both as a standalone routine tool as well as an adjunct to speed up whole genome short read sequencing assemblies. As a matter of fact, we might soon have a new product on the market with roots in

WGM technology that is being commercialized by BioNano genomics (https://bionanogenomics.com/).

Subchapter 2.7 describes how long read single molecule real-time sequencing (SMRT-PacBio) was utilised to identify novel *de novo* antibiotic-resistant mechanisms to nitrofurantoin. Nitrofurantoin-resistant *E. coli* were generated *in vitro* under nitrofurantoin pressure, and resistance could be attributed to a 12-bp deletion in *ribE,* encoding lumazine synthase, an essential enzyme involved in the biosynthesis of flavin mononucleotide (FMN). Using knockout and complementation techniques along with long-read sequencing, we established *ribE* as a novel gene target mediating nitrofurantoin resistance in *E. coli*.

The sudden surge in the discovery of many novel colistin resistance genes in a short span of time is not surprising because these genes have existed in animals and environment for a long time. The continuous use of colistin not only as an antibiotic but also as a growth promoter has helped strains adapt to the environment without any biological cost for the strains. The technology advancement primarily in NGS enables us to find these genes rapidly. However, most of the resistance gene harbouring strains are prevalent in livestock and the environment and are restricted to a few clones of human infection-causing organisms. The most alarming finding was that the main route of transmission of resistance is through transposons and plasmids.

We have delineated colistin resistance mechanisms in **Chapter 3.0**. Colistin has become an extremely important last-line antibiotic, and the emergence of resistance to colistin in multi-drug resistant hospital-acquired *K. pneumoniae* is of great concern. At the same time, the identification of plasmid-based colistin resistance in animal-associated *E. coli* has also raised huge concerns about its transmissibility to humans and the lack of possible treatment options available for patients infected with such pan-drug resistant organisms. As yet, novel *mcr* genes continue to be discovered in animal strains although their transfer to human strains has been relatively restricted.

In **subchapter 3.1,** we describe the identification of a novel plasmid-mediated colistin resistance gene *mcr-2* harboured on the IncX4 plasmid (pKP37-BE) in animal *E. coli* from Belgium. *mcr-2* was harboured on the IS element IS1595 and showed the presence of direct repeats and a complete *tnpA* gene, while inverted repeats were not found. However, the carrier plasmid IncX4 is highly transmissible, showing 102–105-fold higher transfer frequencies in our transconjugation experiments. Importantly, a lack of fitness-cost of the IncX4 carriage in *E. coli* makes this plasmid a highly efficient vehicle for dissemination of *mcr-2.* Interestingly, the pKP37-BE backbone, which likely originated from *Salmonella* spp., harboured a battery of

virulence genes including the *virB4/D4* genes encoding a type-IV secretion system that has been shown to mediate downregulation of host innate immune response genes and to increase bacterial uptake and survival within macrophages and epithelial cells. Outer membrane modifications leading to colistin resistance have been shown to attenuate virulence, and whether these co-harboured virulence genes might compensate the pathogenic abilities of colistin-resistant *E. coli* requires further study.

Subchapter 3.2 describes the molecular mechanisms underlying colistin resistance in *K. pneumoniae*. We identified novel variations in the negative regulator *mgrB* in strains exhibiting varying levels of resistance to colistin. Also, for the first time, we demonstrated that colistin pressure might trigger transposition in *K. pneumoniae*. Our *in vitro* data on colistin resistance selection and stability studies in *K. pneumoniae* suggest that the development of colistin resistance is a rapid process and preferentially occurs through a transposition event and disruption of *mgrB,* a negative regulator of *PhoP/PhoQ* vis-a-vis single nucleotide polymorphisms (SNPs) in other targeted genes (*phop, pmrB,* and phosphoethanolamine transferase). This resistance is also stable as demonstrated by the stability of the phenotype /genotype up to ~6000 generations.

Furthermore, transcriptomics analysis of ST383 *K. pneumoniae* with the *mgrB* gene disrupted in different regions showed individual lists of differentially expressed genes indicating the involvement of different metabolic pathways in mediating colistin resistance via *mgrB* disruption, which seems to be a global gene regulator. Finally, the complexity of colistin-resistant mechanisms observed here warrants a more in-depth molecular analysis.

Subchapters 3.3, 3.4 and 3.5 describe the genetic context of the *mcr-1* gene in Belgium and Vietnam. We found *mcr-1* on an IncFII, and IncP plasmid in *E. coli* strains isolated from Belgian pig farms. The fact that we identified two different genetic associations of *mcr-1* within a small geographic region (Belgium) further underscores the high promiscuity of the mobile element harbouring *mcr-1*. The IncFII plasmid is already known to be highly transmissible and also harbours other resistance genes such as NDM-1 and CTX-M-15. The *mcr-1* is harboured on a multidrug-resistant plasmid. It is the first reported *mcr-1* harbouring complete plasmid sequence from Europe other than *mcr-1* from China. The plasmid also co-harbours tellurite resistance gene clusters indicating possible co-selection. Similarly, we found colistin-resistant *E. coli* harbouring *mcr-1* in food animals in Hanoi, Vietnam. Our findings highlight a high prevalence of the *mcr-1* gene among ESBL-harbouring *E. coli* isolated from a rectal screening of pigs in Vietnam; however, we did not find any evidence of transfer or emergence of the gene among ESBL-harbouring pathogenic *E. coli* of human origin. These data add to recent studies

showing a global emergence of the *mcr-1* harbouring mobile genetic element linked to different plasmids, underscoring the importance of active surveillance in both animal and human populations. It also highlights the importance of a more proactive 'one health' screening approach to identify novel mechanisms and antibiotic resistance vectors from the environment, livestock, and humans.

Subchapter 3.6 describes the first report of isolation of *mcr-1* from a bacterium in a Belgian patient. The strain was an *E. coli* ST88, and *mcr-1* was harboured in the IncHI2 plasmid. The *E. coli* strain also harboured other resistance genes encoding resistance to trimethoprim (*dfrA1*), aminoglycosides (*aadA1* and *aacA4*), sulphonamides (*sul1*), and beta-lactamases (*blaSHV-44*) but not on the same plasmid. To our knowledge, this is the first instance of detection of a bacteria with *mcr-1* from a human with an IncHI2 plasmid. The fact that we identified it in *E. coli* ST88 indicates transmission between human and animals as this clone is prevalent in both animals and humans.

We demonstrate the utility and application of NGS techniques, such as WGS and metagenomics for tracking antibiotic resistance gene transmission via plasmids and understanding the impact of antibiotic use on the commensal flora in **Chapter 4**.

Subchapter 4.1 describes the *in vivo* transfer of a $bla_{\text{CTX-M14}}$-harboring plasmid from *E. coli*, possibly under cephalosporin pressure, into a hyper-virulent *K. pneumoniae* ST23 that could be traced using WGS.

Subchapter 4.2 encapsulates the impact of antibiotic use (moxifloxacin) on the microflora of healthy volunteers. In a proof-of-concept study, we explore the utility of faecal microbial transplantation (FMT) as a solution to facilitate a faster restoration of the host microflora.

5.2 Future perspectives

Biomedical research is data-driven and with the technology advancements in sequencing, the approach "sequence everything & search for" is swiftly overshadowing the conventional hypothesis-driven approach (Greene CS et al. 2016). We are entering into the big-data era, and the data generated from different sequencing technologies and other omics platforms will form the basis for the development of personalised approaches and clinical tools in healthcare and fast-track translational research from the lab to the patient's bedside. The challenge is to accelerate the translation of data and knowledge to define balanced, healthy conditions and to predict and prevent diseases (Mohr C, 2018). During the past 15 years, DNA sequencing technology had advanced rapidly and has now virtually plateaued regarding data output and sequencing time. This rapid development of sequencing technologies has helped us to understand and identify the pitfalls, to make criteria to assist us to train and develop new bioinformatics workflows and improve the quality of bioinformatics analysis. Nonetheless, compared to the amount of data generated, the information extracted remains minuscule, and as technologies and data expand, many new bioinformatics hurdles (algorithms, tools, computational knowledge) need to be addressed. The existing algorithms and workflows are still too technology- or data-specific. Urgently, common algorithms or comprehensive platforms with access to tools for analysis of all types of omics data is needed. Big data storage and integrated omics data generated from different platforms with different formats requires extensive (re-) structuring of the data before any common interpretations can be applied. The challenge we face is to accelerate the translation of data and knowledge to "identification, diagnosis and cure" of diseases.

Quite recently, studies on the molecular mechanisms of antibiotic resistance have undergone a paradigm change. Research is now being directed to harness big data and develop robust algorithms that have a basis in practical experimentation to more confidently predict phenotype-genotype correlations. While these efforts have been escalating in the past few years, the introduction and integration of next generation sequencing (NGS) into the clinical microbiology laboratory and eventually into clinical practice is not smooth. The transition has still not happened primarily due to diagnostic errors resulting from insufficient quality assurance, unacceptably long turnaround times, complex interpretation of data, and last but not least, the need for bioinformatics expertise. Ongoing development of standardised NGS protocols with short sample processing times and high-quality data generated by third generation sequencing technology (Oxford Nanopore and PacBio Single Molecule Real Time:

SMRT) and culture-independent sequencing of samples (metagenomics) is likely to minimise the limitation of turnaround time. Eventually, the development of automated 'black box' NGS pipelines for bioinformatics analysis is expected to facilitate NGS integration in diagnostic clinical microbiology.

One of the last hurdles to taking up culture-independent techniques for patient care would be to accurately predict the bacterial phenotype based on the genotypic information available, i.e., phenotype-genotype correlations. Researchers are evaluating various approaches including machine learning and hidden Markov modelling to translate the vast amount of genotypic information into a coherent bacterial phenotypic signature. While this seems feasible based on WGS on pure bacterial cultures, the challenge is to directly detect pathogens and antibiotic resistance patterns from metagenomics analyses of clinical samples. The impact of such an approach on the timely institution of targeted antibiotic therapy (vis-à-vis empirical) and antibiotic stewardship in hospitals is likely to be immense. A significant caveat to developing a clinical metagenomics approach is our incomplete knowledge of disease pathogenesis and the contribution of specific virulence factors as well the pathogen-pathogen or even pathogen-commensal interactions *in vivo* that might impact the course of the disease. Concerning this, both WGS-based and metagenomics big data analysis are benefitting from using machine learning approaches and from the integration of gene expression levels and metabolite profiles that are helping to capture both genetic and non-genetic sources of variation better.

Another important issue we are/will be facing is privacy and security of omics data linked to patient records from health centres or hospitals. This issue is well-recognised, and the European Union is in the process of adopting a new data regulation, the General Data Protection Regulation (GDPR), to protect personal data. The new regulation would stop discrimination of data and unethical practices and provide transparency when making any of this data public

Thus, we are currently observing a parallel evolution of our understanding of the fundamentals of infectious diseases, antibiotic resistance mechanisms and massive generation of omics data. NGS has already resulted in a paradigm shift from PCR-based to first generation NGS assays that utilise targeted sequencing approaches. Shortly, a coherent merging and interpretation of big data through systems biology approaches and machine trained algorithms are likely to give us the granularity of information required to develop rapid and robust NGS based diagnostics, especially in personalised medicine practice.

5.2.1 References

1. Greene CS, Tan J, Ung M, Moore JH, Cheng C. Big Data Bioinformatics. *Methods (San Diego, Calif)*. 2016; 111:1-2. doi: 10.1016/j.ymeth.2016.11.017.

2. Mohr C, Friedrich A, Wojnar D, et al. qPortal: A platform for data-driven biomedical research. Lisacek F, ed. *PLoS ONE*. 2018;13(1): e0191603. doi: 10.1371/journal.pone.0191603.

5.3 Nederlandse Samenvatting

In de 21ste eeuw heeft antibiotica resistentie zich ontwikkeld tot een globale bedreiging voor de menselijke gezondheid. Hoewel vele nieuwe maatregelen werden getroffen om antibiotica resistentie te controleren, heeft het inherente aanpassingsvermogen van bacteriën gekoppeld met de steeds grotere diversiteit in milieus, die voor een groot deel is toe te schrijven aan ongebreideld en divers antibiotica gebruik, het noodzakelijk gemaakt om de dynamiek van multi-drug resistente bacteriën beter te begrijpen in een poging om antibiotica resistentie onder controle te houden. De introductie van next-generation sequencing technologie (NGS) in de microbiologie is de meest revolutionaire verandering in dit vakgebied geweest sinds de introductie van PCR. NGS biedt vele manieren aan om antibiotica resistentie te voorspellen en reikt mogelijkuitzonderlijke oplossingen aan om de opmars te verhinderen. Aldus, het primaire doel van deze thesis was het gebruik van NGS-technologieën om antibiotica resistentie in belangrijke humane pathogenen te bestuderen.

In inleidend **Hoofdstuk 1**, hebben wij de globale last van antibiotica resistentie en de moleculaire methodes samengevat die momenteel voor detectie worden gebruikt. Daarna concentreerden wij ons op de specifieke mechanismen die ten grondslag liggen aan de ontwikkeling van resistentie tegen colistine in Enterobacteriaceae en het gebruik van next-generation sequencing −technologische achtergrond en veroorloofde voordelen− om antibiotica resistentie te detecteren. Dit hoofdstuk ëindigt met de beschrijving van de algemene en specifieke doelstellingen van de thesis.

Één van de belangrijkste hindernissen voor de toepassing van next-generation sequencing voor het detecteren van ABR is het gebrek aan begrijpbare en up-to-date resistentie genendatabase. In hoofdstuk 2, overlopen wij zowel de bestaande antibiotica resistentie gen databanken als hun bijhorende limitaties, en introduceren BacPipe, een geautomatiseerde pijplijn voor analyse van whole genome data die door ons ontwikkeld werd. Vervolgens, bespreken wijhet nut van whole genome mapping om de bacteriële genoom stabiliteit te bestuderen en als belangrijk hulpmiddel voor second-generation sequencing data analyse.

In **deelhoofstuk 2.1,** bespreekt onze review de huidige staat van vrij beschikbare antibiotica resistentie gen databanken die van substantiële waarde zijn. Nochtans, werden sommige beperkingen genoteerd, met inbegrip van de frequentie van updates, functionaliteit en begrijpbaarheid van de bronnen. Bredere context, consistente gen terminologieën, en gecentraliseerde (vereniging van bodem, humane en andere microbioom/resistoom data) up-

to-date gegevens zullen van cruciaal belang zijn in het traceren van en het identificeren van nieuwe mutaties en genen die verantwoordelijk zijn voor antibiotica resistentie. Aangezien inspanningen om pathogeen genomics te coördineren op internationaal niveau toenemen, hopen wij dat onze studie als platform kan dienen om meer aandacht te besteden aan het ondersteunen van en het uitbreiden van de antibiotica resistentie gen bronnen als kritieke component van onze reactie om multi-drug resistente bacteriën (MDR) en hun verspreiding tegen te gaan.

Deelhoofstuk 2.2 beschrijft een geautomatiseerde whole genome analyse pijplijn, BacPipe, die gericht is op de klinische microbioloog en gezondheidswerkers aangezien de pijplijn weinig of geen bio-informatica vaardigheden vereist. Wij hebben de betrouwbaarheid van de pijplijn geconfirmeerd in het analyseren van verschillende genomen van essentiële klinische pathogenen, die gekarakteriseerd zijn door verschillende grootte, GC-inhoud, en aanwezigheid van repeat regions die een uitdaging vormen voor downstream data analyse. Naast begrijpbaarheid en modulariteit, heeft BacPipe het voordeel van een computationele low-resource, en de pijplijn functionaliteit vereist geen internet connectie of hoge capaciteitscomputers. Een eenvoudige grafische interface maakt de pijplijn zeer gebruikersvriendelijk; een gebruiker kan op directe manier aan de software de input en output map locaties opgeven, de tools specifiëren die opgenomen moeten worden in de analyse en de databank/parameters aanpassen vanuit een drop-down menu of knoppen. BacPipe kan raw reads lopen vanuit verschillende sequencing platforms of de analyse kan herstarten vanuit eender welke stap doorheen de workflow, resulterend in een enorme flexibiliteit.
Door gebruik te maken van ruwe data, bestaande uit contigs vanuit voormalige publicaties die verschillende tools gebruikten om onderscheid te maken tussen ziekenhuis- of gemeenschap-geassocieerde uitbraken of pathogeen transmissie, demonstreerden dat de gehele verzameling aan tools in Bacpipe de volledige analyses kon reproduceren als een ‘one-stop’ platform in enkele uren. Wij geloven dat deze geautomatiseerde pijplijn enkele van de primaire barrières betreffende het analyseren en interpreteren van WGS data zal doorbreken, waardoor toepassingen zullen ontstaan voor routineuze patiëntenzorg in ziekenhuizen, volksgezondheid en monitoring van infectie controle. Toekomstige ontwikkeling van BacPipe omvat de uitbreiding van tools voor de identificatie van profagen, IS (insertion sequence) elements, CRISPR-Cas elementen en, afhankelijk van hun toegankelijkheidsstatus, whole-/core-/pan-genome MLST schema’s mogelijk te maken.

In **deelhoofstuk 2.3,** beschrijven we het gebruik van Whole Genome Mapping (WGM) om multi-drug resistente pathogenen te traceren en karakteriseren, als deel van het European

Public Health initiative (EUPHi). WGM is een op fluorescentie-gebaseerde hoge-resolutie restriction mapping techniek die een belangrijke link vorm tussen PFGE en WGS. De techniek werd gebruikt voor typering van stammen gedurende een uitbraak en kan de genome-wide variaties detecteren. Daarnaast, dient WGM ook als een belangrijk assembly validatie-hulpmiddel van WGS geassembleerde contigs die primair afkomstig zijn van second read sequencing technologie zoals Illumina. Vergeleken met PFGE, genereert het 500 banden en rangschikt deze in het chromosoom (*in situ*). We tonen de bruikbaarheid van WGM aan om de genomische stabiliteit van referentie *S. aureus* stammen te bestuderen, waarbij een grote genomische deletie in één van de twaalf stammen werd geobserveerd. WGS van het gealtereerde genoom en inspectie van de regio, liet toe om de exacte grootte, locatie in het genoom en de aard van deze deletie vast te stellen. WGM kan dus gebruikt worden om zeer snel de genomische stabiliteit van bacteriële stammen over tijd gedurende bijvoorbeeld opslag of interlaboratoir transport in te schatten, gevolgd door identificatie van de verandering in het genoom met behulp van WGS.

Deelhoofdstuk 2.4 beschrijft het gebruik van WGM voor optimale de novo assemblage van bacteriële genomen. Genoom assemblage gebaseerd op de Bruin grafiek omvat mogelijk mis-assemblages wanneer statistieken zoals lengte van de contigs en N50 in beschouwing worden genomen. Doch, werkt WGM als een krachtig hulpmiddel om zulke mis-assemblages op te sporen en voorziet het de optimale k mer groottes die gebruikt kunnen worden om optimaal geassembleerde genomen van short read sequencing technologieën te produceren. Ondanks de extra kost, maken de wetenschappelijke nood voor foutvrije en complete genomen WGM een onmisbare techniek gedurende het proces van genoom assemblage en bijhorende validatie.

Deelhoofdstuk 2.5 beschrijft het gebruik van WGM om een uitbraak van multi-drug resistente *K. pneumoniae* in een Grieks ziekenhuis te traceren en detecteren. De stammen van de uibraak werden gekarakteriseerd door middel van PFGE, WGM, en WGS. Deze studie toont de applicatie van WGM aan in het begrijpen van de epidemiologie van ziekenhuis-geassocieerde *K. pneumoniae*. Daarbovenop, een combinatie van WGM en whole genome sequencing gaf nieuwe inzichten in de genomische kenmerken van de multi-drug resistente ST383 *K. pneumoniae* kloon, die op snel tempo aan belangrijkheid wint inzake prevalentie en klinische significantie in Europa.

Deelhoofdstuk 2.6 gaat over de moleculaire typering bij *A. baumannii* die geïsoleerd werd uit ziekenhuizen in Griekenland en Indië. We hebben de klonaliteit bepaald van *A. baumannii* stammen die werden geïsoleerd uit verschillende ziekenhuizen in Griekenland en 1

ziekenhuis in Indië. Deze stammen werden vergeleken door middel van verschillende typerings methoden, zoals PFGE en MLST. Deze resultaten werden vergeleken met de hoge-resolutie WGM, en de stammen werden ingedeeld op verschillende niveaus. Vergelijking van de hoge-resolutie typeringstechniek WGM met andere typeringsmethoden (MLST en PFGE) toonde de superieure onderscheidingskracht van WGM voor analyse van snel-evoluerende ziekenhuis-geassocieerde *A. baumannii* globale klonen. Ook de hoge diversiteit van de Indische stammen die hier werden bestudeerd was noemenswaardig.

Terwijl WGM een belangrijk hulpmiddel bleek te zijn op verschillende niveaus, hebben de nadelen in termen van hoge kosten geassocieerd met het genereren van whole-genome mappen, gebrek aan high-throughput en het feit dat resultaten steeds geïnterpreteerd moeten worden aan de hand van WGS, deze techniek niet-competitief gemaakt waardoor deze gradueel uit gebruik is genomen in de meeste laboratoria.

Deelhoofdstuk 2.7 beschrijft hoe long read single molecule real-time sequencing (SMRT-PacBio) gebruikt werd om nieuwe *de novo* antibiotic-resistentie mechanismen tegen nitrofurantoin te identificeren. Nitrofurantoin-resistente *E. coli* werden in vitro geconstrueerd onder nitrofurantoin druk, en resistentie kon toegewezen worden aan een 12-bp deletie in het ribE gen, dewelke een lumazine synthase codeert dat een essentieel enzyme vormt in de biosynthese van flavin mononucleotide (FMN). Door gebruik te maken van knockout en complementatie technieken, alsook long-read sequencing, konden we ribE bevestigen als een nieuw gen doelwit, verantwoordelijk voor nitrofurantoin resistentie in *E. coli*.

De plotse opmars van vele nieuwe colistine resistentie genen in een korte tijdsperiode is niet verrassend, aangezien deze genen reeds lange tijd aanwezig zijn in dieren en de omgeving. Door het constante gebruik van colistine als antibioticum, maar ook als groeipromoter, konden stammen zich aanpassen aan de omgeving zonder enige biologische kost. De vooruitgang, voornamelijk in next-generation sequencing technologie, maakt het mogelijk om deze genen onmiddellijk te identificeren. De meeste van deze gen-dragende stammen waren voornamelijk prevalent in vee en het milieu, terwijl slechts enkelen behoorden tot klonen die humane infecties veroorzaken. Het meest schokkende is de hoofdroute van transmissie, welke plaatsvindt door middel van transposons en plasmiden.

In **hoofdstuk 3** besteden we aandacht aan de verschillende resistentiemechanismen tegen colistine. Colistine is het meest belangrijke last-line antibioticum geworden, en het ontstaan van resistentie tegen colistine in multi-drug resistente ziekenhuis-geassocieerde *K. pneumoniae* is van groot belang. Tegelijkertijd, heeft de identificatie van plasmide-gebaseerde colistine resistentie in vee-geassocieerde *E. coli* ook enorme zorgen opgeworpen over hun

overdraagbaarheid naar mensen en het gebrek aan mogelijke behandelingsopties voor patiënten die geïnfecteerd zijn met zulke pan-resistente organismen. Tot op dit moment worden nieuwe *mcr* genen ontdekt in stammen geïsoleerd van vee, alhoewel hun transfer naar stammen in de mens relatief beperkt is gebleven.

In **deelhoofdstuk 3.1**, beschrijven we de identificatie van een nieuw plasmide-gemedieerde colistine resistentie gen, *mcr2*, dewelke gedragen wordt in de IncX4 plasmide (pKP37-BE) in vee-geassocieerde *E. coli* uit België. Mcr-2 werd gedragen op het IS element IS1595 en toonde de aanwezigheid van direct repeats alsook een compleet *tnpA* gen, terwijl inverted repeats niet werden gevonden. De carrier plasmide IncX4 toonde een hoge overdraagbaarheid, met 102-105-voudig hogere transfer frequenties in onze transconjugatie experimenten. Het gebrek aan fitness-kost van het dragen van de IncX4-plasmide in *E. coli* zorgt ervoor dat deze plasmide een zeer efficiënt middel is voor de verspreiding van mcr-2. De pKP37-BE backbone, dewelke afkomstig is van Salmonella spp., droeg een wijde variëteit aan virulentiegenen, waaronder de virB4/D4 genen die een type-IV secretiesysteem encoderen. Van dit secretiesysteem werd aangetoond dat de response genen van het aangeboren immuunsysteem in de gastheer neerreguleerd worden en dat het zorgt voor een verhoogde bacteriële opname alsook overleving in macrofagen en epitheliale cellen. Er werd aangetoond dat aanpassingen in de buitenste membraan die leiden tot colistine resistentie, virulentie verzwakken. Of de bijhorende virulentiegenen op de pKP37-BE backbone de pathogene kwaliteiten van colistine-resistente *E. coli* bevorderen vereist verder onderzoek.

Deelhoofdstuk 3.2 beschrijft de moleculaire mechanismen die aan de basis liggen van colistine resistentie in *K. pneumoniae*. We identificeren nieuwe variaties in de negative regulator mgrB in stammen die verschillende niveaus van resistentie vertonen tegen colistine. We demonstreerden ook voor de eerste keer dat colistine druk conversie in *K. pneumoniae* kan triggeren. Onze in vitro data betreffende colistine resistentie selectie en stabiliteit in *K. pneumoniae* suggereren dat de ontwikkeling van resistentie tegen colistine een snel proces is dat voornamelijk plaats vind door middel van een conversie-event en disruptie van mgrB, een negative regulator van PhoP/PhoQ tegenover single nucleotide polymorphisms (SNPs) in andere genen (phop, pmrB, en phosphoethanolamine transferase). Deze resistentie is ook stabiel, zoals gedemonstreerd werd door de stabiliteit van het fenotype/genotype gedurende 6000 generaties.

Transcriptoom analyse van ST383 *K. pneumoniae* waarbij het mgrB gen werd onderbroken in verschillende regio's resulteerde in individuele lijsten van verschillende geexpresseerde genen. Dit vormt een indicatie van de betrokkenheid van verschillende

metabolische pathways in het hanteren van colistine resistentie via disruptie van mgrB, dewelke een globale gen regulator lijkt te zijn. Tenslotte, rechtvaardigt de complexiteit van de colistine-resistentie mechanismen die hier werden geobserveerd een meer gedetailleerdere moleculaire analyse.

Deelhoofdstukken 3.3, 3.4, en 3.5 beschrijven de genetische context van het *mcr-1* gen in België en Vietnam. We vonden *mcr-1* op een IncFII en IncP plasmide in *E. coli* stammen die geïsoleerd werden uit Belgische varkensboerderijen. Het feit dat we twee verschillende genetische associaties van mcr-1gevonden werden in een beperkte geografische regio (België) onderlijnt de hoge promiscuïteit van het mobiele element dat dit gen draagt. De IncFII plasmide is alreeds gekend als sterk overdraagbaar en draagt ook andere resistentiegenen zoals NDM-1 en CTX-M-15. Het *mcr-1* gen werd gedragen op een multi-drug resistente plasmide. Buiten mcr-1 in China, is dit de eerste complete sequentie van een plasmide in Europa die mcr-1 draagt. De plasmide draagt ook het tellurite resistentiegen die mogelijke co-selectie suggereren. We vonden ook colistine-resistente *E. coli* die *mcr-1* droegen in vee uit Hanoi, Vietnam. Onze bevindingen tonen een hoge prevalentie aan van het *mcr-1* gen in *E. coli* die werden geïsoleerd uit rectale screenings van varkens in Vietnam. We vonden echter geen bewijs van overdracht of aanwezigheid van het gen in pathogene *E. coli* van menselijke oorsprong. Deze data draagt bij tot recente studies die een globale verspreiding aantonen van verschillende plasmiden gelinkt aan het mobiele element dat *mcr-1* draagt. Dit onderstreept het belang van actief toezicht in zowel menselijke als dierlijke populaties en brengt de belangrijkheid van een meer proactieve 'one health' screening benadering naar voren om nieuwe mechanismen en vectoren van antibiotica resistentie van milieu, vee en mens te identificeren.

Deelhoofdstuk 3.6 beschrijft het eerste rapport betreffende de isolatie van mcr-1 uit een Belgische patiënt. De stam *E. coli* ST88 werd geïsoleerd, en droeg mcr-1 op de IncHI2 plasmide. De *E. coli* stam droeg ook andere resistentiegenen zoals de genen die resistentie encoderen tegen trimethoprim (*dfrA1*), aminoglycosides (*aadA1* en *aacA4*), sulfonamides (*sul1*), en beta-lactamase genen (*blaSHV-44*). Deze waren echter niet gelokaliseerd op dezelfde plasmide. Naar onze kennis, is dit de eerste beschrijving van mcr-1 op een IncHI2 plasmide geïsoleerd uit een mens. Het feit dat we de E. coli ST88 stam konden identificeren indiceert transfer tussen mensen en dieren aangezien deze kloon prevalent is in zowel dieren als mensen.

In **hoofdstuk 4** tonen we de bruikbaarheid en toepassing van NGS-technieken, zoals WGS, transcriptomics, en metagenomics aan voor het traceren van antibiotica resistentie gen transmissie via plasmiden en het begrijpen van de impact van antibiotica-gebruik op de commensale flora.

Deelhoofdstuk 4.1 beschrijft de in vivo transfer van een *E. coli* plasmide die *bla*$_{CTX\text{-}M14}$ overdraagt naar een hyper-virulente *K. pneumoniae* ST23 onder cefalosporine druk die getraceerd kon worden door middel van WGS.

Deelhoofdstuk 4.2 omvat de impact van antibiotica-gebruik (moxifloxacin) op de microflora van gezonde vrijwilligers. In een proof-of-concept studie, verkennen we het gebruik van faecal microbial transplantation (FMT) als een oplossing voor beter en sneller herstel van de gastheer microflora.

5.4 Toekomstige perspectieven

Biomedisch onderzoek wordt gedreven door data, en met de technologische vooruitgang van de sequencing technologie, overschaduwt de aanpak van "sequeneer alles en zoek dan" stilaan de conventionele hypothese-gedreven aanpak (Greene CS et al. 2016). We stappen het tijdperk van 'big data' binnen, en deze data, afkomstig van verschillende sequencing technologiën en andere omics platformen, zal de basis vormen voor de ontwikkeling van een gepersonaliseerde aanpak en klinische tools in de gezondheidszorg, en zal translationeel onderzoek snel van het lab naar het bed van de patient brengen. De uitdaging hierbij is om het omzetten van data en kennis te versnellen zodat men gebalanceerde, gezonde condities kan definiëren en ziektes kan voorspellen en voorkomen (Mohr C, 2018). Tijdens de afgelopen 15 jaar heeft de DNA sequencing technologie snel vooruitgang geboekt, en nu heeft de technologie zijn virtueel plateau bereikt in termen van data output en sequencing tijd. Deze snelle ontwikkeling van sequencing technologieën heeft ons geholpen om valkuilen te begrijpen en te identificeren, alsook om criteria op te stellen die ons helpen bij de training en ontwikkeling van nieuwe bioinformatica workflows en de kwaliteit van bioinformatica analyses te verbeteren. Niettegenstaande, vergeleken met de hoeveelheid gegenereerde data, is de informatie die geselecteerd wordt nog steeds miniscuul en aangezien technologieën en data toenemen, zullen vele nieuwe bioinformatica problemen, zoals algoritmen, tools en computationele kennis, overwonnen moeten worden. De bestaande algoritmen en workflows zijn momenteel nog te technologisch of data specifiek. Er is momenteel dringend nood aan gemeenschappelijke algoritmen of uitgebreide platformen met toegang tot tools voor de analyse van alle types omics data. De opslag van 'big data' en de integratie van omics data afkomstig van verschillende platformen met verschillende formats vereisen een uitgebreide (her)structurering van data alvorens enige gemeenschappelijke interpretatie kan worden toegepast. De uitdaging waarmee

we geconfronteerd worden is om de omzetting van data en kennis naar " identificatie, diagnose en behandeling" van ziektes te versnellen.

Recent hebben studies over moleculaire mechanisms van antibiotica resistentie een structurele verandering ondergaan. Onderzoek wordt nu gericht op het aanwenden van big data en op het onwikkelen van robuuste algoritmen die een basis hebben in praktische experimenten om op meer betrouwbare wijze fenotype-genotype correlaties te voorspellen. Hoewel deze inspanningen zijn toegenomen in de afgelopen jaren, is de introductie en integratie van next-generation sequencing (NGS) in de laboratoria van klinische microbiologie en dus ook de klinische praktijk niet soepel verlopen. De overstap is nog steeds niet gebeurd, voornamelijk omwille van diagnostische fouten ten gevolge van onvoldoende kwaliteitsbewaking, onaanvaardbaar lange doorlooptijden, complexe interpretatie van data alsook de noodzaak aan expertise in bioinformatica. De voortdurende ontwikkeling van gestandardiseerde NGS protocollen met korte verwerkingstijden van het staal en data van hoge kwaliteit gegenereerd door derde generatie sequencing technologieën (Oxford Nanopore and PacBio Single Molecule Real Time: SMRT) en cultuur onafhankelijke sequencing van stalen (metagenomics) zal hoogstwaarschijnlijk de limitatie van de doorlooptijd tot een minimum beperken.

Verwacht wordt dat de ontwikkeling van geautomatiseerde 'black box' NGS pijplijnen voor bioinformatica analyses uiteindelijk de integratie van NGS in de diagnostische klinische microbiologie zal vergemakkelijken. Een van de laatste hindernissen om cultuuronafhankelijke technieken op te nemen in de patiëntenzorg, is het accuraat voorspellen van het bacteriële fenotype op basis van de beschikbare genotypische informatie, namelijk de fenotype-genotype correlaties. Onderzoekers evalueren verschillende aanpakken, bestaande uit machine learning en verborgen Markov modelling om de enorme hoeveelheid aan genotypische informatie te vertalen naar een coherent bacterieel fenotype. Hoewel dit haalbaar lijkt op basis van WGS van zuivere bacteriële culturen, is het de uitdaging om pathogenen en antibiotica resistentiepatronen direct te detecteren uit de metagenomics-analyse van klinische stalen. De impact van een dergelijke aanpak op het tijdige instellen van een gerichte antibioticatherapie (vis-à-vis empirisch) en op het antibiotica beheer in ziekenhuizen zal hoogstwaarschijnlijk enorm zijn. Een mogelijke valkuil bij het ontwikkelen van een klinische metagenomics aanpak is de beperkte kennis over de pathogenese van ziekte en de bijdrage van specifieke virulentiefactoren, alsook pathogeen-pathogeen of zelfs pathogeen-commensale interacties *in vivo* die het verloop van de ziekte kunnen beïnvloeden. Dit betreffende, hebben zowel WGS-gebaseerde als metagenomics big data analyse baat bij het gebruik van de machine learning-

aanpak en van de integratie van genexpressie niveaus en van de metaboliet profielen die bijdragen tot een beter begrip van zowel de genetische als niet-genetische bronnen van variatie.

Een ander belangrijk probleem waarmee we geconfronteerd (zullen) worden, is de privacy en de veiligheid van omics gegevens gekoppeld aan patiëntendossiers van gezondheidscentra of ziekenhuizen. Dit probleem wordt algemeen erkend en de Europese Unie is bezig met de goedkeuring van een nieuwe wetgeving om persoonsgegevens te beschermen, namelijk de General Data Protection Regulation, GDPR). Deze nieuwe wetgeving zou de discriminatie van gegevens alsook de onethische praktijken stoppen en zorgen voor transparantie wanneer deze data publiek worden gemaakt.

Momenteel kunnen we dus een parallelle evolutie waarnemen in ons begrip van de basisprincipes van infectieziektes, antibioticaresistentie mechanismen en de massale aanmaak van omics-gegevens. NGS heeft reeds geresulteerd in een paradigmaverschuiving van PCR-gebaseerde naar eerste generatie NGS assays die een gerichte sequencing aanpak hanteren. In de nabije toekomst, zal een coherente samenvoeging en interpretatie van big data via systeembiologie benaderingen en machine getrainde algoritmen ons waarschijnlijk de granulariteit van informatie verschaffen die nodig is om snelle en robuuste NGS-gebaseerde diagnostiek te ontwikkelen, voornamelijk in de gepersonaliseerde geneeskunde praktijk.

CURRICULUM VITAE

Name Basil Britto Xavier
Place of Birth Tiruchirappalli, Tamil Nadu, India
Current Affiliation Laboratory of Medical Microbiology, VAXINFECTIO
Campus drie eiken, S6.25, University of Antwerp,
Universiteitsplein 1, 2610 Wilrijk, Belgium
Tel: +32 3 265 25 50, mobile: 0474425115
Email: basilbritto.xavier@uantwerpen.be

EDUCATION

- PhD in Medical Sciences, October 2012 - May 2018, Laboratory of Medical Microbiology (LMM), VAXINFECTIO, University of Antwerp, Antwerp, Belgium.
- Masters in Biology, Institute of Biology, University of Iceland, Reykjavik, Iceland.
- Masters in Bioethics, (Erasmus Mundus-EMMB), Katholieke University: KU Leuven), Belgium, Radbound University, Netherlands and University of Padova, Italy.
- Master of Philosophy in Microbiology, Bharathidasan University, Tiruchirappalli, Tamil Nadu, India.
- Masters in Microbiology, Bharathidasan University, Tiruchirappalli, Tamil Nadu, India.
- Post Graduate Diploma in Bioinformatics, Bharathidasan University, Tiruchirappalli, Tamil Nadu, India.
- Bachelors in Microbiology, Bharathidasan University, Tiruchirappalli, Tamil Nadu, India.

PUBLICATIONS PART OF THIS THESIS

1. **Xavier BB**, Plantinga N, Coppens J, Zarkotou O, Minh N, Lammens C, Janssens L, Lal S, Kumar-Singh S, Bonten M, Tsakris A, Pournaras S, Goossens H, Malhotra-Kumar S. 2018. A systems biology approach identifies pathways mediating colistin resistance in *Klebsiella pneumoniae* (*In submission*).
2. **Xavier BB**, Bonten M, Goossens H, Malhotra-Kumar S. 2018. Identification of *mcr1* gene in *Escherichia coli* ST88 from human isolate harboured in the IncHI2 plasmid. (*In preparation*).
3. **Xavier BB***, Mysara M*, Bolzan M, Lammens C, Kumar-Singh S, Goossens H, Malhotra-Kumar S. 2018. BacPipe: A rapid, user-friendly whole genome sequencing pipeline for clinical diagnostic bacteriology and outbreak detection (*In revision*).
   ***equal first author contribution**
4. **Xavier BB**, Lammens C, Ruhal R, Butaye P, Kumar-Singh S, Goossens H, Malhotra-Kumar S. Identification of a novel plasmid-mediated colistin resistant gene *mcr-2* in *E. coli*, Belgium, June 2016. *Euro surveillance. 2016*. **IF:7.2**
5. **Xavier BB**, Lammens C, Butaye P, Goossens H, Malhotra-Kumar S. Complete

sequence of an IncFII plasmid harbouring the colistin resistance gene mcr-1 isolated from Belgian pig farms. Journal of Antimicrobial Chemotherapy. 2016. **IF: 5.313**

6. Malhotra-Kumar S, **Xavier BB**, Das AJ, Lammens C, Hoang HTT, Pham NT, et al. Colistin-resistant *Escherichia coli* harbouring *mcr-1* isolated from food animals in Hanoi, Vietnam. The Lancet Infectious Diseases. 2016. **IF: 22.43**
7. Malhotra-Kumar S, **Xavier BB**, Das AJ, Lammens C, Butaye P, Goossens H. Colistin resistance gene *mcr-1* harboured on a multidrug-resistant plasmid. The Lancet Infectious Diseases. 2016. **IF: 22.43**
8. **Xavier BB**, Das JA, C Guy, De Ganck S, Kumar-Singh S, Aarestrup FM, Goossens H, Malhotra-Kumar S. Effective use of antibiotic resistance gene databases for the whole genome and metagenomic sequencing data Journal of Clinical Microbiology. 2016. **IF: 3.993**
9. Sabirova JS*, **Xavier BB***, Coppens J, Lammens C, Wagner T, Pournaras S, Goossens H, Malhotra-Kumar S. Whole-genome typing and characterisation of bla_{VIM19}-harbouring ST383 *Klebsiella pneumoniae* by PFGE, whole genome mapping and sequencing. *Journal of Antimicrobial Chemotherapy*. 2016. **IF: 5.313**

   ***equal first author contribution**
10. **Xavier BB,** Coppens J, Vanjari L, Sabirova JS, Lammens C, Vemu L, Wagner T, Pournaras S, Goossens H, Malhotra-Kumar S. 2018Comparison of typing methods for hospital-associated *Acinetobacter baumannii* clones isolated in Greece and India. (*In preparation*)
11. **Xavier BB**, Gesuele R, Lammens C, Abdessalam C, Goossens H, Schrenzel J, Harbarth S, Malhotra-Kumar S. 2018. Evidence of an *in vivo* transfer of a $bla_{CTX-M14}$-harboring plasmid under antibiotic pressure, Geneva, Switzerland (*Under revision*).
12. **Xavier BB,** Mazlom H, El Hadidi M, Mysara M, Brems P, Lammens C, Huson D, Malhotra-Kumar S, Goossens H. 2018. Faster restoration of the major microbial taxa in the human gut with fecal microbial transplantation (FMT) after treatment with an antibiotic (Moxifloxacin) (*In preparation*)
13. **Xavier BB**, Sabirova J, Moons P, Goossens H, Malhotra-Kumar S. 2014. Employing whole genome mapping for optimization of assembly parameters during *de novo* assembly of *Staphylococcus aureus* genomes BMC Research Notes 07/2014; 7(484). doi:10.1186/1756-0500-7-484
14. Sabirova JS, **Xavier BB,** Ieven M, Goossens H, Malhotra-Kumar S. 2014. Whole genome mapping as a fast-track tool to assess genomic stability of sequenced *Staphylococcus aureus* strains BMC Research Notes doi:10.1186/1756-0500-7-704.
15. **Xavier BB,** Vervoort J, Stewardson A, Adriaenssens N, Coenen S, Harbarth S, Goossens H, Malhotra-Kumar S. 2014. Complete genome sequences of nitrofurantoin-sensitive and -resistant *Escherichia coli* ST540 and ST2747 strains. Genome Announc. 2(2): e00239-14. doi:10.1128/genomeA.00239-14

PUBLICATIONS <u>NOT</u> PART OF THIS THESIS

1. Vervoort J*, **Xavier BB***, Stewardson A, Adriaenssens N, Coenen S, Harbarth S, Goossens H, Malhotra-Kumar S. 2015. Metagenomic analysis of the impact of nitrofurantoin treatment on the human faecal microbiota. Journal of Antimicrobial

Chemotherapy. March 12, 2015 **IF: 5.439** ***equal contributions**.

2. Hotterbeekx, A*, **Xavier BB*,** Bielen K, Lammens C, Moons P, Schepens T, Goossens H, Malhotra-Kumar S (2016). The endotracheal tube microbiome associated with *Pseudomonas aeruginosa* or *Staphylococcus epidermidis*. *Scientific Reports*, *6*, 36507. http://doi.org/10.1038/srep36507 **IF: 5.52 *equal contributions**.
3. Vervoort J, **Xavier BB,** Stewardson A, Adriaenssens N, Coenen S, Harbarth S, Goossens H, Malhotra-Kumar S. An *in vitro* deletion in *ribE* encoding lumazine synthase contributes to nitrofurantoin resistance in *Escherichia coli*. Antimicrob Agents Chemother 2014 Sep 22. **IF: 4.451**
4. Vervoort J, **Xavier BB,** Joossens M, Versporten A, Lammens C, Goossens H, Malhotra-Kumar S. Characterization of the naso- and oropharyngeal resistome in the European population. (*In preparation*)
5. Sabirova J, **Xavier BB,** Hernalsteens J-P, de Greve H, Ieven M, Goossens H, Malhotra-Kumar S. 2014. Complete genome sequences of two prolific biofilm-forming *Staphylococcus aureus* isolates belonging to USA300 and EMRSA-15 clonal lineages. Genome Announc. 2(3) : e00610-14. Doi :10.1128/genomeA.00610-14.
6. **Xavier BB,** Miao VP, Jonsson ZO, Andresson OS. 2012. Mitochondrial genomes from the lichenized fungi *Peltigera membranacea* and *Peltigera malacea*: Features and phylogeny, *Fungal Biology* 116: 802-814. **IF: 2.01**
7. Dafopoulou K, **Xavier BB**, Hotterbeekx A, Janssens L, Lammens C, De E, et al. Colistin-resistant Acinetobacter baumannii clinical strains deficient in biofilm formation. *Antimicrob Agents Chemother. 2015* Dec 14. **IF: 4.76**
8. Julia S, Hernalsteens J-P, de Backer S, **Xavier BB,** Moons P, Turlej-Rogacka A, De Greve H., Goossens H, Malhotra-Kumar S. Fatty acid kinase A is an important determinant of biofilm formation in Staphylococcus aureus USA300 BMC genomics-issn 1471-2164-16 (2015) **IF: 3.986**.
9. Tomislav K, Strateva T, **Xavier BB**, Proevska J, Lammens C, Markova B, Goossens H, Malhotra-Kumar S. Detection, and characterization of two NDM-1 producing *Klebsiella pneumoniae* strains from Bulgaria. *Journal of Antimicrobial Chemotherapy*. 2016. **IF: 5.313**
10. Turlej-Rogacka A, **Xavier BB,** Janssens L, Lammens C, Zarkotou O, Pournaras S, et al. Evaluation of colistin stability in agar and comparison of four methods for MIC testing of colistin. European Journal of Clinical Microbiology & Infectious Diseases. 2017. **IF: 2.72**
11. Singhal L, Sharma M, Verma S, Kaur R, **Xavier BB**, Malhotra-Kumar S, Ray P, Gautam V. 2018. Comparative evaluation of broth microdilution with polystyrene and glass-coated plates, agar dilution, E-test, Vitek and disk diffusion for susceptibility testing of colistin and polymyxin B on carbapenem-resistant clinical isolates of *Acinetobacter baumannii*. Microb. Drug Resist. **IF: 2.306**
12. Partridge SR, Pilato VD, Doi Y, Feldgarden M, Haft DH, Klimke W, Kumar-Singh S, Liu JH, Malhotra-Kumar S, Prasad A, Rossolini GM, Schwarz S, Shen J, Walsh T, Wang Y, **Xavier BB**. 2018. Proposal for assignment of allele numbers for mobile colistin resistance (*mcr*) genes. *Journal of Antimicrobial Chemotherapy*. 2018. **IF: 5.071**

PRESENTATIONS IN INTERNATIONAL MEETINGS

1. **Xavier BB** Invited Speaker at Illumina Technical User Experience Day, Brussels, Belgium (4th, June 2015, UZ Brussels).
2. **Xavier BB,** Plantinga N, Lammens C, Poirel L, Nordmann P, Bonten M, Goossens H, Malhotra-Kumar, S. 2016. Novel *mgrB* variants identified in colistin-resistant *Klebsiella pneumoniae* Eur. Soc. Clin. Microbiol. Infect. Dis. (ECCMID, 26th Ed) Amsterdam, The Netherlands (Talk).
3. **Xavier BB**. Studies on Antibiotic resistance & NGS applications in Clinical Microbiology at Christian Medical College (CMC) Vellore, Tamil Nadu, India on December 20th, 2016 (Talk).
4. **Xavier BB.** Emerging Pan Drug-Resistant Gram-negative bacteria at Aster MIMS Hospital, Calicut, Kerala India on 6th Jan 2017 (Talk).
5. **Xavier BB,** Mazlom H, El Hadidi M, Mysara M, Lammens C, Huson D, Malhotra-Kumar S, Goossens H. 2017. Faster restoration of the major microbial taxa in the human gut with faecal microbial transplantation (FMT) after treatment with an antibiotic (Moxifloxacin) Eur. Soc. Clin. Microbiol. Infect. Dis. (ECCMID, 27th Ed), Vienna, Austria (Talk)
6. **Xavier BB,** Zarkotou O, Coppens J, Lammens C, Tsakris A, Pournaras S, Goossens H, Malhotra-Kumar S. 2018. Genetic characterization and transcriptomic profiling of colistin-resistant phenotypes in *Klebsiella pneumoniae* ST383. Eur. Soc. Clin. Microbiol. Infect. Dis. (ECCMID, 28th Ed), Madrid, Spain (Talk)

POSTER PRESENTATION IN INTERNATIONAL CONFERENCES

1. **Xavier BB,** Janssens L, Zarkotou O, Lammens C, Tsakris A, Pournaras S, Goossens H, Malhotra-Kumar, S. 2015. Genome-wide analysis of colistin-resistant ST383 *Klebsiella pneumoniae*. Eur. Soc. Clin. Microbiol. Infect. Dis. (ECCMID, 25th Ed) **Poster**
2. **Xavier BB,** Coppens J, Zarkotou O, Lammens C, Tsakris A, Pournaras S, Goossens H, Malhotra-Kumar, S. 2015. High-Resolution Clonality of Outbreak-Causing *Acinetobacter baumannii* studied by Whole Genome Mapping. Eur. Soc. Clin. Microbiol. Infect. Dis. (ECCMID, 25th Ed) **Poster**
3. **Xavier BB,** Vanjari L, Vemu L, Lammens C, Goossens H, Malhotra-Kumar S. 2015. Whole genome analysis of methicillin-resistant *Staphylococcus aureus* from tertiary care hospital in Southern India) Eur. Soc. Clin. Microbiol. Infect. Dis. (ECCMID, 25th Ed) **Poster**
4. **Xavier BB,** Vervoort J, Stewardson A, Adriaenssens N, Coenen S, Harbarth S, Goossens H, Malhotra-Kumar S. A deletion in lumazine synthase contributes to nitrofurantoin resistance in *Escherichia coli* ESCMID Conference on Reviving Old Antibiotics, Oct 2014, Vienna, Austria. **Poster**
5. **Xavier BB,** Janssens L, Zarkotou O, Lammens C, Tsakris A, Pournaras S, Goossens H, Malhotra-Kumar, S Heteroresistance to colistin in *Klebsiella pneumoniae* ESCMID

Conference on Reviving Old Antibiotics, Oct 2014, Vienna, Austria **Poster**

6. **Xavier BB**, Sabirova J, Moons P, Goossens H, Malhotra-Kumar S. Parameter optimization for genome assembly: for non-model organisms with the assistance of whole genome mapper (optical mapping) ECCMID 2014 Barcelona, Spain (May 10-13, 2014) **Poster**
7. **Xavier BB**, Julia S, Zarkotou O, Lammens C, Tsakris A, Pournaras S Goossens H, Malhotra-Kumar S. Analysis of colistin (CL) resistance (CR) in Klebsiella pneumoniae (KPN) ICAAC 2013 (Sep 10-13) at Denver, CO, USA. **Poster**

OTHER PRESENTATIONS

1. **Xavier BB**, Hadi Ameen, Jasmine Coppens, Lavanya Vanjari, Runak Hawrame, Azad O. Maged, Youri Glupczynski, Christine Lammens, Lakshmi Vemu, Herman Goossens and Surbhi Malhotra-Kumar Genetic context of ***bla***$_{\text{NDM-1}}$ isolated from *Acinetobacter baumannii* and *Klebsiella pneumoniae* from India, Belgium, and Iraq. Eur. Soc. Clin. Microbiol. Infect. Dis. (ECCMID, 26th Ed) **Poster**
2. Anupam J Das, **Xavier BB**, Guy Cochrane, Sandra De Ganck Samir Kumar-Singh, Frank Møller Aarestrup, Herman Goossens, Surbhi Malhotra-Kumar. Effective use of antibiotic resistance gene databases for the whole genome and metagenomic sequencing data Eur. Soc. Clin. Microbiol. Infect. Dis. (ECCMID, 26th Ed) **Talk**
3. Jasmine Coppens, **Xavier BB,** Anna Gorska, Biljana Carevic, Christine Lammens, Jacques Schrenzel, Daniel Huson, Herman Goossens, Evelina Tacconelli, Surbhi Malhotra-Kumar on behalf of SATURN WP1 and 4 study groups. Genetic adaptations in hospital-acquired MRSA upon switch from a colonizing to infecting lifestyle in patients developing pneumonia Eur. Soc. Clin. Microbiol. Infect. Dis. (ECCMID, 26th Ed) **Talk**
4. De Backer S, **Xavier BB,** Vanjari L, Coppens J, Lammens C, Vemu L, Schrenzel J, Tacconelli E, Goossens H, Malhotra-Kumar S. Genomic context of a colonization and pathogenicity determinant, *sasX*, in methicillin-resistant *Staphylococcus aureus* Eur. Soc. Clin. Microbiol. Infect. Dis. (ECCMID, 26th Ed) **Talk**.
5. Agata Turlej-Rogacka, **Xavier BB**, Lore Janssens, Christine Lammens, Olympia Zarkotou, Youri Glupczynski, Herman Goossens, Surbhi Malhotra-Kumar.Colistin stability and MIC testing in agar dilution in comparison to E-test, micro- and macrobroth. Eur. Soc. Clin. Microbiol. Infect. Dis. (ECCMID, 26th Ed) **Poster**
6. **Xavier BB,** Coppens J, Vanjari L, Dafopoulou K, Sabirova J, Lammens C, Zarkotou O, Wagner T, Tsakris A, Vemu L, Pournaras S, Goossens H, Malhotra-Kumar S. Comparison of Typing Methods for Hospital-Associated *Acinetobacter baumannii* Clones isolated in Greece and India. ESCMID ESGEM Conference 9th IMMEM, Estoril, Portugal (March 2016) **Poster**
7. Janssens L, **Xavier BB,** Zarkotou O, Dafopoulou K, Lammens C, Tsakris A, Pournaras S, Goossens H, Malhotra-Kumar, S. 2015. Colistin resistance and biofilm formation in international *Klebsiella pneumoniae* ST383 Eur. Soc. Clin. Microbiol. Infect. Dis. (ECCMID, 25th Ed) **Poster**
8. **Xavier BB,** Vanjari L, Vemu L, Lammens C, Goossens H, Malhotra-Kumar S. 2015. Whole genome analysis of methicillin-resistant *Staphylococcus aureus* from tertiary

care hospital in Southern India) Eur. Soc. Clin. Microbiol. Infect. Dis. (ECCMID, 25th Ed) **Poster**

9. De Backer S, Vanmarsenille C., Hernalsteens J, **Xavier BB**, De Greve H, Goossens H & Malhotra-Kumar S. Stage-dependent changes in matrix composition of biofilms formed by methicillin-resistant *Staphylococcus aureus* USA300. Eurobiofilms, 2015. **Poster**
10. **Xavier BB.** Heteroresistance to colistin in *Klebsiella pneumoniae* ESCMID Conference on "Reviving Old Antibiotics" Oct 2014, Vienna, Austria. **Poster**
11. **Xavier BB.** Parameter optimisation for genome assembly: for non-model organisms with the assistance of whole genome mapper (optical mapping) ECCMID May 10-13, 2014 Barcelona, Spain. **Poster**
12. Hotterbeekx A, **Xavier BB,** Zarkotou O, Dafopoulou K, Ruhal R, Lammens C, Tsakris A, Pournaras S, Goossens H, Malhotra-kumar S. Changes in the capsular polysaccharide synthesis locus correlate to increased biofilm formation in colistin-resistant *Klebsiella pneumoniae* ST383 at ECCMID 2017, Vienna, Austria. **Poster**
13. Ruhal R, **Xavier BB**, Lammens L, Goossens.H, Singh SK, Malhotra-Kumar, S. Multiple signalling systems regulate biofilm dispersal in *Pseudomonas aeruginosa* Eur. Soc. Clin. Microbiol. Infect. Dis. (ECCMID, 28th Ed), April 21st -24, 2018, Madrid, Spain
14. Ruhal R, **Xavier BB**, Lammens L, Goossen H, Singh SK, Malhotra-Kumar, S Transcriptomic of biofilm of Pseudomonas aeruginosa. 5th European Congress on Biofilms: Eurobiofilm 19-22 Sep, 2017 Amsterdam, The Netherlands.
15. Ruhal R, **Xavier BB**, Lammens L, Goossens.H, Singh SK, Malhotra-Kumar, S. Role of surface adhesins in biofilm dispersal and increase in virulence in pathogenic bacteria Pseudomonas aeruginosa. FWO Research Community 4th Workshop on Bacterial and Fungal Biofilms, Leuven, Belgium, 2017
16. Hotterbeekx A, **Xavier BB**, Moons P, Lammens C, Ieven M, Vandenbroeck E, Jorens P, Kumar-Singh S, Goossens H, Malhotra-Kumar S. Identification of a core microbiome of endotracheal tubes associated with *Pseudomonas aeruginosa* and/or *Staphylococcus epidermidis*, ECCMID 2014, Barcelona, Spain. **Poster**

INTERNATIONAL PRESS COVERAGE FOR RESEARCH & PUBLISHED ARTICLES

- https://www.statnews.com/2016/07/07/superbug-new-gene-discovery/
- http://news.nationalgeographic.com/2016/07/germination-superbug-colistin-second-case/
- https://www.washingtonpost.com/news/to-your-health/wp/2016/07/11/superbug-gene-detected-in-a-second-person-in-the-u-s/
- http://www.hcpro.com/ACC-327394-4634/Second-gene-discovered-that-makes-bacteria-resistant-to-lastresort-antibiotic.html
- https://tlarremore.wordpress.com/2016/07/11/biological-hazard-antibiotic-resistance-colistin-mcr-2-plasmid-e-coli-animal-human-european-continent-belgium/
- http://antibiotic-action.com/new-plasmid-mediated-colistin-resistance-gene-found-

mcr-2/

- http://www.chicagotribune.com/news/nationworld/ct-superbug-second-person-20160711-story.html
- https://www.nrdc.org/experts/david-wallinga-md/spreading-resistance-last-resort-drugs-us
- http://www.upi.com/Health_News/2016/07/08/Second-gene-causing-resistance-to-last-resort-antibiotic-found-by-scientists/6251467982893/
- Linked to antibiotic use and FMT to the general public, for instance, television program, was telecast as PANO "hoe-schadelijk-zijn-antibiotics" on een tv on 19th October 2016 at 22.00 PM https://www.een.be/pano/hoe-schadelijk-zijn-antibiotica later on in the newspaper http://www.demorgen.be/wetenschap/experiment-in-pano-pil-van-eigen-stoelgang-kan-darmflora-herstellen-b4e0ac0f/;

AWARDS AND GRANTS

- Erasmus Mundus Scholarship by European Commission for the Academic year of 2007-2008 for Masters in Bioethics (EMMB).
- Workmanship award by Aster MIMS, Calicut, Kerala, India on 6th Jan 2017, for work on "Emerging Pan Drug Resistant Gram-negative bacteria"
- Travel grant awarded by British Mycological Society to participate "International Mycological Congress (IMC 9- 2010)" held at Edinburgh, Scotland UK
- ESCMID Attendance Grant (2014) for attending a conference on "Reviving Old Antibiotics" at Vienna, Austria
- European Molecular Biology Laboratory (EMBL) registration fee waiver grant for course on "Next Generation Sequencing: Total RNA sequencing library preparation and validation" October, 2014.
- Travel grant awarded by FWO to participate "ECCMID 2015" held at Copenhagen, Denmark.
- ESCMID Grant for Postgraduate Technical Capacity-building Workshop: Rapid NGS for Characterization and Typing of Resistant Gram-Negative Bacilli, at Groningen, Netherlands, October 7th - 9th 2015.
- Travel grant awarded by FWO to participate "ECCMID 2017" held at Vienna, Austria.
- Travel grant awarded by FWO to participate "ECCMID 2018" held at Madrid, Spain.

PROFESSIONAL MEMBERSHIP

- European Society of Clinical Microbiology and Infectious Diseases
- European Study Group for Antimicrobial Resistance (ESGARS)
- European Study Group for Clostridium difficile (ESGCD)

COURSES COMPLETED DURING PhD

- One Day training on Whole Genome Mapping (Optical Mapping) by OpGen, USA at LMM, University of Antwerp, Antwerp, Belgium – November, 2012.
- International Course in Automated Functional Annotation and Data Mining (Blast2GO) at Centro de Investigacion Principe Felipe, Valencia, Spain from October 23rd - 25th, 2013.

- Research training on Next Generation Sequencing: Total RNA sequencing library preparation and validation at European Molecular Biology Laboratory (EMBL), Heidelberg, Germany, from October 8 - 10th, 2014.
- Research training on Microbial Single Cell Genomics Workshop at Bigelow Laboratory for Ocean Sciences, Single Cell Genomics Center, Spruce Point Inn, Boothbay Harbor, Maine, USA, from June 8 – 14th, 2015.
- Research training on Capacity-building Workshop: Rapid NGS for Characterization and Typing of Resistant Gram-Negative Bacilli at University of Groningen, Groningen, The Netherlands, from October 7 - 9th, 2015.
- Research training on Whole Transcriptomics Data Analysis at European Molecular Biology Laboratory (EMBL), Heidelberg, Germany from October 4 - 7th, 2016.

ACKNOWLEDGEMENTS

This research thesis would not have been possible without the support of many people. I would like to extend my deepest gratitude primarily to the following persons. First and foremost, I wish to thank my promoter prof. dr. Surbhi Malhotra-Kumar whose patience, kindness, wisdom, commitment, as well as her academic excellence, have been invaluable to me. She has supported me by not only providing me with this research opportunity but also gave me the freedom to be a part of many projects which helped me to gain knowledge beyond my topic and also taught me the state of the art of scientific writing. I gratefully acknowledge my co-promoter prof. dr. Herman Goossens who allowed me to be part of many projects and always trusted my work. Deepest gratitude to the members of the internal doctoral committee prof. dr. Geert Mortier and prof. dr. Jurgen Del-Favero and external examiners prof. dr. João André Carriço and prof. dr. Jan Kluytmans without whose assistance this would have not been successful. My special thanks to prof. dr. Samir Kumar Singh for offering me valuable advice and for the encouragement in improving my interpersonal skills. My special thanks go to Christine Lammens for always being there to clarify all work-related issues. I would also like to give my appreciation to our collaborators and co-authors. This thesis and research would not be realistic without funding, so my thanks to the University of Antwerp Research Grants BOF (2012-27450) for funding my PhD fellowship. I would like to say special thanks to Kristin Deby for being kind and answer all my administrative questions during this thesis preparation. I would like to thank all LMM members (Sabine, Liesbet V, An, Sarah, Rohit, Minh, Mandy, Kitty, Geert, Liesbeth, Anouk, Karen, Kathleen, Pieter and Stalin) and ex-colleagues (Veronique, Mohamed A, Mattia and Lore) for their support. My special thanks to Jasmine for being a friend and workmate. Finally, my deep gratefulness goes to my family members Amma, Appa, Sisters (Helen and Emaly), brother-in-laws, parent-in-law, especially to my wife Sheeba, little darling Benita and Marshall (for sacrificing his classes in order to take care of my daughter so that I can concentrate on my work) for their unconditional love and support throughout my life. My special heartfelt thanks to my uncle, Rev. Fr. dr. G. Amaldoss, who is always my moral support. Last but not least, I thank the Almighty for the wisdom and perseverance that he has bestowed upon me during this time, and indeed, throughout my life through all tests in the past years. You have made my life more bountiful. May your name be exalted, honoured, and glorified.