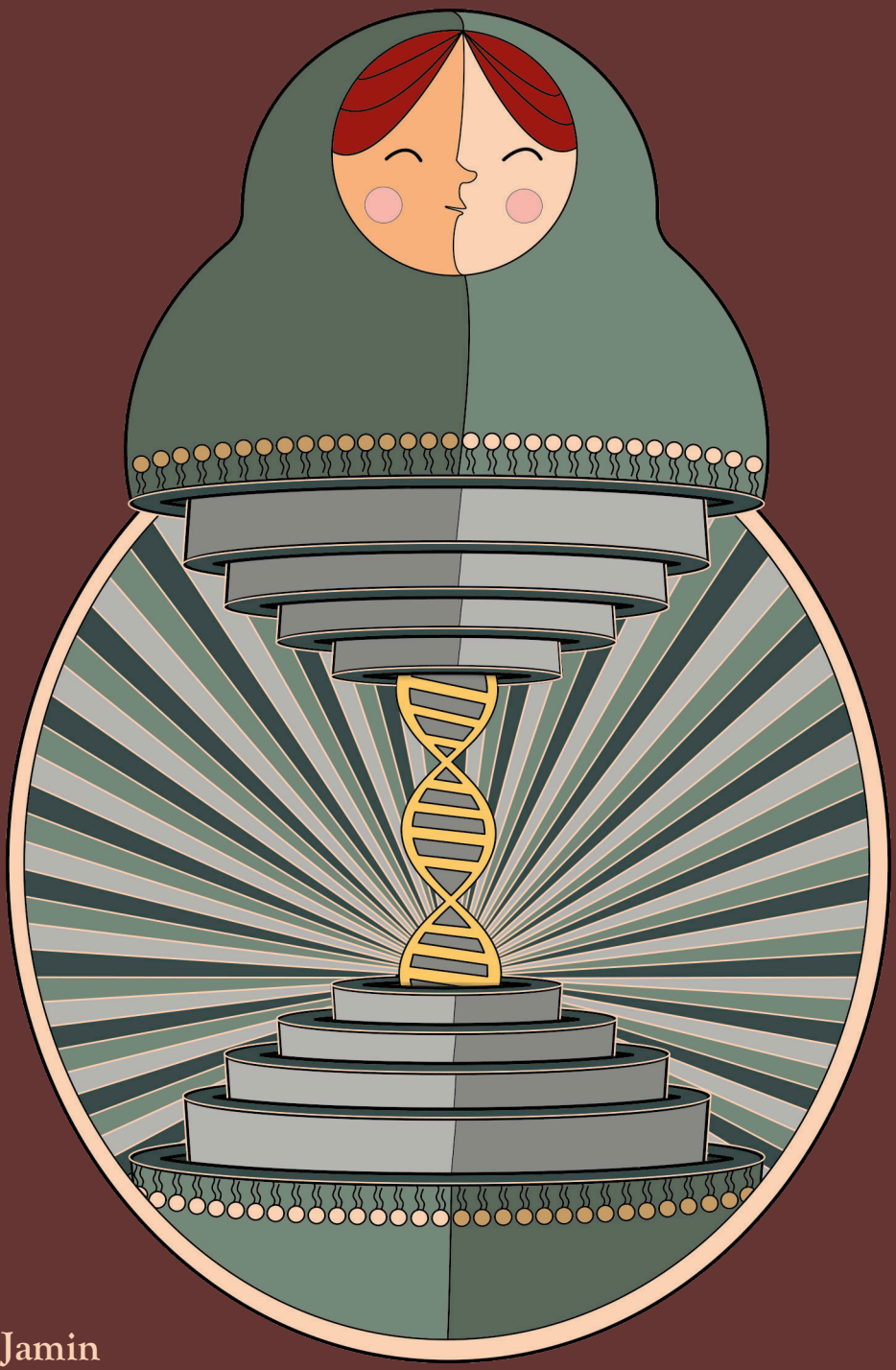
The molecular-matryoshka phenomenon

The Molecular Matryoshka Phenomenon

Peeling the layers of spread of antimicrobial resistance genes



Casper Jamin

THE MOLECULAR-MATRYOSHKA PHENOMENON
PEELING THE LAYERS OF SPREAD OF ANTIMICROBIAL RESISTANCE GENES

Printing of this thesis was financially supported by the Netherlands Society of Medical Microbiology (NVMM) and the Royal Netherlands Society for Microbiology (KNVM)

The molecular-Matryoshka phenomenon
Peeling the layers of spread of antimicrobial resistance genes
2023, Casper Jamin

Layout by Textcetera, Den Haag
Cover art by Mayk Luchessi

THE MOLECULAR-MATRYOSHKA PHENOMENON

PEELING THE LAYERS OF SPREAD OF ANTIMICROBIAL RESISTANCE GENES

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Maastricht,
op gezag van de Rector Magnificus, Prof. dr. Pamela Habibović,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op woensdag 29 Maart 2023 om 16:00 uur

door

Casper Jamin

Supervisors:	Prof. dr. Paul H.M. Savelkoul
	Prof. dr. Christian J.P.A Hoebe
Co-supervisor:	Dr. Lieke B. van Alphen
Assesment Committee:	Prof. dr. C.D.A. Stehouwer (chair)
	Prof. dr. R. Peters
	Prof. dr. C. Schultsz, Amsterdam UMC
	Dr. A. Schürch, Universitair medisch Centrum Utrecht

TABLE OF CONTENTS

1

GENERAL INTRODUCTION

Antimicrobial resistance: A global problem

Antimicrobial resistance (AMR) has become a global problem. In 2015 the World Health Organisation (WHO) has declared AMR as an immediate threat to mankind, requiring urgent and harmonized efforts to curb moving into a post-antibiotic age, where previously treatable common infections become untreatable. This will have serious consequences, such as a higher healthcare requirement, increased morbidity and pre-mature death[1–3]. Even though the impact of AMR seems to have evolved in the past decades, antibiotic resistance is a natural defense mechanism of bacteria present on our planet for already millions of years[4]. Microorganisms have been in a constant arms race to defend and fight for their respective niche in nature. Therefore, it's beneficial for them to assert their space within these niches by outcompeting others. This outcompeting can be achieved by the production of specific compounds like toxins and antimicrobial agents which inhibit the growth or kill other microorganisms. Discovery of these compounds led to the further development, improvement and industrial production of antibiotics for clinical treatment of infections.

In more recent years, different specific AMR bacterial species have taken the stage worldwide, such as methicillin resistant *Staphylococcus aureus* (MRSA), vancomycin resistant *Enteroccocus* (VRE), extended spectrum beta-lactamase (ESBL) producing *Enterobacterales*, and carbepenemase producing *Enterobacterales* (CPE). These microorganisms are often associated within a nosocomial setting, and together with *Pseudomonas aeruginosa* and *Acinetobacter baumannii* they are part of the group of so called ESKAPE pathogens (*Enterococcus, Staphylococcus, Klebsiella, Acinetobacter Pseudomonas* and *Enterobacter*)[5]. When these pathogens are resistant to antibiotics, infections with these bacteria are associated with doubling the chance of adverse outcomes for patients, which includes increased hospital stay, morbidity and mortality[3]. Spread of these resistant microorganisms can also require the implementation of strict infection control measures and even closure of units in the hospitals[6]. Recently, it has been estimated that 1,27 million deaths are attributable to AMR bacteria in 2019 worldwide[7]. Therefore, urgency is required to combat the increasing problem of AMR. On the contrary, the development of new antimicrobial agents has unfortunately been outpaced by the emergence of AMR in the 20th and 21st century[8]. This hampers further moving to a sustainable solution against AMR. As a consequence, the WHO has outlined a global action plan to combat AMR[1]. This plan is outlined by five objectives: I) improve understanding and awareness of AMR. II) Strengthen knowledge through research and surveillance. III) decrease the number of infections, by proper infection prevention measures. IV) optimize the use of antibiotics and V) ensure a way to improve sustainable investments for developments of new drugs, diagnostics, and vaccines to combat AMR.

Mechanisms for resistance

To investigate the emergence of AMR it is first important to understand the mechanisms behind antibiotic resistance. In general, four mechanisms of actions can be recognized (Figure 1). Firstly, modification/inactivation of the attacking antibiotic molecule, where bacteria remain susceptible to the original antibiotic, but the modified antibiotic can no longer reach and inhibit its target or the antibiotic molecule is degraded. Examples are the hydrolysis of the beta-lactam ring by beta-lactamases[9] or the modification of aminoglycosides by aminoglycoside acetyltransferases[10]. Secondly, altering flux of the antibiotic compound in or out of the bacterial cell, by either mitigating influx, e.g. by porin loss, so the antibiotic cannot enter the cell or by increasing efflux of antibiotics out of the cells by efflux pumps such as the tetracycline efflux pump[11,12]. Thirdly, target-site modification, such as the vancomycin resistance gene cluster *vanA* and *vanB*[13], where the D-alanyl-D-alanine terminus of the peptidoglycan layer is replaced by a D-alanyl-D-lactate. It is this lactate group for which vancomycin no longer has affinity and prevents maturation of the peptidoglycan layer. Fourthly, antibiotic target amplification by gene duplication, a phenomenon often associated with IS26 elements[14–16].

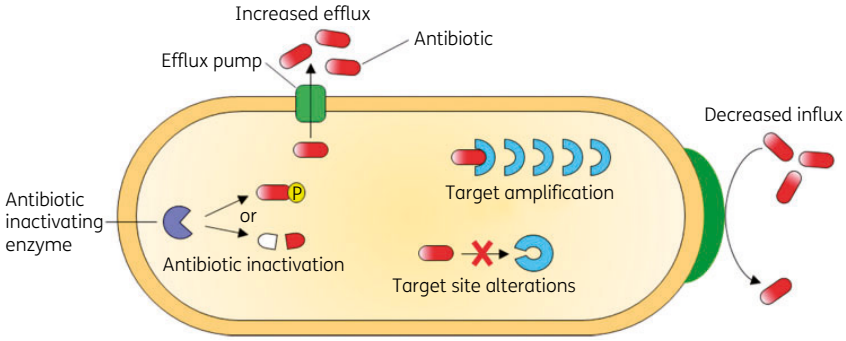


Figure 1. Mechanisms for antibiotic resistance. I) modification or inactivation of the antibiotic II) altering flux in or out of the cell. III) modifications of the target-site of the antibiotic, this can also include the shielding of the target site by other molecules. IV) target amplification by gene duplication. Image, with permission, used from [12].

How do bacteria become resistant?

AMR can either be a natural or an acquired trait in bacterial cells[17]. Natural resistance can either be intrinsic, e.g. antibiotics cannot enter a bacterium because of its default cell wall properties. Or natural AMR can be induced, such as common efflux pumps in some bacteria[18]. One way of acquiring AMR, is by the introduction of mutations in the genes targeted by the antibiotic. These mutations lead to an alteration of the subsequent protein or ribosomal complex, affecting the binding affinity of the antibiotic[19].

A potent way of acquiring of AMR is typically achieved by exchange of genetic material between bacterial cells. This is achieved through exchange of genetic material by horizontal gene transfer (HGT)[17]. HGT consists of 4 main routes (Figure 2)[17]. A) Conjugation of plasmids, Integrative & mobilizable elements (IME)[20], Integrative & conjugative elements (ICE, previously known as conjugative transposons)[21], B) Natural transformation by competent bacteria to take up DNA from their environment[22], C) Transduction by bacteriophages[23] and D) gene transfer agents[24].

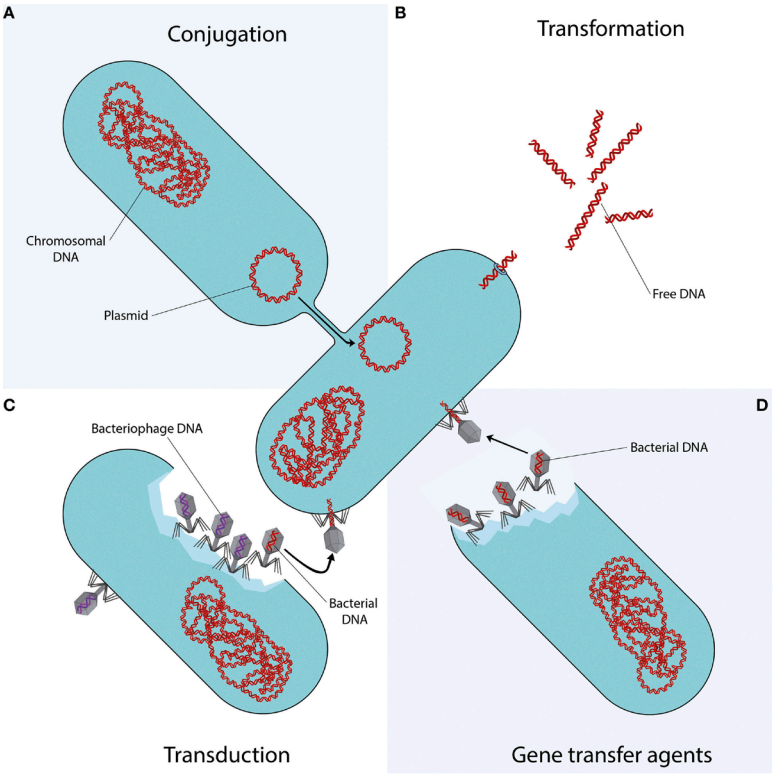


Figure 2. Four ways of horizontal gene transfer. A) Conjugation of plasmids. B) Uptake of free DNA by transformation. C) Transduction by bacteriophages. D) Gene transfer agents by bacteriophage-like particles. Image used from[17], with permission via the Creative Commons Attribution License (CC BY) associated with this article.

Dissecting the layers that contribute to the dissemination of AMR

As outlined in the by WHO's global action plan[1], one of the objectives to combat AMR is to improve research and surveillance of AMR. Therefore, it is paramount to better understand the different aspects (layers) responsible the spread of antibiotic resistance and how to investigate these. Throughout this thesis, three layers are considered. The first layer: The resistant bacterium itself can spread between persons, animals and their environment (clonal expansion of resistance). The second layer, AMR genes often resides on plasmids, which can be transferred to other bacteria, even across different genera of bacteria[25]. And the third layer, AMR genes are often part of transposons: genetic elements that can move within the chromosome, to and from plasmids and in the case of ICE (previously known as conjugative transposons) and IME, also to other bacteria.

Spread of antibiotic resistant bacteria

The **first molecular layer** of the dissemination of AMR is the spread of resistant bacterial isolates/clones. Specific bacterial strains can carry genetic markers for AMR either inherited vertically from a parental cell or acquired horizontally (Figure 2). Bacterial lineages evolve over time and may acquire additional AMR markers. Spread of specific resistant bacteria can be problematic in healthcare centers due to negative outcomes for patients. To curb the spread of (resistant) pathogens it is essential to be able to identify when and how these pathogens have spread.

Bacterial typing

The goal of bacterial typing is to be able to distinguish between different strains belonging to the same species. Thereby asserting if a particular strain might have spread. Two levels of interpretation should be considered for typing. On the one hand, the comparison of the actual bacterial isolates by classifying if these belong to the same clonal group. On the other hand, the interpretation of the epidemiological and/or clinical data from where these isolates originate from. If two bacterial isolates are considered identical and were derived from patients who were situated on the same department in the hospital, an outbreak of this strain has likely occured[26].

Bacterial typing

Prior to the introduction of bacterial whole genome sequencing (WGS), bacterial typing has been performed by a plethora of different methods[26,27]. Most, but not all of these methods are either based on band-based- and sequence based methods[26]. Pulsed field gel electrophoresis (PFGE). Multi locus variable-number tandem repeat analysis (MLVA) and amplified fragment length polymorphism (AFLP) rely on

the restriction of genomic DNA (and subsequent amplification for AFLP) and visualization of banding patterns. For sequence-based typing such as multi locus sequence typing (MLST), a handful, usually seven housekeeping genes are amplified by PCR and subsequently sequenced using Sanger-sequencing based methods. Each allele of these seven loci is assigned a unique number which is coupled to the unique sequence of this allele. Based on the combination of these seven alleles, a sequence type (ST) is assigned. Compared to WGS several disadvantages should be considered. I) discriminatory power (all). II) low throughput (PFGE)[26]. III) assay needs to be optimized for each species (MLVA).

Advantages of WGS

Compared to previous DNA based methods WGS has several advantages. I) ease of data-sharing of consensus genomes or core genome MLST (cgMLST) or whole genome MLST (wgMLST) profiles, as the data footprint of these files are relatively low. II) genotypic information on antibiotic resistance and virulence factors can be determined from the genome. III) Theoretically, the highest discriminatory power compared to other methods, as all information of the genome is used.

Disadvantages of WGS

However, a few disadvantages of WGS should be pointed out. I) capital investment of DNA sequencers. II) Relatively long time to result. A single sequencing run can take 24 to 48 hours, before data analysis even takes place, without even considering the hands-on time required of DNA isolation and the preparation of DNA to sequencing libraries, required by the platform for DNA sequencing. III) trade-off between either short, high accurate sequence reads or inaccurate long reads. IV) Bioinformatic expertise required for data analysis. V) data-storage of large raw sequencing files. VI) resolving entire genomes & plasmids require either an inaccurate long-read sequencing or a hybrid approach combining short accurate reads with the long inaccurate reads, further increasing costs.

Determining clonality using WGS

Using WGS, the clonality between strains can be determined by either allele based approaches, or by single nucleotide polymorphisms (SNP) based approaches[28,29]. Allele based methods for WGS based typing have a lot in common with normal multi-locus sequencing typing (MLST). For WGS based bacterial typing using alleles, a similar approach is taken. First, the genome needs to be reconstructed by bioinformatic algorithms that perform *de novo* assembly. Then, all coding sequences (loci) are extracted and each individual locus is assigned the unique allele number based on the sequence. Now instead of seven alleles, bacterial isolates are compared based on the core genes (called cgMLST), which utilizes 200 to 2000 conserved alleles or

all genes (called wgMLST), which uses next to the cgMLST alleles, also all the more variable accessory genes. For wgMLST up to 6000 alleles are included, depending on the species[28].

A different approach for bacterial typing is based on identifying the number of SNPs in the genome. SNP based approaches can be performed using I) reference genome alignment, II) genome to genome alignment or III) alignment free approaches using k-mers, where genomes are analysed for their 'split k-mer' (a subsequence of DNA of length k, with a variable middle base pair) content [30,31]. If two genomes have the same split k-mer, the middle basepair is examined. If this basepair is different between the two genomes, it is considered a SNP between the two genomes. All these methods rely on SNPs throughout the genome. By comparing if isolates have the same SNPs on specific locations, a genetic distance can be determined which is used to assess if two isolates are clonal or not. Compared to cgMLST and wgMLST, SNP based approaches generally have a higher resolution for phylogenetic reconstruction as, in theory, all base pairs in the bacterial genome can be evaluated. However, having to manually pick a reference genome and data-storage footprint make these SNPs methods less versatile than cg/wgMLST based methods.

Spread of AMR genes via mobile genetic elements

The spread of antibiotic resistance can also occur via mobile genetic elements such as plasmids, transposons and (conjugative) transposons (now known as Integrative & conjugative elements: ICE[32]).

Plasmids

The **second molecular layer** driving the spread of AMR are plasmids. Plasmids are independent genetic entities which encode for their own replication machinery and use the origin of replication (ORI) as an initiation site for self-replication. Genes encoded on plasmids can augment the bacteria by adjusting their metabolic capacity, increase resistance to antibiotics[33] and (heavy) metals[34], and improve pathogenicity and virulence capacity[35]. So, bacteria can benefit from carrying plasmids, but the replication of these plasmids also can come at a metabolic cost for the bacterium. Genes expressed from these plasmids and the replication of these elements require resources from the bacterial cell, which could otherwise be used for the bacterial cell itself. Whether plasmids are maintained within a bacterial cell is thus dependent on the benefit/cost ratio of the genes carried on this plasmid. Plasmids occur in varying sizes and can be present in one or more copies per cell, further influencing the level of expression of their cargo genes[36–38].

Although more difficult to identify, plasmids have been the cause of outbreaks of resistant bacteria in hospitals, where multiple different bacterial species were involved expressing the same resistant (plasmid derived) phenotype like carbapen-

emase producing *Enterobacterales*[39–42]. Not surprisingly, these kinds of outbreaks will usually only be identified when a high increase occurs in a normally low prevalent environment in hospitals.

Before the widespread use of WGS, plasmids could already be typed using a replicon specific PCR-Based replicon typing (PBRT)[43]. All plasmids have an origin of replication, which can be categorized in several incompatibility groups. These incompatibility groups are distinguished based on the phenomenon that multiple plasmids of a same incompatibly group are unable to co-exist within one bacterial cell. Since the introduction of WGS, also plasmid MLST (pMLST, which utilizes a handful of genes for each plasmid type) and replicon typing can now be performed directly from sequence data, facilitating the identification of various and multiple plasmids in a bacterial host[44]. Unfortunately, pMLST is only available for a few replicon types[44].

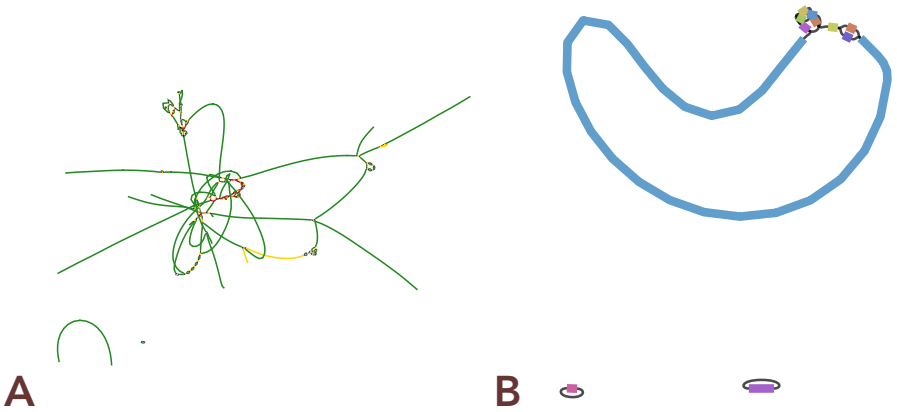


Figure 3A. Example of de novo assembly graph of *Klebsiella pneumoniae* using only paired end short reads. No clear distinction can be made between chromosomal contigs and plasmid contigs. **Figure 3B.** De novo assembly graph of the same *Klebsiella pneumoniae* using short and long reads together to generate a hybrid assembly. Separate genomic elements (1 chromosome and 2 plasmids) are clearly visible.

Horizontal gene transfer by conjugation of plasmids is a general way of bacterial cells for sharing gene content (Figure 2A). In this process bacterial cells connect to each other via a pilus from the donor cell. Via this pilus, plasmid DNA is transferred and subsequently after recirculation and off-strand synthesis, the recipient cell can utilize genes encoded on the plasmid. For conjugation to occur, the plasmid itself encode for type IV secretion systems, or can utilize these from other plasmids or from the chromosome[45,46]. Plasmids can also undergo recombination, where, for example, two plasmids can combine into a large multi-ORI or hybrid plasmids. These can be the product of recombination or transposition events caused by insertion sequences, such as the *IS6/IS26* family[47,48]. These hybrid plasmids, also called cointegrates, can subsequently be resolved by homologous recombination in *RecA* competent

bacteria[47,49,50] into separate plasmids that differ from the original plasmids before recombination. Moreover, these cointegrate genomic structures can remain stable and thus unresolved for long periods of time (Dr. Christopher Harmer, personal communication). Because of the high fluidity of gene content on some plasmids, it becomes increasingly difficult to track these MGEs in a more clinical setting. One of the main problems with resolving plasmid-genomic content is the presence of repeating sequences throughout the bacterial genome, such as insertion sequences (IS)[51]. Multiple copies of IS sequences can be present on both the chromosome as on plasmids. These IS sequences are generally longer (>1kb) than sequencing reads (150-500bp) generated from the traditional NGS sequencing systems (e.g. Illumina- and IonTorrent platforms). Therefore, subsequent *de novo* assemblers are unable to rebuild genomes and plasmids around these elements (Figure 3). Using long-read sequencing (e.g. Oxford Nanopore- or PacBio's single-molecule real-time sequencing) the reads (> 10kb) will span the complete IS genes (or other repeating regions) and neighboring regions, so the location of these IS can be uniquely pinpointed in the genome and the genomes can be completely circularized.

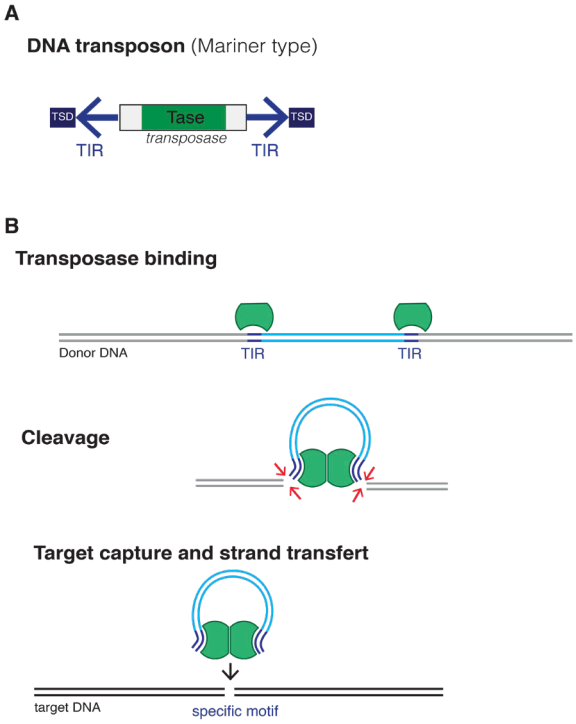


Figure 4A. Genetic structure of a DNA transposon. B) Transposition visualized by transposases which bind terminal inverted repeats (TIR), excise DNA and move to a new location in the genome for insertion. Image used from https://en.wikipedia.org/wiki/Transposable_element#/media/File:DNA_Transposon.png.

Insertion sequences and transposons

The **third molecular layer** in the dissemination of antimicrobial resistance is the excision and re-insertion of AMR genes by transposons and their insertion sequences (IS). Transposons, also known as 'jumping genes' are genetic elements which can move throughout the genome by excising itself and nearby cargo genes and pasting itself on a new location in the genome. Two categories of transposons exist: I) class I transposons, or retrotransposons which forms RNA intermediates during transposition (copy-paste transposition) and II) class II transposons, or DNA transposons (cut and paste transposition). Class I transposons are mainly found in eukaryotes and class II is found in bacteria, archaea and eukaryotes[52]. Of main interest in this thesis here, are the DNA transposons as these are responsible for the dissemination of antibiotic resistance in bacteria (Figure 4). One of the main drivers behind the translocation of AMR genes are transposases of the family *IS6*[53]. This includes *IS26*, a transposase element which is widely disseminated in *Enterobacterales* and a key player in the spread of AMR genes. Simple transposition of AMR genes (cut and paste) by IS elements can occur via linear and circular intermediates. However, recently it was discovered that *IS26* does not seem to move by itself alone and take along flanking DNA and has an extreme tendency to co-localize (>50x likely than random insertion) onto other *IS26*, creating arrays with multiple AMR genes, also sometimes called resistance gene islands[47,49,54,55] utilizing so called translocatable units (TU). These TUs are small, circular DNA molecules only encoding the *IS26* and some cargo genes such as AMR genes. These TUs can subsequently reinsert somewhere in the genome (Figure 5). In general, the role of IS sequences should not be underestimated in jumping of AMR genes in the genome to and from plasmids and subsequent spread to other bacterial species.

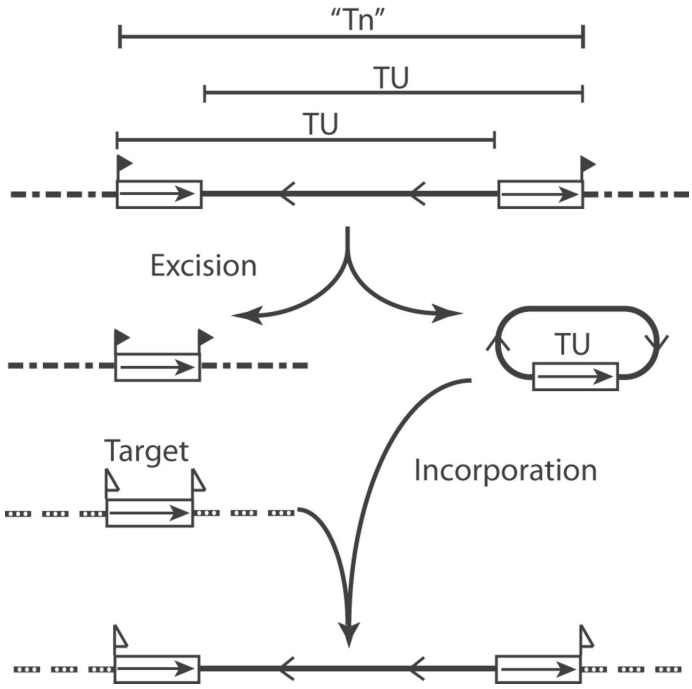


Figure 5. IS26 mediated formation of translocatable units (TU) and subsequent preferential insertion on another IS26. The translocatable unit can contain AMR genes which lay next to the IS26.

Scope & outline of this thesis

The main goal of this thesis is to study the different layers contributing to the spread of antimicrobial resistance in *Enterobacterales* species, using whole genome sequencing. For continuous patient care and optimal treatment, it is important to determine which AMR genes are important and cause a resistant phenotype in bacteria and how this resistance is transferred to other bacteria. In this thesis three molecular layers are classified: 1) the bacterial DNA of the isolate, 2) plasmids in bacteria and 3) independent transposons and other mobile genetic elements that either reside on plasmids or on the chromosome of bacteria. This nested nature of AMR genes resembles the way Russian nesting dolls or Matryoshka dolls can be stacked inside each other on a molecular-scale. Furthermore, multiple sets of these dolls can be rearranged with each other, recombining into new unique sets, much like bacteria share their genetic content via conjugation of plasmids and 'Integrative & conjugative elements' (ICE).

In **chapter 2**, the outcome of harmonizing WGS for outbreak typing of bacterial isolates was assessed by means of a three-center ring trial (layer 1). In **chapter 3**, a proficiency test was performed by thirteen Dutch health-care centers to determine if and how the use of different bioinformatic workflows may impact the interpretation of identified outbreak clusters (layer 1). In **chapter 4** we describe WGS datasets generated for the benchmarking of bioinformatic tools for AMR gene identification. In **chapter 5**, the prevalence and relevance for the clinic of newly described *mcr-9* colistin resistance gene was determined**.** In **chapter 6,** Mobile colistin resistance *mcr-1* encoding plasmids derived from chicken retail meat were analyzed and compared to clinically derived *mcr-1* plasmids retrieved from public databases (layer 2)**. In chapter 7,** we describe the appearance of a mcr-4 encoding plasmids in animals and humans in The Netherlands (layer 2). Finally, in **chapter 8**, all three layers that compose the dissemination of antibiotic resistance come back. In this chapter the spread of a resistance gene island is described in hospitals in Limburg, The Netherlands. This resistance gene island was encoded on plasmids and on the chromosome. The spread of isolates and plasmids were both identified.

References

1. Global action plan on antimicrobial resistance. https://www.who.int/antimicrobial-resistance/global-action-plan/development_process/en/.
2. Livermore, D. M. Bacterial resistance: Origins, epidemiology, and impact. *Clin. Infect. Dis.* **36**, (2003).
3. Friedman, N. D., Temkin, E. & Carmeli, Y. The negative impact of antibiotic resistance. *Clinical Microbiology and Infection* vol. 22 (2016).
4. Barlow, M. & Hall, B. G. Phylogenetic analysis shows that the OXA β-lactamase genes have been on plasmids for millions of years. *J. Mol. Evol.* **55**, (2002).
5. Pendleton, J. N., Gorman, S. P. & Gilmore, B. F. Clinical relevance of the ESKAPE pathogens. *Expert Review of Anti-Infective Therapy* vol. 11 (2013).
6. Girmenia, C. *et al.* Infections by carbapenem-resistant Klebsiella pneumoniae in SCT recipients: A nationwide retrospective survey from Italy. *Bone Marrow Transplant.* **50**, (2015).
7. Murray, C. J. *et al.* Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* **6736**, (2022).
8. Årdal, C. *et al.* Antibiotic development — economic, regulatory and societal challenges. *Nat. Rev. Microbiol.* **18**, 267–274 (2020).
9. Majiduddin, F. K., Materon, I. C. & Palzkill, T. G. Molecular analysis of beta-lactamase structure and function. *Int. J. Med. Microbiol.* **292**, (2002).
10. Vong, K. & Auclair, K. Understanding and overcoming aminoglycoside resistance caused by N-6′-acetyltransferase. *MedChemComm* vol. 3 (2012).
11. Roberts, M. C. Tetracycline resistance determinants: Mechanisms of action, regulation of expression, genetic mobility, and distribution. *FEMS Microbiology Reviews* vol. 19 (1996).
12. Alav, I., Sutton, J. M. & Rahman, K. M. Role of bacterial efflux pumps in biofilm formation. *J. Antimicrob. Chemother.* **73**, (2018).
13. Courvalin, P. Vancomycin resistance in gram-positive cocci. *Clinical Infectious Diseases* vol. 42 (2006).
14. Shropshire, W. C. *et al.* IS26-mediated amplification of blaOXA-1and blaCTX-M-15with concurrent outer membrane porin disruption associated with de novo carbapenem resistance in a recurrent bacteraemia cohort. *J. Antimicrob. Chemother.* **76**, (2021).
15. Bontron, S. *et al.* Increased resistance to carbapenems in proteus mirabilis mediated by amplification of the blaVIM-1-carrying and IS26-associated class 1 integron. *Microb. Drug Resist.* **25**, (2019).
16. Harmer, C. J., Lebreton, F., Stam, J., Mcgann, P. T. & Hall, R. M. Mechanisms of IS 26 -Mediated Ampli fi cation of the aphA1 Gene Leading to Tobramycin Resistance in an Acinetobacter baumannii Isolate.
17. Von Wintersdorff, C. J. H. *et al.* Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer. *Front. Microbiol.* **7**, 1–10 (2016).
18. Cox, G. & Wright, G. D. Intrinsic antibiotic resistance: Mechanisms, origins, challenges and solutions. *International Journal of Medical Microbiology* vol. 303 (2013).
19. C Reygaert, W. An overview of the antimicrobial resistance mechanisms of bacteria. *AIMS Microbiol.* **4**, 482–501 (2018).
20. Guédon, G., Libante, V., Coluzzi, C., Payot, S. & Leblond-Bourget, N. The obscure world of integrative and mobilizable elements, highly widespread elements that pirate bacterial conjugative systems. *Genes (Basel).* **8**, (2017).
21. Burrus, V. & Waldor, M. K. Shaping bacterial genomes with integrative and conjugative elements. *Research in Microbiology* vol. 155 (2004).
22. Domingues, S., Nielsen, K. M. & da Silva, G. J. Various pathways leading to the acquisition of antibiotic resistance by natural transformation. *Mob. Genet. Elements* **2**, (2012).
23. Brown-Jaque, M., Calero-C??ceres, W. & Muniesa, M. Transfer of antibiotic-resistance genes via phage-related mobile elements. *Plasmid* vol. 79 1–7 (2015).
24. Marrs, B. Genetic recombination in Rhodopseudomonas capsulata. *Proc. Natl. Acad. Sci. U. S. A.* **71**, (1974).
25. Acman, M., van Dorp, L., Santini, J. M. & Balloux, F. Large-scale network analysis captures biological features of bacterial plasmids. *Nat. Commun.* **11**, 1–11 (2020).
26. van Alphen, L. B., von Wintersdorff, C. J. H. & Savelkoul, P. H. M. Epidemiological Typing Using WGS. in *Application and Integration of Omics-powered Diagnostics in Clinical and Public Health Microbiology* (2021). doi:10.1007/978-3-030-62155-1_5.
27. Li, W., Raoult, D. & Fournier, P. E. Bacterial strain typing in the genomic era. *FEMS Microbiology Reviews* vol. 33 (2009).
28. Schürch, A. C., Arredondo-Alonso, S., Willems, R. J. L. & Goering, R. V. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene–based approaches. *Clinical Microbiology and Infection* (2018) doi:10.1016/j.cmi.2017.12.016.

29. Quainoo, S. *et al.* Whole-genome sequencing of bacterial pathogens: The future of nosocomial outbreak analysis. *Clinical Microbiology Reviews* vol. 30 (2017).
30. Gardner, S. N., Slezak, T. & Hall, B. G. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* (2015) doi:10.1093/bioinformatics/btv271.
31. Harris, S. R. SKA: Split Kmer Analysis Toolkit for Bacterial Genomic Epidemiology. *bioRxiv* (2018) doi:10.1101/453142.
32. Wozniak, R. A. F. & Waldor, M. K. Integrative and conjugative elements: Mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nature Reviews Microbiology* vol. 8 (2010).
33. Bennett, P. M. Plasmid encoded antibiotic resistance: Acquisition and transfer of antibiotic resistance genes in bacteria. *Br. J. Pharmacol.* **153**, 347–357 (2008).
34. Baker-Austin, C., Wright, M. S., Stepanauskas, R. & McArthur, J. V. Co-selection of antibiotic and metal resistance. *Trends Microbiol.* **14**, 176–182 (2006).
35. Thorne, C. B. Bacillus anthracis. *BSL3 BSL4 Agents Epidemiol. Microbiol. Pract. Guidel.* **2**, 291–297 (1993).
36. van Mastrigt, O., Mager, E. E., Jamin, C., Abee, T. & Smid, E. J. Citrate, low pH and amino acid limitation induce citrate utilization in Lactococcus lactis biovar diacetylactis. *Microb. Biotechnol.* **11**, (2018).
37. Millan, A. S. *et al.* Small-plasmid-mediated antibiotic resistance is enhanced by increases in plasmid copy number and bacterial fitness. *Antimicrob. Agents Chemother.* **59**, (2015).
38. Cook, L. C. & Dunny, G. M. Effects of biofilm growth on plasmid copy number and expression of antibiotic resistance genes in Enterococcus faecalis. *Antimicrob. Agents Chemother.* **57**, (2013).
39. Ledda, A. *et al.* Hospital outbreak of carbapenem-resistant Enterobacteriales associated with an OXA-48 plasmid carried mostly by Escherichia coli ST399. *bioRxiv* 1–30 (2020) doi:10.1101/2020.06.15.148189.
40. Adhiratha Boonyasiri, Ashleigh C. Myall, Yu Wan, Frances Bolt, Alice Ledda, Siddharth Mookerjee, Andrea Y. Weiße, Jane F. Turton, Hala Abbas, Ruta Prakapaite, Akshay Sabnis, Alireza Abdolrasouli, Kenny Malpartida-Cardenas, Luca Miglietta, Hugo Donaldson, F. D. Integrated patient network and genomic plasmid analysis reveal a regional, multi-species outbreak of carbapenemase-producing Enterobacterales carrying both blaIMP and mcr-9 genes. *medRvix* (2021).
41. Yamagishi, T. *et al.* A prolonged multispecies outbreak of IMP-6 carbapenemase-producing Enterobacterales due to horizontal transmission of the IncN plasmid. *Sci. Rep.* **10**, (2020).
42. De Man, T. J. B. *et al.* Multispecies Outbreak of Verona Integron-Encoded Metallo-ß-Lactamase-Producing Multidrug-Resistant Bacteria Driven by a Promiscuous Incompatibility Group A/C2 Plasmid. *Clin. Infect. Dis.* **72**, (2021).
43. Carattoli, A. Plasmids in Gram negatives: Molecular typing of resistance plasmids. *International Journal of Medical Microbiology* vol. 301 (2011).
44. Carattoli, A. *et al.* In Silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* (2014) doi:10.1128/AAC.02412-14.
45. Llosa, M., Gomis-Rüth, F. X., Coll, M. & De la Cruz, F. Bacterial conjugation: A two-step mechanism for DNA transport. *Mol. Microbiol.* **45**, (2002).
46. Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C. & de la Cruz, F. Mobility of Plasmids. *Microbiol. Mol. Biol. Rev.* **74**, (2010).
47. Harmer, C. J., Moran, R. A. & Hall, R. M. Movement of IS26-Associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS26 and preferentially inserts adjacent to another IS26. *MBio* **5**, (2014).
48. He, S. *et al.* Insertion sequence IS26 reorganizes plasmids in clinically isolated multidrug-resistant bacteria by replicative transposition. *MBio* **6**, (2015).
49. Harmer, C. J. & Hall, R. M. Targeted conservative formation of cointegrates between two DNA molecules containing IS26 occurs via strand exchange at either IS end. *Mol. Microbiol.* **106**, (2017).
50. He, D. D. *et al.* Antimicrobial resistance-encoding plasmid clusters with heterogeneous MDR regions driven by IS26 in a single Escherichia coli isolate. *J. Antimicrob. Chemother.* **74**, 1511–1516 (2019).
51. Paszkiewicz, K. & Studholme, D. J. De novo assembly of short sequence reads. *Brief. Bioinform.* **11**, (2010).
52. Babakhani, S. & Oloomi, M. Transposons: the agents of antibiotic resistance in bacteria. *Journal of Basic Microbiology* vol. 58 (2018).
53. Varani, A., He, S., Siguier, P., Ross, K. & Chandler, M. The IS6 family, a clinically important group of insertion sequences including IS26. *Mob. DNA* **12**, 1–18 (2021).
54. Harmer, C. J. & Hall, R. M. IS26 cannot move alone. *J. Antimicrob. Chemother.* **76**, (2021).
55. Harmer, C. J., Pong, C. H. & Hall, R. M. Structures bounded by directly-oriented members of the IS26 family are pseudo-compound transposons. *Plasmid* vol. 111 (2020).

2

HARMONISATION OF WHOLE GENOME SEQUENCING FOR OUTBREAK SURVEILLANCE OF *ENTEROBACTERIACEAE* AND *ENTEROCOCCI*

Casper Jamin[1], Sien De Koster[2], Stefanie van Koeveringe[3], Dieter De Coninck[4], Klaas Mensaert[4], Katrien De Bruyne[4], Natascha Perales Selva[3], Christine Lammens[2], Herman Goossens[2,3], Christian Hoebe[1,5], Paul Savelkoul[1], Lieke van Alphen[1], on behalf of the i-4-1-Health Study Group

1. Department of Medical Microbiology, Care and Public Health Research Institute (CAPHRI), Maastricht University Medical Center+, Maastricht, The Netherlands.
2. Laboratory of Medical Microbiology, Vaccine & Infectious Disease Institute, University of Antwerp, Belgium
3. Laboratory of Clinical Microbiology, Antwerp University Hospital, Antwerp, Belgium
4. bioMérieux SA, Sint-Martens-Latem, Belgium
5. Department of Sexual Health, Infectious Diseases and Environment, South Limburg Public Health Service, Heerlen, The Netherlands

Published: **Jamin, C.,** De Koster, S., van Koeveringe, S., De Coninck, D., Mensaert, K., De Bruyne, K., ... & van Alphen, L. (2021). Harmonization of whole-genome sequencing for outbreak surveillance of *Enterobacteriaceae* and *Enterococci*. Microbial *genomics, 7*(7).

Abstract

Introduction

Whole genome sequencing (WGS) is becoming the de facto standard for bacterial typing and outbreak surveillance of resistant bacterial pathogens. However, interoperability for Whole Genome Sequencing (WGS) of bacterial outbreaks is poorly understood.

Hypothesis/gap statement

We hypothesized that harmonisation of WGS for outbreak surveillance is achievable through the use of identical protocols for both data generation and data analysis.

Aim

We assessed if inter-laboratory harmonisation of WGS for outbreak surveillance is achievable.

Methodology

A set of 30 bacterial isolates, comprising of various species belonging to the *Enterobacteriaceae* family and *Enterococcus* genera, were selected and sequenced using the same protocol on the Illumina MiSeq platform in each individual centre. All generated sequencing data were analysed by one centre using BioNumerics (6.7.3) for i) genotyping origin of replications & antimicrobial resistance genes, ii) core-genome Multi-locus sequence typing (cgMLST) for *Escherichia coli* and *Klebsiella pneumoniae* & whole-genome multi-locus sequencing typing (wgMLST) for all species. Additionally, a split k-mer analysis was performed to determine the number of SNPs between samples.

Results

A precision of 99.0% and an accuracy of 99.2% was achieved for genotyping. Based on cgMLST, a discrepant allele was called only in 2/27 and 3/15 comparisons between two genomes, for *E. coli* and *K. pneumoniae,* respectively. Based on wgMLST, the number of discrepant alleles ranged from 0 to 7 (average 1.6). For SNPs, this ranged from 0-11 SNPs (average 3.4). Furthermore, we demonstrate that using different *de novo* assemblers to analyse the same dataset introduces up to 150 SNPs, which surpasses most thresholds for bacterial outbreaks.

Conclusion

This shows the importance of harmonisation of data processing surveillance of bacterial outbreaks. In summary, multi-centre WGS for bacterial surveillance is achievable, but only if protocols are harmonised.

Impact statement

Whole Genome Sequencing (WGS) for typing bacterial outbreaks has surged in recent years. We performed an inter-laboratory ring-trial by sending out 30 bacterial isolates to assess the reproducibility of WGS. We demonstrated that the use of different *de novo* assemblers for a single outbreak analysis will lead to bacterial isolates being misclassified as not related to the outbreak cluster. Additionally, we show that implementing WGS for regional or (inter)national surveillance of bacterial pathogens is feasible if identical laboratory procedures and data analysis workflows are used.

Introduction

The dissemination of antimicrobial resistance (AMR) has grown to an issue of worldwide proportions. Routine surveillance by molecular typing can aid in the fight against AMR, as outlined by the global action plan of the World Health Organization[1]. ESKAPE pathogens (*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa and Enterobacter species*) are of major interest as they are the leading cause of hospital-related infections and outbreaks. Furthermore, reports show that the number of infections by resistant microorganisms have been on the rise in recent years. Infections by multidrug resistant (MDR) bacteria are associated with an increase in economic burden[2] and negative patient outcomes such as morbidity and mortality[3,4].

To determine the spread of resistance and of resistant microbes, different molecular typing methods are being applied. Older, established typing methods for outbreak surveillance, such as pulsed field gel electropheresis (PFGE), amplification fragment length polymorphism (AFLP), multi-locus sequencing typing (MLST) and multi-locus variable-number tandem repeat analysis (MLVA) are slowly being replaced by whole genome sequencing (WGS). The introduction of WGS to the field of bacterial typing and spread of AMR has set a new standard for discriminatory power and accuracy, as it encompasses a comprehensive view of the bacterial core and accessory genome. This gives rise to the possibility to determine clonal relatedness in a more discriminatory fashion, and at the same time provide data on resistance genes, plasmids and virulence-potential, which would otherwise require a combination of other methods[5–8]. Current methods to determine phylogeny are based on core/whole genome multi-locus sequence typing (cgMLST, wgMLST)[9,10] or single nucleotide polymorphisms (SNPs)[11–13].

Approaches like cgMLST and wgMLST determine the phylogeny among bacterial isolates based on differences in allelic profile in either the core genome or the entire genome, respectively. All coding sequences (CDS) or loci are identified using tools

such as Prodigal[14]. Then, all variants of each locus are assigned a unique allele number and the complete set of allele numbers is called the allelic profile. The genetic distance is calculated by counting the number of discrepant alleles between two isolates. A relative genetic distance can also be calculated by dividing the number of discrepant alleles by the number of alleles that were compared. Next to commercial packages for cgMLST and wgMLST analyses, such as BioNumerics or SeqSphere, open source options are available as well, such as ChewBBACA[10] and Enterobase[15].

Inferring phylogeny based on SNPs can be performed by three different methods. i) Alignment to a reference genome (Snippy[11]). ii) (core-) genome alignment (MAUVE[16] or Harvest Suite[17]). iii) alignment-free methods based on using the entire collection of subsequences of a sequence of length k: k-mer (kSNP[18] or SKA[13]).

Currently, only few studies have described clonal cluster thresholds definitions using cgMLST, wgMLST or SNP-based methods. Generally, these studies determine the thresholds based on either i) previous or ongoing bacterial outbreaks in hospitals and in the food production chain, or ii) by means of follow-up on human carriers of these pathogens over time. Furthermore, most of these studies only describe single clone outbreaks, which can hamper the interpretation when these thresholds are applied to different lineages of a specific species. Some clinically relevant lineages might be more clonal than others, and so require different thresholds. One of the first reports on the use of WGS for bacterial outbreak analysis were on methicillin-resistant *Staphylococcus aureus* (MRSA) in 2013, in a neonatal intensive care unit. Next to standard assessment of epidemiological data and antibiograms, WGS was performed to resolve this putative outbreak[19]. In that study, a maximum of 20 SNPs was observed among the MRSA isolates found in the outbreak. For the foodborne pathogen *E. coli* O157:H7, the Public Health Agency Canada evaluated WGS for outbreak detection[20]. To this end, they retrospectively performed WGS for 250 isolates, from eight different outbreaks and analyzed using wgMLST and SNP analyses. These 250 isolates were previously typed using MLVA or PFGE. WGS based typing was in excellent concordance with MLVA and PFGE and also had higher discriminatory power to resolve outbreak clusters. Additionally, they reported that all isolates from for each outbreak fell within a cutoff of 5 SNPs or 10 allele differences (on wgMLST basis). In their review, Schürch *et al.* suggested various clonal cluster thresholds based on wgMLST or SNP analyses for a few common bacterial pathogens in outbreak situations[9].

Kluytmans-van den Bergh *et al.* recently determined clonal-cutoffs based on cgMLST and wgMLST for four extended-spectrum beta-lactamase-producing *Enterobacteriaceae* (ESBL-E)*: Escherichia coli, K. pneumoniae*, *Citrobacter* species and *Enterobacter sp.*[21]. In their study, isolates were classified as epidemiologically linked when these were cultured from a single patient in a 30-day time window and when they belonged to the same seven-gene sequence type. Subsequently, the genetic distance (here defined as number of discrepant alleles divided by the num-

ber of alleles compared) was compared among all isolates, and clonal thresholds were determined by the lowest genetic distance possible that included all epidemiologically linked isolates.

The goal of the i-4-1-Health study is to assess the prevalence and spread of resistant bacteria among humans and animals in the Dutch-Belgian border[22]. Across a one-year period, we screened patients in hospitals and in long-term healthcare facilities, infants at day-care facilities, and broilers and weaned pigs for gut or rectal carriage of ESBL-producing, ciprofloxacin-resistant or carbapenemase-producing *Enterobacteriaceae* and vancomycin-resistant *Enterococci*. This One-Health approach could provide insights into the prevalence and spread of resistant bacteria between and within these separate domains. In the i-4-1 Health study, WGS data was generated in three independent locations, and thus inter-laboratory reproducibility needed to be assessed to allow the comparison of this data. To standardize the WGS results and interpretation, we made efforts to harmonise the WGS protocols, both for the wet-lab procedures and the bioinformatics analysis.

Here, we harmonised the inter-laboratory reproducibility of WGS for outbreak surveillance and genotyping of AMR and origin of replication (ORI) of plasmids for a selection of AMR bacteria frequently encountered in hospital-related infections and AMR surveillance within the I-4-1-Health project. As the implementation of WGS for routine outbreak surveillance is particularly dependent on standardized methodology, we evaluated the technical variation in phylogenetic comparison using a commercially available wgMLST tool in BioNumerics and an open-source reference-free SNP-based tool called SKA[13].

Materials and Methods

Selection of isolates

In total 30 resistant bacterial isolates were selected based on their extended-spectrum beta-lactamase (ESBL) or carbapenemase activity, or based on ciprofloxacin or vancomycin resistance phenotype. The complete collection of isolates consisted of nine *Escherichia coli*; five *K. pneumonia*; four *Citrobacter sp.*; four *Enterobacter sp.;* two *Klebsiella oxytoca;* two *Klebsiella aerogenes;* two *Enterococcus faecalis* and two *E. faecium*. Six isolates (two *E. coli,* two *K. pneumoniae* and two *Enterobacter sp.*) were collected previously[21] and kindly provided by the SoM study-group, and 20 isolates were collected during the i-4-1-Health study[22]. The *E. faecium* and *E. faecalis* isolates were from a previous collection, stored at Antwerp University. The isolates were collected from perianal swabs of hospitalized patients (21) and clients in nursing homes (6), and from feces from broilers (2) and weaned pigs (1) by selective culturing. The culturing methods are described elsewhere[21,22]. An overview of isolates and their

origin is available in table 1. Isolates were inoculated from -80°C on Mueller Hinton II agar (BD, Franklin Lakes, NJ, USA) and sent to the participating institutes. The 30 isolates were divided in three sets of ten isolates. Each set was sequenced once by each center, with a six-month interval between each set.

DNA isolation and WGS

The DNA isolation and WGS procedure was performed as follows: DNA was extracted using the MasterPure DNA isolation kit (Lucigen) or MasterPure Gram Positive DNA

Table 1. Metadata of all isolates used in this study

name	species	origin	study
Citrobacter sp. 1	*Citrobacter sp.*	hospital	i-4-1-health
Citrobacter sp. 2	*Citrobacter sp.*	long term healthcare facility	i-4-1-health
Citrobacter sp. 3	*Citrobacter sp.*	long term healthcare facility	i-4-1-health
Citrobacter sp. 4	*Citrobacter sp.*	hospital	i-4-1-health
Entembacter sp. 1	*Enterobacter sp.*	long term healthcare faciliy	i-4-1-health
Enterobacter sp. 2	*Enterobacter sp.*	hospital	i-4-1-health
Enterobactersp. 3	*Enterobacter sp.*	hospital	SoM
Enfetobacter sp. 4	*Enterobacter sp.*	hospital	SoM
E. coli 1	*E. coli*	hospital	SoM
E. coli 2	*E. coli*	hospital	SoM
E. coli 3	*E. coli*	hospital	i-4-1-health
E. coli 4	*E. coli*	hospital	i-4-1-health
E. coli 5	*E. coli*	broiler	i-4-1-health
E. coli 6	*E. coli*	weaned pig	i-4-1-health
E. coli 7	*E. coli*	long term healthcare faciliy	i-4-1-health
E. coli 8	*E. coli*	broiler	i-4-1-health
E. coli 9	*E. coli*	hospital	i-4-1-health
E. faecalis 1	*E. faecalis*	hospital	
E. faecalis 2	*E. faecalis*	hospital	
E. faecium 1	*E. faecium*	hospital	
E. faecium 2	*E. faecium*	hospital	
K. aerogenes 1	*E. aerogenes*	hospital	i-4-1-health
K. aerogenes 2	*E. aerogenes*	hospital	i-4-1-health
K. oxytoca 1	*K. oxytoca*	hospital	i-4-1-health
K. oxytoca 2	*K. oxytoca*	hospital	i-4-1-health
K. pneumoniae 1	*K. pneumoniae*	hospital	i-4-1-health
K. pneumoniae 2	*K. pneumoniae*	long term healthcare facility	i-4-1-health
K. pneumoniae 3	*K. pneumoniae*	long term healthcare facility	i-4-1-health
K. pneumoniae 4	*K. pneumoniae*	hospital	SoM
K. pneumoniae 5	*K. pneumoniae*	hospital	SoM

purification kit (Lucigen). Sequencing libraries were prepared using NexteraXT (Illumina). Libraries were sequenced on the Illumina MiSeq platform in paired end 2x250 base pairs (bp) reads using the MiSeq V2 cartridge. Where possible, each set of isolates was subjected to WGS in a single run. Acceptance criteria for WGS were a *de novo* assembly with an average coverage higher than 30 and less than 1000 contigs, as reported in BioNumerics (7.6.3). Samples not fulfilling acceptance criteria were re-sequenced. The accession numbers for the raw sequencing data are available in table 1. Analysis of the generated datasets (n=90) was performed in one institute.

county	accession number centre 1	accession number centre 2	accession number centre 3
Netherlands	ERS5219870	ERS5219871	ERS5219872
Netherlands	ERS5219873	ERS5219874	ERS5219875
Netherlands	ERS5219876	ERS5219877	ERS5219878
Netherlands	ERS5219879	ERS5219880	ERS5219881
Netherlands	ERS5219882	ERS5219883	ERS5219884
Netherlands	ERS5219885	ERS5219886	ERS5219887
Netherlands	ERS5219888	ERS5219889	ERS5219890
Netherlands	ERS5219891	ERS5219892	ERS5219893
Netherlands	ERS5219828	ERS5219829	ERS5219830
Netherlands	ERS5219831	ERS5219832	ERS5219833
Netherlands	ERS5219834	ERS5219835	ERS5219836
Netherlands	ERS5219837	ERS5219838	ERS5219839
Netherlands	ERS5219840	ERS5219841	ERS5219842
Netherlands	ERS5219843	ERS5219844	ERS5219845
Netherlands	ERS5219846	ERS5219847	ERS5219848
Netherlands	ERS5219849	ERS5219850	ERS5219851
Netherlands	ERS5219852	ERS5219853	ERS5219854
Belgium	ERS5219894	ERS5219895	ERS5219896
Belgium	ERS5219897	ERS5219898	ERS5219899
Belgium	ERS5219900	ERS5219901	ERS5219902
Belgium	ERS5219903	ERS5219904	ERS5219905
Belgium	ERS5219912	ERS5219913	ERS5219914
Belgium	ERS5219915	ERS5219916	ERS5219917
Belgium	ERS5219906	ERS5219907	ERS5219908
Netherlands	ERS5219909	ERS5219910	ERS5219911
Belgium	ERS5219855	ERS5219856	ERS5219857
Netherlands	ERS5219858	ERS5219859	ERS5219860
Netherlands	ER55219861	ER55219862	ERS5219863
Netherlands	ERS5219864	ERS5219865	ER35219866
Netherlands	ERS5219867	ER55219868	ERS5219869

Statistical analysis

Statistical analyses were done using scipy.stats module (V1.3.1)[23] and the statsmodel.api package in Python (v3.7).

cgMLST and wgMLST allele calling and genotyping

Raw sequencing reads were assembled using a custom pipeline in BioNumerics (7.6.3) employing SPAdes[24] (v3.7.0) for its *de novo* assembly. From the raw reads and the *de novo* assembly, alleles were called for core genome and whole genome MLST (cgMLST/wgMLST). In BioNumerics, cgMLST schemes were only available for *E. coli* and *K. pneumoniae* consisting of 2513 and 634 fixed loci, respectively. Pairwise allelic distance was determined by counting the number of discrepant allele variants between two datasets, ignoring loci that were not present in both datasets. Resistance genes and Origins of replication (ORI) were determined using BLAST[25] and two custom databases based on Resfinder[26] and PlasmidFinder[27]. AMR genes were called with a using 90% identity and 60% length cutoff. ORIs were called using 95% identity and 60% length cutoff. In total, 90 WGS datasets were generated. As no gold standard with regard to true genotype of each isolate was available, the following rules were applied: (i) If either two or three out of three datasets of an isolate had a specific genotype, this was considered as a true positive observation; (ii) If only one out of three datasets of an isolate had a specific genotype, this was considered as a false positive observation; (iii) If a different allelic variant was observed (i.e two blaTEM-1B and one blaTEM-116) this was noted as a discrepancy and counted as a false positive.

wgSNP analysis

To determine the best *de novo* assembler to use for wgSNP analysis, we chose the assembler generating the least amount of pairwise SNPs (using SKA), among assemblies of the same isolate. To avoid complexity, only the *E. coli* dataset of this study was used. The following assemblers were used: I) SPAdes (v3.14.0)[24], II) SKESA (parameters: '—use-paired_end', v2.3.0)[28], III) Megahit[29] (v1.2.9). All tools were used in default settings, unless otherwise specified. Additionally, the assembly-free method to determine SNPs straight from the raw reads, using "SKA fastq", was also used in this comparison. The complete workflow is available at "https://github.com/MUMC-MEDMIC/assemblercompare" (v1.0). SKA[13] was used to determine SNPs on a whole genome level, using a split k-mer length of 31. In short, pairwise SNPs were determined by generating a profile of split k-mers, in which the middle base may vary ("SKA fasta" for assembly- or "SKA fastq" for read based SNP profiling). The number of SNPs, between two datasets, was determined by comparing the split k-mer files ("SKA distance"). All data preprocessing for the SNP-based data analysis was performed using Snakemake[30] as workflow manager.

Data Summary

The authors confirm that all supporting data, code and protocols have been provided within the Article. All raw sequencing data was deposited at EBI-ENA under BioProject PRJEB40571.

Results

The assembly coverage, or the depth of coverage, of all isolates ranged from 30 to 203 (figure 1A). The N50 score, indicative for how fragmented a *de novo* assembly is, ranged from 33.712bp (*E. faecium*) to 942.715 bp (*K. pneumoniae*) and showed clear species dependence (Figure 1B). Assemblies of *E. coli, E. faecalis* and *E. faecium* showed a lower N50 score, indicating the difficulties of assembling such genomes (Figure 1B). The number of contigs also varied per species, and overall had a significant negative correlation with the sequencing depth ($P < 0.01$, Spearman rank correlation, Figure 1F).

The number of wgMLST alleles called ranged from 1933 (*Citrobacter sp.)* alleles to 5493 (*K. pneumoniae*, Figure 1C). Furthermore, the average number of alleles per kilobase (kb) ranged from 0.41 to 0.98. A significant positive correlation between the normalized allele count and sequencing depth was observed ($P < 0.05$, Spearman rank correlation Figure 1G). Surprisingly, the *Citrobacter sp.* datasets seemed to showed a low coding density (range 0.41 to 0.65) compared to the median of the entire dataset (0.83). Further inspection of the *Citrobacter sp.* genome assemblies using BLAST webservice (https://blast.ncbi.nlm.nih.gov/Blast.cgi, accessed 1-4-2020), showed low homology (~85% DNA identity score) to known *Citrobacter sp.* isolates available in the NCBI database (accessed on 1-4-2020, data not shown).

One dataset of *E. faecium-1* had an unusually large genome size of 5.4Mb (Figure 1E). This dataset also had a higher number of contigs; (636, median of 274 for *E. faecium,* Figure 1F*)*, and showed a lower number of alleles per kb (0.43, median of 0.84. Figure 1D) compared to the other *E. faecium* datasets. This indicates contamination in the NGS dataset of a non- *E. faecium* microbe. Manual inspection of the assembly, using BLAST webservice, showed the presence of contigs belonging to *Cutibacterium* (formerly known as *Propionibacterium*), a skin commensal and previously described as a common contaminant of NGS datasets[31–33].

Resistance genes and plasmid ORIs

Overall, a good consensus was obtained for the genotyping of plasmid ORIs and AMR genes (Figure 2A and 2B). A total of 973 AMR genes and ORIs were called with a precision of 99.0% and sensitivity of 99.2%. For four isolates, a genotype was not called in one of the datasets. The missed genotypes were for *E. cloacae-2* a

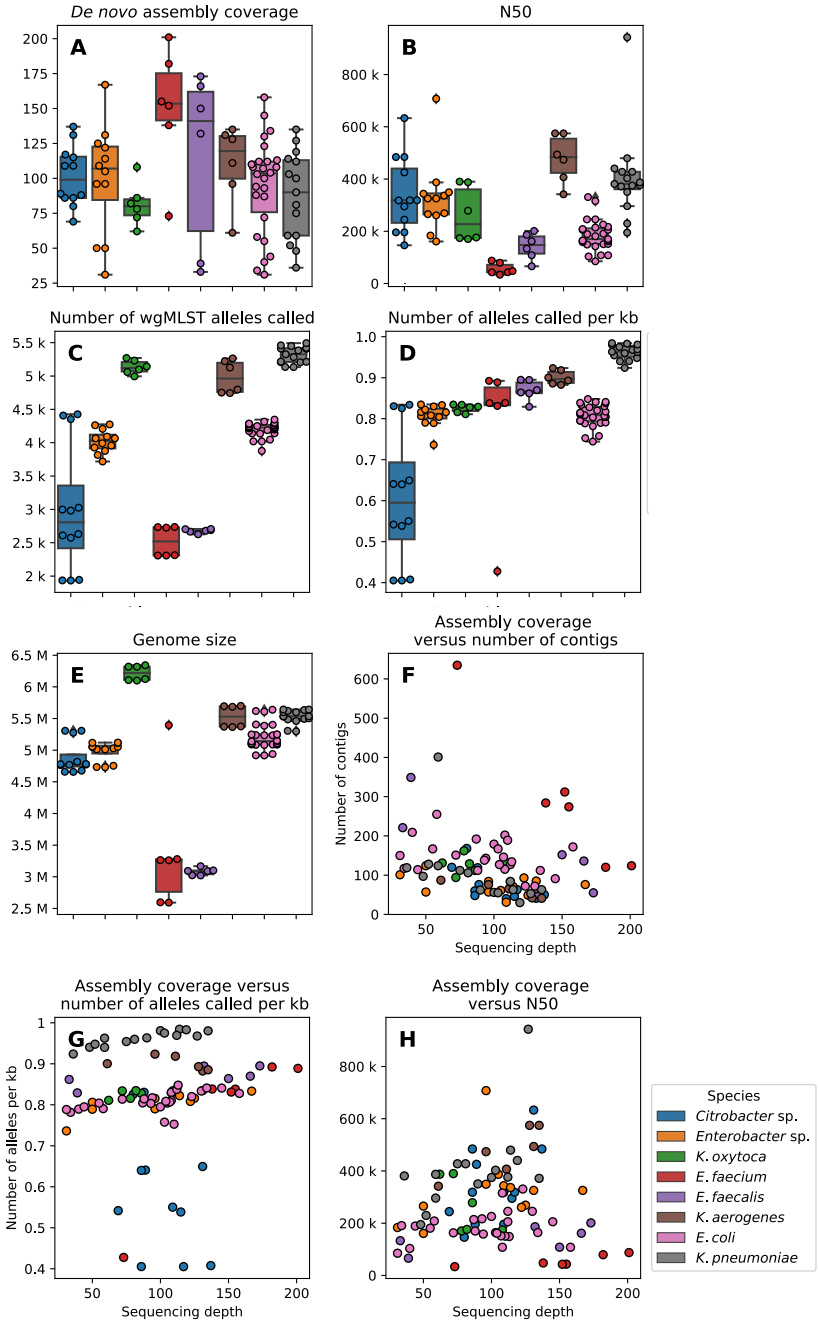


Figure 1. Distribution of various quality parameters pre- and post de novo assembly. Subplot A to E show boxplots with interquartile (IQ) range. Whiskers range up to 1.5 times the IQ range. All single datapoints are represented as single dots. Subplot F to G show scatterplots of relations between two quality metrics.

sul1 gene, for *E. coli-6* a *tet(A)* gene, for *E. faecalis*-1 an *aac(6')-aph(2'')* gene, and for *E. faecium-2* an *aph(2'')-Ia* gene. For *Citrobacter sp.-2* and *K. oxytoca-2,* a false discovery of a *blaTEM-116* was observed, as this genotype was not called in either of the other two datasets of these isolates. For four isolates, a discrepant genotype was called. These discrepancies were observed for *K. aerogenes-2* (*blaTEM*), for *E. cloacae-2* (*aadA*), and for *K. oxytoca-1* (*blaOXY* and *blaTEM*). Twice, an unexpected ColpVC was found in a *K. oxytoca-2* and *K. pneumoniae-4* dataset, which were from two different centers, indicating either loss of this plasmid in the other dataset of this isolate, or contamination during DNA isolation or library preparation (Figure 2A).

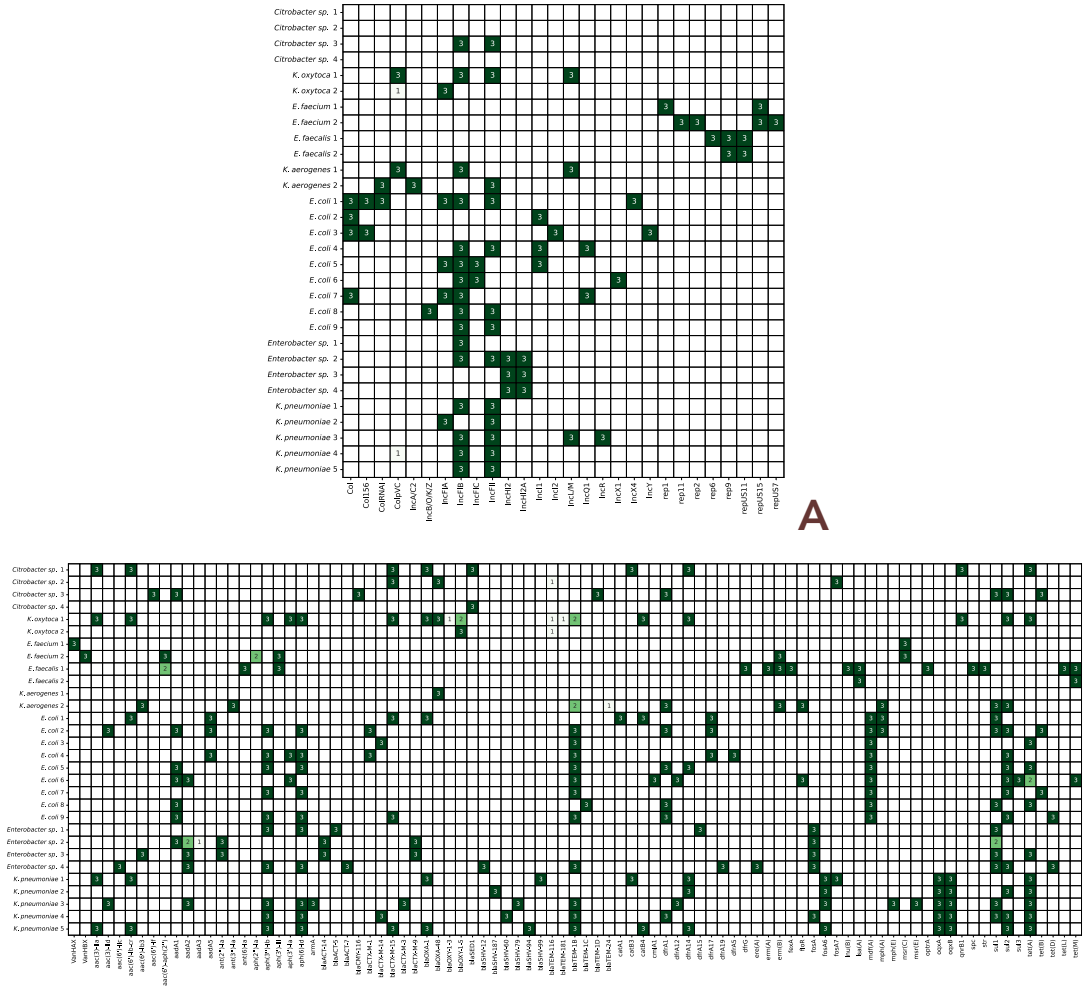


Figure 2. Heatmap of the number of genotype calls for various origins of replication **(Figure 2A)** and AMR genes **(Figure 2B)** among the isolates subjected to WGS. Genotype calls per locus was summed up for each center's isolate if this locus was detected in their dataset.

Inter-laboratory variation in cgMLST profiles

To assess the baseline genetic variation of identical isolates when these isolates were sequenced in different sequencing institutes, we compared the cgMLST and wgMLST profiles among the isolates from the three participating institutes. Only for *E. coli* and *K. pneumoniae* cgMLST schemes were available for use in BioNumerics. On average, 2441 (97.1%) and 615 (97.1%%) core genome alleles were called for *E. coli* and *K. pneumoniae* respectively (Supplemental Figure 1). In total, 27 and 15 pairwise allelic distances were calculated among the nine *E. coli* and five *K. pneumoniae* isolates. In 25/27 (93%) and 12/15 (80%) comparisons, a perfect concordance of cgMLST profiles was observed. If no concordance in cgMLST profiles was observed, only one allele was differently called (Supplemental figure 2).

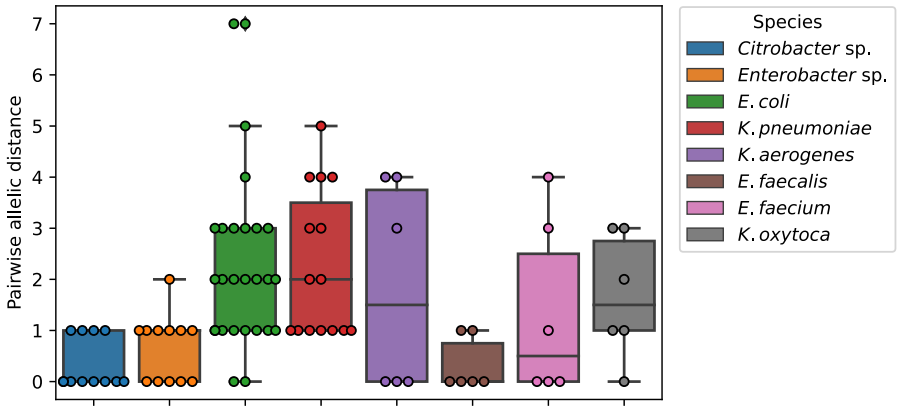


Figure 3. Boxplot of the allelic distance based on wgMLST between the triplicates that were selected for WGS. Boxes show interquartile (IQ) range and whiskers range up to 1.5 times the IQ range. All single pairwise observations were plotted as dots.

Inter-laboratory variation in wgMLST profiles

In total 90 pairwise comparisons were made for *K. oxytoca* (6), *Citrobacter sp.* (12), *E. coli* (27), *K. pneumoniae* (15), *E. cloacae* (12), *K. aerogenes* (6), *E. faecalis* (6) and *E. faecium* (6). Perfect concordance in wgMLST profiles was obtained in 26/90 (29%) comparisons (Figure 3). In 44/90 (49%) pairwise comparisons, one or two discrepant alleles were observed. Only 23/90 (22%) comparisons showed more than two discrepant alleles, with a maximum of seven alleles different for an *E. coli*. For *E. faecium-1* with the contamination of *Cutibacterium* had a perfect concordance of wgMLST profiles was observed (data not shown). This indicates the robustness of allele-based typing despite contamination with bacterial DNA from different species. For all species, an average allelic distance of 1.6 alleles (standard deviation 1.6) was observed.

For the four *Citrobacter sp.*, a highly diverse number of wgMLST alleles were called, ranging from 1933 to 4426. The genome size did not vary strongly (mean 4.88Mb, range 4.66Mb to 5.31Mb). The normalized allele counts were lower for *Citrobacter sp.* (mean 0.61, range 0.41 to 0.83) than in other species in this study (mean 0.84, range 0.43 to 0.98). Therefore, the variation in the number of alleles in the wgMLST scheme for *Citrobacter sp.* cannot be determined in this study, as an incomplete set of alleles were called.

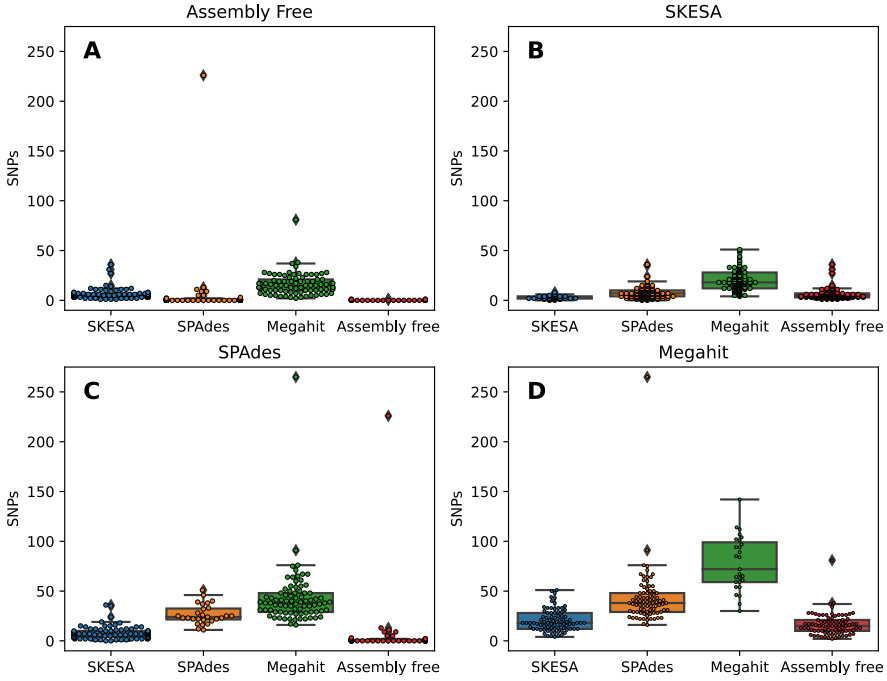


Figure 4. Boxplots of the inter and intra assembly difference in de novo assemblies based on SNPs, using SKA for the E. coli dataset. De novo assembly method compared to is indicated above each box. A. Assembly free, B SKESA, C SPAdes and D Megahit. Boxes show interquartile (IQ) range. Whiskers range up to 1.5 times the IQ range. All single pairwise observations were plotted as dots.

Reference free wgSNP

As mutations in the genome can also arise in intergenic regions (which are not taken into account in MLST based methods), all assemblies of each isolate were screened using pairwise SNPs. First, the most optimal assembler for this task was chosen. For this, we determined the inter- and intra-assembler variation introduced on the number of pairwise SNP between two *de novo* assemblies. The best assembler was chosen based on the one that introduced the least number of pairwise SNPs in the datasets from the same isolates with the intra-assembler comparison. To reduce complexity,

only the *E. coli* dataset was used. Secondly, the number of pairwise SNPs was determined for the entire dataset using the best suited assembler. Additionally, we also used the assembly-free method for determining SNPs, as in implemented by SKA.

The mean intra-assembly variation was 0.2 SNPs (assembly free), 2.7 SNPs (SKESA), 26.6 SNPs (SPAdes) and 77.8 SNPs (Megahit) (Figure 4A, B, C and D). The mean Inter-assembler variation ranged from 3.9 (assembly free compared to SPAdes) up to 43.0 SNPs ("SPAdes to megahit"). All combinations, except the "assembly-free to assembly-free" and "SKESA to SKESA", revealed pairwise comparisons with over 20 SNPs for the *E. coli* dataset. Therefore, only these two methods were used to analyze the complete dataset.

Using the assembly free approach, 63/90 (70%) and 21/90 (21%) comparisons show zero or one pairwise SNPs respectively (Figure 5A). Only for *K. pneumoniae*, *E. faecium, K. oxytoca*, and *K. aerogenes* was more than one pairwise SNP observed, with a maximum of five SNPs for *K. oxytoca*. Using the assembly-based approach zero SNPs were observed among assemblies in 10/90 (10%) comparisons (Figure 5B). Less than five pairwise SNPs were observed in 72/90 (80%) of the comparisons. Interestingly, in the *K. aerogenes* and *K. oxytoca* datasets, more than eight pairwise SNPs were observed. However, on wgMLST no more than four alleles difference was observed. On average, 3.4 (standard deviation 2.6) pairwise SNPs were observed between assemblies of the same isolates (but sequenced in different institutes). Overall, more pairwise SNPs were observed when assemblies were used for SNP analysis compared to screening raw reads for SNPs. The difference in number of k-mers between the assembly free and assembly-based methods ranged from -2.1% to 1.2% (Supplemental figure 3), indicating that a similar amount of k-mers were compared in both methods.

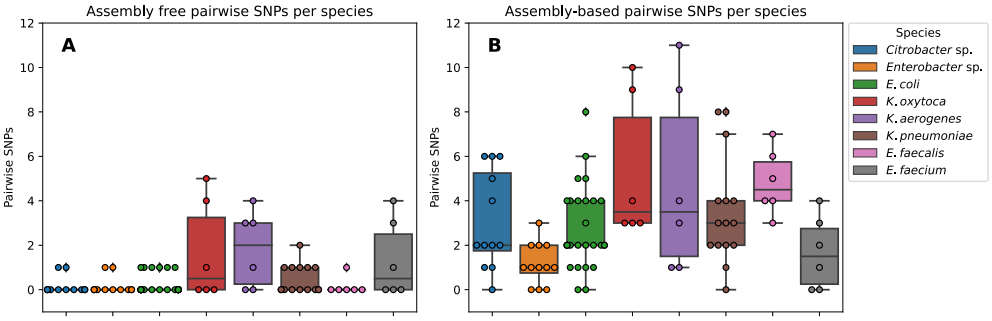


Figure 5. Boxplot of the SNP distance between the triplicates that were selected for WGS. Boxes show interquartile (IQ) range and whiskers range up to 1.5 times the IQ range. All single pairwise observations were plotted as dots. Panel A shows SNP distances using the raw reads as input for SKA. Panel B shows the SNP distances based on the de novo assembly using SKESA.

Discussion

Using an inter-laboratory ring trial we evaluated the reproducibility of whole genome sequencing for outbreak surveillance purposes. Participating institutes subjected the same set of 30 bacterial isolates of various *Enterobacteriaceae* and Enterococci species to whole genome sequencing. As a first step, we assessed various QC measures and observed a slight positive trend of the sequencing depth on the normalized number of alleles called in the sequencing depth range of 30 to 207-fold. It remains unclear what sequencing depth is needed to correctly reconstruct the maximum possible number of correct alleles in the genome. Kluytmans van den Bergh et al.[21] demonstrated an increase in resolution for phylogenetic reconstruction of *Enterobacteriaceae* if wgMLST is implemented compared to cgMLST. This would indicate that making more alleles available for comparison will improve the surveillance of outbreaks by cgMLST or wgMLST methods. Therefore, it is advisable to generate WGS data of sufficient depth to maximize the number of loci in the *de novo* assembly. On the other hand, deeper sequencing after a certain depth may not improve the phylogenetic signal any further, and does increase the run-time of subsequent *de novo* assembly.

Prokaryotes show a coding density of 1 CDS per 1 kb[34], however we observed a lower allele density. The majority of our datasets showed a lower number allele density (0.83 per kb, Figure 1D), which could be caused by the quality filtering step in allele calling. However, the low number of called alleles for most *Citrobacter sp.* may be explained by incomplete allele schemes, which do not contain the complete diversity of alleles. This indicates that the diversity of *Citrobacter sp.* genome assemblies present in public databases is incomplete, and our data may represent the discovery of a new antibiotic-resistant *Citrobacter sp.* in The Netherlands.

Genotyping AMR genes and ORIs

We next performed identification of AMR genes and plasmid ORIs. Overall, an excellent reproducibility was achieved, as a precision of 99% was obtained. Most discrepancies could be explained by the variation in the variant calling of a specific resistance gene. There was an unexplained absence of a resistance gene four times in 973 genotype calls. Although the DNA isolation method used here showed good results for the application of WGS[35], some loci could still be missed due to inefficient isolation of plasmid DNA, where these AMR genes can be located. Only twice, and in different institutes, an unexpected ColpVC ORI was found in one of the sequencing datasets, which may indicate contamination during DNA isolation or library preparation. Strauß and co-authors reported a 1.7% discordance between WGS and micro-array for the detection of resistance and virulence genes[36]. In this study, a similar reproducibility in typing resistance genes and ORIs was obtained and previously described by Kozyreva *et al.,* which found a reproducibility rate of 99.97%[37].

Genetic variation

It is of great importance that the genetic distance between technical duplicates does not surpass commonly used thresholds to classify isolates into clusters. In this study some variation among the wgMLST allelic profiles was observed, translating to an average of 0.49 discrepant allele per 1000 alleles. Kluytmans-Van den Bergh *et al.*[21] reported a variation in genetic distance based on wgMLST in a range of 0 to 0.001 (which translates to 5 alleles difference, based on 5000 alleles compared) for five *E. coli* and three *K. pneumoniae* which were sequenced in duplicate[21]. This is in concordance with our study, where 88/90 comparisons differed by no more than five alleles. Additionally, clonal thresholds reported by these authors were roughly 26 and 2 alleles difference for *E. coli* and *K. pneumoniae* on cgMLST respectively. For wgMLST this was 29, 23, 8, and 14 alleles difference for *E. coli, K. pneumoniae*, *Citrobacter sp.* and *Enterobacter sp.* respectively. These clonal thresholds are higher by a safe margin than the variation between any of the replicates in our study presented here. Although variation on a genetic level was observed, the level of disparity remained below other thresholds commonly employed for hospital outbreak surveillance purposes[9]. Previous work suggested a cut-off of 10 alleles for MDR *E. coli* and *K. pneumoniae* based on cgMLST[38,39]. Therefore, it is safe to assume that if harmonised protocols are used, the technical genetic variation will remain within these previously described thresholds.

In the wgSNP analysis, all methods except for the "assembly-free to assembly-free" and "SKESA to SKESA" showed pairwise comparisons with more than 20 SNPs. This indicates that using SPAdes or Megahit in combination with a SNP based method is unsuitable for outbreak surveillance, as datasets from identical isolates have more SNPs than commonly used outbreak thresholds[9], indicating that these isolates would be considered not clonally related, thus not belonging to the same outbreak. Furthermore, this also held true when comparing two assembly methods, which implies that comparing bacterial assemblies should be avoided at all costs if centers employ different methodologies to generate *de novo* assemblies for WGS outbreak surveillance. Potential outbreaks could be missed due to the large number of SNPs detected, resulting in identical isolates not being flagged as clonally related. This would subsequently have implications for infection prevention and control. For the assembly-free method, we observed most replicates to have no SNPs between each other (70%), which is in line with the GenomeTrakr proficiency-test study, which found a similar fraction of datasets showed having no SNPs (73%)[40].

Variation in SNPs among isolates showed a lower number of SNPs based on the assembly-free method compared to the assembly-based method. It is unlikely that this is caused by different numbers of k-mers that were compared for SNPs, as there was only a modest difference for the number of k-mers compared between the assembly free and the assembly based SNP analysis, which ranged from -2.1% to 1.2% dif-

ference in compared k-mers (Supplemental figure 3). Therefore, it is more likely that *de novo* assembly introduces phylogenetic noise in regions difficult to assemble, like regions such as mobile elements (transposons and plasmids). Previously described work employing SNP-based methodologies to infer phylogeny among bacterial isolates often mask regions in the genome that are sensitive for non-informative SNPs for phylogenetic reconstruction, such as mobile genetic elements (MGE). Masking of these regions requires specialized tools such as Gubbins[41] that are able to recognize regions with elevated numbers of base substitutions in the genome. Unfortunately, using this reference-free methodology makes this masking impossible to perform in an unbiased and automated fashion like in the Gubbins pipeline. Therefore, we must assume the possibility of overestimation of SNPs among isolates in our study.

Study limitations

For this study, only three centers participated in this ring-trial, all part of the I-4-1-Health study group. Here, ESBL-producing and ciprofloxacin-resistant *Enterobacteriaceae* and vancomycin-resistant *Enterococcus* were defined of primary interest, however other important nosocomial bacterial pathogens such as *Pseudomonas* sp., *Staphylococci* and A*cinetobacter sp.* were not included in the study. Furthermore, all three centers used the same protocols for the extraction and library preparation for sequencing on the Illumina MiSeq. Recommendations for future research would therefore be to determine if these harmonised wet-lab protocols and subsequent bioinformatic data processing are indeed required for the reconstruction of outbreak clusters.

Conclusion

Overall, the work presented here demonstrated that whole genome sequencing generates reproducible results when comparing results across laboratories that use identical wet-lab and dry-lab methodologies for WGS. Furthermore, to make multi-center outbreak surveillance feasible in the future, we recommend that laboratories share raw sequencing reads, because systematic errors were introduced in the *de novo* assemblies by different assemblers. Finally, work presented here lays the foundation for routine proficiency testing in clinical microbiology laboratories.

Acknowledgements

We are grateful to David Barnett for proof reading this manuscript and the collaborators in the participating laboratories for their contribution to the collection of whole genome sequence data. These collaborators are part of the i-4-1-Health Study Group:

Lieke van Alphen (Maastricht University Medical Center+, Maastricht, the Netherlands), Nicole van den Braak (Avans University of Applied Sciences, Breda, the Netherlands), Caroline Broucke (Agency for proofreading ;)Care and Health, Brussels, Belgium), Anton Buiting (Elisabeth-TweeSteden Hospital, Tilburg, the Netherlands), Liselotte Coorevits (Ghent University Hospital, Ghent, Belgium), Sara Dequeker (Agency for Care and Health, Brussels, Belgium and Sciensano, Brussels, Belgium), Jeroen Dewulf (Ghent University, Ghent, Belgium), Wouter Dhaeze (Agency for Care and Health, Brussels, Belgium), Bram Diederen (ZorgSaam Hospital, Terneuzen, the Netherlands), Helen Ewalts (Regional Public Health Service Hart voor Brabant, Tilburg, the Netherlands), Herman Goossens (University of Antwerp, Antwerpen, Belgium and Antwerp University Hospital, Antwerp, Belgium), Inge Gyssens (Hasselt University, Hasselt, Belgium), Casper den Heijer (Regional Public Health Service Zuid- Limburg, Heerlen, the Netherlands), Christian Hoebe (Maastricht University Medical Center+, Maastricht, the Netherlands and Regional Public Health Service Zuid-Limburg, Heerlen, the Netherlands), Casper Jamin (Maastricht University Medical Center+, Maastricht, the Netherlands), Patricia Jansingh (Regional Public Health Service Limburg Noord, Venlo, the Netherlands), Jan Kluytmans (Amphia Hospital, Breda, the Netherlands and University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands), Marjolein Kluytmans-van den Bergh (Amphia Hospital, Breda, the Netherlands and University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands), Stefanie van Koeveringe (Antwerp University Hospital, Antwerp, Belgium), Sien De Koster (University of Antwerp, Antwerp, Belgium), Christine Lammens (University of Antwerp, Antwerp, Belgium), Isabel Leroux-Roels (Ghent University Hospital, Ghent, Belgium), Hanna Masson (Agency for Care and Health, Brussel, Belgium), Ellen Nieuwkoop (Elisabeth-TweeSteden Hospital, Tilburg, the Netherlands), Anita van Oosten (Admiraal De Ruyter Hospital, Goes, the Netherlands), Natascha Perales Selva (Antwerp University Hospital, Antwerp, Belgium), Merel Postma (Ghent University, Ghent, Belgium), Stijn Raven (Regional Public Health Service West-Brabant, Breda, the Netherlands), Veroniek Saegeman (University Hospitals Leuven, Leuven, Belgium), Paul Savelkoul (Maastricht University Medical Center+, Maastricht, the Netherlands), Annette Schuermans (University Hospitals Leuven, Leuven, Belgium), Nathalie Sleeckx (Experimental Poultry Centre, Geel, Belgium), Arjan Stegeman (Utrecht University, Utrecht, the Netherlands), Tijs Tobias (Utrecht University, Utrecht, the Netherlands), Paulien Tolsma (Regional

Public Health Service Brabant Zuid-Oost, Eindhoven, the Netherlands), Jacobien Veenemans (Admiraal De Ruyter Hospital, Goes, the Netherlands), Dewi van der Vegt (PAMM Laboratory for Pathology and Medical Microbiology, Veldhoven, the Netherlands), Martine Verelst (University Hospitals Leuven, Leuven, Belgium), Carlo Verhulst (Amphia Hospital, Breda, the Netherlands), Pascal De Waegemaeker (Ghent University Hospital, Ghent, Belgium), Veronica Weterings (Amphia Hospital, Breda, the Netherlands), Clementine Wijkmans (Regional Public Health Service Hart voor Brabant, Tilburg, the Netherlands), Patricia Willemse-Smits (Elkerliek Hospital, Helmond, the Netherlands), Ina Willemsen (Amphia Hospital, Breda, the Netherlands).

Funding

The i-4-1-Health project was financed by the Interreg V Flanders-The Netherlands program, the cross-border cooperation program with financial support from the European Regional Development Fund (ERDF). Additional financial support was received from the Dutch Ministry of Health, Welfare and Sport, the Dutch Ministry of Economic Affairs, the Province of Noord-Brabant, the Belgian Department of Agriculture and Fisheries, the Province of Antwerp and the Province of East-Flanders." The authors are free to publish the results from the project without interference from the funding bodies. Selective and non-selective agar plates, ETEST® stips and VITEK® 2 AST cards were provided by bioMérieux (Marcy l'Etoile); FecalSwabs® and tryptic soy broths were provided by Copan Italy (Brescia, Italy). The authors are free to publish the results from the project without interference from bioMérieux or Copan Italy.

Conflicts of interest

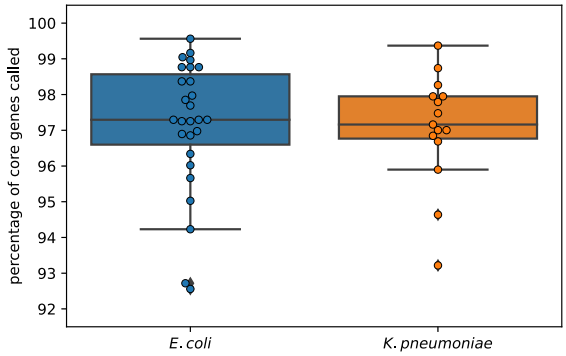
The author(s) declare that there are no conflicts of interest.

References

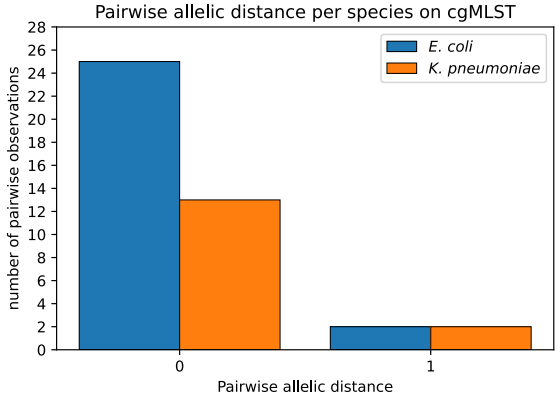
1. Global action plan on antimicrobial resistance. https://www.who.int/antimicrobial-resistance/global-action-plan/development_process/en/.
2. Thaden, J. T. *et al.* Increased costs associated with bloodstream infections caused by multidrug-resistant gram-negative bacteria are due primarily to patients with hospital-acquired infections. *Antimicrob. Agents Chemother.* (2017) doi:10.1128/AAC.01709-16.
3. Founou, R. C., Founou, L. L. & Essack, S. Y. Clinical and economic impact of antibiotic resistance in developing countries: A systematic review and meta-analysis. *PLoS ONE* (2017) doi:10.1371/journal.pone.0189621.
4. Mariappan, S., Sekar, U. & Kamalanathan, A. Carbapenemase-producing Enterobacteriaceae: Risk factors for infection and impact of resistance on outcomes. *Int. J. Appl. Basic Med. Res.* (2017) doi:10.4103/2229-516x.198520.
5. Harris, S. R. *et al.* Evolution of MRSA during hospital transmission and intercontinental spread. *Science (80-.).* (2010) doi:10.1126/science.1182395.
6. Pendleton, S., Hanning, I., Biswas, D. & Ricke, S. C. Evaluation of whole-genome sequencing as a genotyping tool for Campylobacter jejuni in comparison with pulsed-field gel electrophoresis and flaA typing. *Poult. Sci.* (2013) doi:10.3382/ps.2012-02695.
7. Leekitcharoenphon, P., Nielsen, E. M., Kaas, R. S., Lund, O. & Aarestrup, F. M. Evaluation of whole genome sequencing for outbreak detection of salmonella enterica. *PLoS One* (2014) doi:10.1371/journal.pone.0087991.
8. Salipante, S. J. *et al.* Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology. *J. Clin. Microbiol.* (2015) doi:10.1128/JCM.03385-14.
9. Schürch, A. C., Arredondo-Alonso, S., Willems, R. J. L. & Goering, R. V. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene–based approaches. *Clinical Microbiology and Infection* (2018) doi:10.1016/j.cmi.2017.12.016.
10. Silva, M. *et al.* chewBBACA: A complete suite for gene-by-gene schema creation and strain identification. *Microb. genomics* (2018) doi:10.1099/mgen.0.000166.
11. Seemann, T. Snippy. https://github.com/tseemann/snippy.
12. Wu, T. D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, (2010).
13. Harris, S. R. SKA: Split Kmer Analysis Toolkit for Bacterial Genomic Epidemiology. *bioRxiv* (2018) doi:10.1101/453142.
14. Hyatt, D. *et al.* Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* (2010) doi:10.1186/1471-2105-11-119.
15. Zhou, Z., Alikhan, N. F., Mohamed, K., Fan, Y. & Achtman, M. The EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and Escherichia core genomic diversity. *Genome Res.* (2020) doi:10.1101/gr.251678.119.
16. Darling, A. E., Mau, B. & Perna, N. T. Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* (2010) doi:10.1371/journal.pone.0011147.
17. Treangen, T. J., Ondov, B. D., Koren, S. & Phillippy, A. M. The harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* (2014) doi:10.1186/s13059-014-0524-x.
18. Gardner, S. N., Slezak, T. & Hall, B. G. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* (2015) doi:10.1093/bioinformatics/btv271.
19. Harris, S. R. *et al.* Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: A descriptive study. *Lancet Infect. Dis.* (2013) doi:10.1016/S1473-3099(12)70268-2.
20. Rumore, J. *et al.* Evaluation of whole-genome sequencing for outbreak detection of Verotoxigenic Escherichia coli O157:H7 from the Canadian perspective. *BMC Genomics* (2018) doi:10.1186/s12864-018-5243-3.
21. Kluytmans-Van Den Bergh, M. F. Q. *et al.* Whole-genome multilocus sequence typing of extended-spectrum-beta-lactamase-producing enterobacteriaceae. *J. Clin. Microbiol.* (2016) doi:10.1128/JCM.01648-16.
22. Bergh, M. K. Den *et al.* Microbiological methods to detect intestinal carriage of highly-resistant microorganisms (HRMO) in humans and livestock in the i-4-1-Health Dutch- Belgian cross-border project. *Preprints.org* 1–16 (2019) doi:10.20944/preprints201912.0216.v1.
23. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* (2020) doi:10.1038/s41592-019-0686-2.

24. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* (2012) doi:10.1089/cmb.2012.0021.
25. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
26. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* (2012) doi:10.1093/jac/dks261.
27. Carattoli, A. *et al.* In Silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* (2014) doi:10.1128/AAC.02412-14.
28. Souvorov, A., Agarwala, R. & Lipman, D. J. SKESA: Strategic k-mer extension for scrupulous assemblies. *Genome Biol.* (2018) doi:10.1186/s13059-018-1540-z.
29. Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* (2015) doi:10.1093/bioinformatics/btv033.
30. Köster, J. & Rahmann, S. Snakemake-a scalable bioinformatics workflow engine. *Bioinformatics* **34**, 3600–3600 (2012).
31. Tae, H., Karunasena, E., Bavarva, J. H., McIver, L. J. & Garner, H. R. Large scale comparison of non-human sequences in human sequencing data. *Genomics* **104**, 453–458 (2014).
32. Sangiovanni, M., Granata, I., Thind, A. S. & Guarracino, M. R. From trash to treasure: Detecting unexpected contamination in unmapped NGS data. *BMC Bioinformatics* (2019) doi:10.1186/s12859-019-2684-x.
33. Glassing, A., Dowd, S. E., Galandiuk, S., Davis, B. & Chiodini, R. J. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut Pathog.* (2016) doi:10.1186/s13099-016-0103-7.
34. Koonin, E. V. & Wolf, Y. I. Genomics of bacteria and archaea: The emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* **36**, 6688–6719 (2008).
35. Nouws, S. *et al.* Impact of DNA extraction on whole genome sequencing analysis for characterization and relatedness of Shiga toxin-producing Escherichia coli isolates. *Sci. Rep.* (2020) doi:10.1038/s41598-020-71207-3.
36. Strauß, L. *et al.* Detecting staphylococcus aureus virulence and resistance genes: A comparison of whole-genome sequencing and DNA microarray technology. *J. Clin. Microbiol.* (2016) doi:10.1128/JCM.03022-15.
37. Kozyreva, V. K. *et al.* Validation and implementation of clinical laboratory improvements act-compliant whole-genome sequencing in the public health microbiology laboratory. *J. Clin. Microbiol.* (2017) doi:10.1128/JCM.00361-17.
38. Roer, L. *et al.* WGS-based surveillance of third-generation cephalosporin-resistant Escherichia coli from bloodstream infections in Denmark. *J. Antimicrob. Chemother.* (2017) doi:10.1093/jac/dkx092.
39. Zhou, H., Liu, W., Qin, T., Liu, C. & Ren, H. Defining and evaluating a core genome multilocus sequence typing scheme for whole-genome sequence-based typing of Klebsiella pneumoniae. *Front. Microbiol.* (2017) doi:10.3389/fmicb.2017.00371.
40. Timme, R. E. *et al.* GenomeTrakr proficiency testing for foodborne pathogen surveillance: An exercise from 2015. *Microb. Genomics* (2018) doi:10.1099/mgen.0.000185.
41. Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* (2015) doi:10.1093/nar/gku1196.

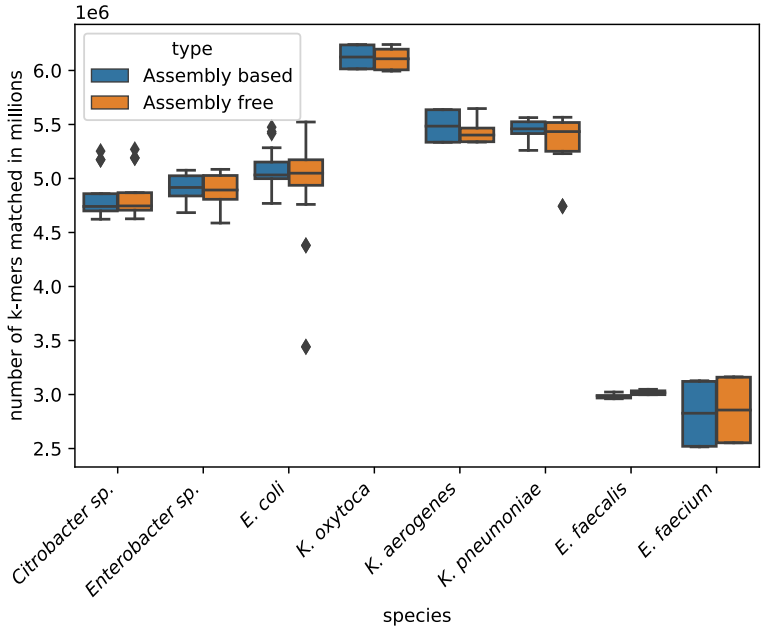
Supplemental figures



Supplemental figure 1. Boxplot of the number of core genes called for cgMLST in percentage for *K. pneumoniae* and *E. coli* respectively. Boxes range the interquartile (IQ) range. Whiskers range up to 1.5 times the IQ range. All single datapoints are represented as single dots. Only cgMLST schemes were available for *E. coli* and *K. pneumoniae*.



Supplemental figure 2. Barplot of the pairwise number of alleles that were different between two strains.



Supplemental figure 3. boxplot of the number of k-mers compared for the assembly free and the assembly based method. Boxes range the interquartile (IQ) range. Whiskers range up to 1.5 times the IQ range. Outliers were shown as single datapoints.

3

CENTRE SPECIFIC BACTERIAL PATHOGEN TYPING AFFECTS INFECTION CONTROL DECISION MAKING

Coolen J.P.M.[1#], **Jamin C.**[2#], Savelkoul P.H.M.[2,3], Rossen J.W.A.[4,5], Wertheim H.F.L.[1], Matamoros S.P.[3^], van Alphen L.B.[2^] on behalf of SIG Bioinformatics in Medical Microbiology NL Consortium*

1. Department of Medical Microbiology and Radboudumc Center for Infectious Diseases, Radboud University Medical Center, Nijmegen, The Netherlands.
2. Department of Medical Microbiology, Care and Public Health Research Institute (CAPHRI), Maastricht University Medical Center+, Maastricht, The Netherlands.
3. Department of Medical Microbiology & Infection Control, Amsterdam University Medical Centers, location VUmc, Amsterdam, The Netherlands
4. Department of Medical Microbiology and Infection Prevention, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands.
5. Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT, United States of America.

[#] These authors share first authorship
[^] These authors share senior authorship

Special Interest Group Bioinformatics in Medical Microbiology (SIG-BIMM) NL Consortium*
AmsterdamUMC, Amsterdam, The Netherlands
Amsterdam Institute for Global Health and Development, Amsterdam, The Netherlands
ETZ, Tilburg, The Netherlands
LUMC, Leiden, The Netherlands
Maasstad Ziekenhuis, Rotterdam, The Netherlands
Microvida, Breda, The Netherlands
MUMC+, Maastricht, The Netherlands
PAMM, Veldhoven, The Netherlands
RadboudUMC, Nijmegen, The Netherlands
RIVM, Bilthoven, The Netherlands
Streeklab Haarlem, Haarlem, The Netherlands
UMCG, Groningen, The Netherlands
UMCU, Utrecht, The Netherlands

ABSTRACT

Whole-genome sequencing is becoming the *de facto* standard for bacterial outbreak surveillance and infection prevention. This is accompanied by a variety of bioinformatic tools and needs bioinformatics expertise for implementation. However, little is known about the concordance of reported outbreaks when using different bioinformatic workflows. In this multi-centre proficiency testing among thirteen major Dutch health care affiliated centres, bacterial whole-genome outbreak analysis was assessed. Centres who participated obtained two randomized bacterial datasets of Illumina sequences, a *Klebsiella pneumoniae* and a Vancomycin-resistant *Enterococcus faecium,* and were asked to apply their bioinformatic workflows. Centres reported back on antimicrobial resistance, Multi-locus sequence typing (MLST), and outbreak clusters. The reported clusters were analysed using a method to compare landscapes of phylogenetic trees and calculating Kendall-Colijn distances. Furthermore, fasta files were analysed by state-of-the-art single nucleotide polymorphism (SNP) analysis to mitigate the differences introduced by each centre and determine standardized SNP cut-offs. Thirteen centres participated in this study. The reported outbreak clusters revealed discrepancies between centres, even when almost identical bioinformatic workflows were used. Due to stringent filtering, some centres failed to detect Extended-spectrum beta-lactamase genes and MLST loci. Applying a standardized method to determine outbreak clusters on the reported *de novo* assemblies, did not result in uniformity of outbreak-cluster composition among centres.

Impact statement

Bacterial typing and outbreak analyses are essential for performing appropriate infection prevention control. Whole genome sequencing (WGS) is quickly becoming the gold standard in the field, notwithstanding the bioinformatic tools used to process the data and interpret the phylogenetic relation between the bacterial pathogens are currently not standardized. To date, it remains unclear what impact the use of these different tools has on the typing outcome and interpretation of outbreaks between different centres. In this study, we performed a proficiency test that focuses on the impact of different bioinformatic tools applied by centres on interpretation and possible infection prevention decision making. The results of this study contribute to the community by: i) exposing the extend of variations in WGS analysis resulting from usage of different bioinformatics tools, parameters and interpretation thresholds; ii) highlighting the shortcomings of certain bioinformatic tools and decisions; iii) provide insights on how to improve bacterial typing. We bring to light that it is essential to apply identical bioinformatic workflows to make it possible to implement inter-laboratory surveillance on regional or national level and thus improve future outbreak analysis.

Keywords

Whole genome sequencing; Outbreak analysis; bacterial typing; Bioinformatics; Proficiency test; Infection Prevention Control

Data summary

K. pneumoniae and *E. faecium* Illumina sequence data is available via BioProject PRJEB15226 and PRJEB25424, respectively. For a full list of the accession numbers, see Supplementary Table S1. Proficiency test template sheets and associated code are available at "https://github.com/MUMC-MEDMIC/SIGBIO-proficiencytest" under an MIT license.

Introduction

Dissemination of pathogenic bacteria is a significant contributor to healthcare-associated infections (HAI) and a global problem. For intensive care (IC) admitted patients, 11787 (8.3%) patients acquired an HAI in Europe in 2017 alone[1]. Infections by Antimicrobial resistant bacteria are an increased risk for mortality[2].

Of significant interest are the ESKAPE pathogens (Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa, and Enterobacter sp.), as they are associated with a burden on the economy and adverse outcomes for hospitalized patients[2,3]. Therefore, it is essential to curb the dissemination and infections of these nosocomial pathogens by employing proper infection prevention measurements and typing strategies to strengthen surveillance in and around healthcare facilities.

Conventional typing methods such as Pulsed-field Gel Electrophoresis (PFGE)[4], Multi-locus sequence typing (MLST)[5], and Amplification fragment length polymorphism (AFLP)[6] have been used for many years to perform outbreak analysis and made bacterial epidemiology possible. These methods are robust and have well-defined guidelines[7]. Nowadays, whole-genome sequencing (WGS) has become common, as an increased number of laboratories have adopted it. The versatility, backward compatibility and ability to measure at a detailed genomic level are significant contributors to its increased implementation[8–13], thereby phasing out conventional typing methods.

WGS provides genomic data which can be used to find genetic sample-to-sample relations[14–16]. WGS outbreak analysis is more and more applied by hospital Infection Prevention Control (IPC) teams to trace and monitor pathogenic infections[17–20] but also to traceback the source of transmission[21,22]. Additionally, with WGS, one can detect antimicrobial resistance (AMR) genes and virulence factors, which is a beneficial add-on for clinicians and IPC[11]. To perform bacterial whole-genome based

outbreak analysis, WGS for data needs pre-processing using either one of three strategies or a combination. i) Reference-based: By mapping sequence reads to a reference genome and detect single-nucleotide polymorphisms (SNP). ii) Allelic-based: by determining the allelic content and compare these alleles between strains, commonly referred to as core genome (cg)- or whole genome (wg) MLST. iii) K-mer based: Genomic data is grouped into smaller sequences of length k, and the composition of those shorter sequences is used to detect SNPs. To accompany these strategies, a vast amount of bioinformatic tools are available[10]. To date, guidelines or quality markers for WGS outbreak analysis in nosocomial settings are still in its infancy. However, minimal sequencing quality requirements and well-defined quality markers are needed to harmonize laboratories and make inter-laboratory comparisons possible[23].

Previous studies have provided insights into the inter-laboratory comparison of WGS data. A study that assessed the reproducibility of WGS-based typing by performing a ring trial with multiple centres concluded that WGS-based typing is reproducible for Staphylococcus aureus[24]. Studies show that the identification of AMR genes is reproducible[25]. However, the translation to phenotype is inconsistent[26].

A third initiative is ongoing and initiated by The Swiss Institute of Bioinformatics. They are performing a nationwide quality assessment ring trial focusing on bacterial phylogeny to eventually start a nationwide WGS outbreak surveillance platform[23].

The variety in bioinformatic workflows for outbreak analyses applied by these studies only reflects a small portion of the total diversity of procedures used among centres. However, little is known about the congruence of identifying bacterial outbreaks among these various bioinformatic workflows.

This study assessed the comparability of bacterial outbreak analyses and outcomes performed by multiple centres in the Netherlands. We aim to i) expose the differences in bioinformatic workflows applied by centres and their effect on cluster composition, ii) present a strategy to assess performance between centres by using an advance analysis methodology that is easy to implement and interpret, and iii) provide guidelines for bioinformatic workflows to perform outbreak analyses.

Methods

Sequence datasets

Illumina paired-end sequencing data was obtained from the Sequence Read Archive (SRA) and extracted using fasterq-dump (-F -S) (https://github.com/ncbi/sra-tools/tree/master/tools/fasterq-dump). For both *K. pneumoniae* and *E. faecium,* 40 random datasets were selected from BioProject PRJEB15226[27] and PRJEB25424, respectively.

To the best of our knowledge, no publicly available outbreak analysis was conducted previously on these samples. File names and FASTQ headers were anonymised before distribution to the centres. For a full list of the accession numbers, sample details, and metadata, see Supplementary Table S1.

Standardisation of reporting

A secure data transferring service (www.surffilesender.nl) was used to provide each participating centre three standardized excel report files including an instruction manual. The first excel file is a pipeline report file in which the participants describe their pipeline(s), QC rejection parameters, and cluster cut-offs applied on the datasets. The second and third files are sheets for KP and VRE, respectively, in which the participants report genome coverage, Multi-locus sequence type (MLST), and presence of AMR genes for each sample in the dataset, as well as the sample-to-sample relation based on clonal relatedness. Participants used their routine methods and thresholds for analysis. Participating centres were asked to fill in their analysis results in the report sheets and fill out their contact details. All excel sheets were parsed using python (version 3.7.6) and jupyter (version 4.6.1) using pandas (version 0.25.3) and NumPy (version 1.17.3). When necessary, manual inspection of assemblies was done using ABRicate (version 0.9.8)[28] or mlst (version 2.19.0)[29], and inspection of reads was done using KMA (version 1.2.26)[30] and the Resfinder database (accessed 18 June 2020). These template sheets and associated code are available at https://github.com/MUMC-MEDMIC/SIGBIO-proficiencytest.

Reporting of outbreak clusters

Participants registered the outbreak clusters by inserting values in the lower triangle of a similarity matrix by placing either 0, 0.5, or 1, which indicates for "not related", "probably related", or "related", respectively. The lower triangle similarity matrix was converted to a square similarity matrix using python. The instruction manual explicitly stated that all strains in a cluster should be related to each other to be part of a cluster. A custom python script (available at https://github.com/MUMC-MEDMIC/SIGBIO-proficiencytest), implementing networkx (version 2.4), was used to correct these missing relations. With this script, sample-to-sample relations were represented in a network graph, and missing edges were restored between samples to complete the outbreak clusters. For example, if sample A is clonally related to sample B, and sample B clonally related to sample C, sample A and C are also part of the cluster. This missing edge was added in the graph between samples A and C to complete all edges within a cluster. The resulting network graph was converted into a dissimilarity matrix for subsequent analyses. This process was manually inspected before applying to all reported results. Sample-to-sample relations, as reported by

all centres, are aggregated and visualized using Cytoscape (version 3.7.2) using Prefuse Force Directed OpenCL Layout[31].

Creating additional matrixes

A summed distance matrix (SDM) was calculated by summing all the reported dissimilarity matrices per species. A majority distance matrix (MDM) was constructed by selecting values in the SDM that scored higher than half of the number of participating centres (>6.5). Thereby, maintaining only the sample-to-sample relations which represent the majority vote.

Compare outbreak clusters among centres

Dissimilarity matrices per centre and MDM were imported in R (version 3.6.3). Dissimilarity trees were inferred by using UPGMA with hclust (version 3.6.3). A geometric median of all dissimilarity trees, according to the Kendall-Colijn metric, was calculated by using the function medTree of the R package treespace (version 1.1.3.2)[32,33]. Additionally, all trees, including the MDM tree, were compared using the multiDist function of R package treespace per species. This resulted in a pairwise distance matrix of all trees calculated using the Kendall-Colijn metric[32]. The pairwise distance matrix was used as input for hclust to create a UPGMA tree-of-centres. Visualization of the trees and metadata was done using iTOL (version 5.5.1).

Perform SNP-cutoff sweep

Pairwise core- and whole-genome SNPs (cgSNPs, wgSNPs) was used to determine if standardized cut-offs mitigate cluster composition variation. The fasta files of all *de novo* assemblies provided by each centre were used as input. The pairwise SNPs were calculated by split k-mer analyses as implemented by SKA (version 1.0)[34]. In short, split k-mer files (.skf) were generated for each assembly (ska fasta, default parameters). For cgSNPs we only maintained split k-mers that were present in 90% of all assemblies per dataset. Pairwise alignments were made ("ska align -p 0.9"), and the SNP distance was determined using snp-dists (version 0.7.0)[35]. For wgSNP analysis, pairwise SNP distance was determined directly from the .skf files ("ska distance"). The pairwise cg- and wgSNPs were imported into R (version 3.6.3), and a sweep cut-off was applied by setting the cut-off to a range from 0 to 150 SNPs. Samples equal to or within this cut-off were set to be part of an outbreak cluster. Additionally, all strains in an outbreak cluster were related to each other to be consistent with the proficiency test method using R package igraph (version 1.2.5)[36]. Centres were compared to each other by calculating the Kendall-Colijn distance metric using the multiDist function of R package treespace.

Results

Thirteen centres who are members of the Special Interest Group Bioinformatics in Medical Microbiology (SIG-BIMM) NL Consortium participated in this study.

Sequence types

Participating centres were asked to report on conventional MLST. All thirteen centres reported on sequence types (ST). Good concordance among centres on the reported STs was observed for both the KP and the VRE dataset, and for 35/40 (87.5%) and 38/40 (95%) samples, no discrepant ST were reported for KP and VRE, respectively.

For the KP dataset, Centre 3 may have switched sequence data of KP12 with KP13 and KP23 with KP24 (Supplementary Table S2). Centre 5 reported on the least number of STs for *K. pneumoniae* 32/40 (80%) and was the only centre using BioNumerics (Applied Maths, Belgium). This centre mentioned that for some of the isolates no ST could be identified because not all seven required alleles were called.

Centre 2, 5, 7, 11, and 13 mentioned sample KP23 not belonging to the *K. pneumoniae* species but to *Klebsiella variicola*, a different species in the *Klebsiella pneumoniae* complex[37]. Interestingly, of the seven centres (2, 3, 4, 6, 8, 9, 10, and 12) using Ridom SeqSphere+, only Centre 2 identified sample KP23 as *K. variicola*. For KP33, two Centres (2, 11) reported it as ST33, and five centres (1, 2, 9, 10, and 13) appointed it to a novel sequence type.

For the VRE set, only Centre 9 reported on a discordant ST for VRE18 and VRE33 (Supplementary Table S3). Manual inspection of the assembled contigs for these two datasets from Centre 9 revealed for VRE18, an incomplete *pstS* gene (548bp/583bp) at the end of a contig. For VRE33, no *pstS* was identified. The absence of this allele leads to an entirely new ST[38].

AMR reporting

We focus on beta-lactamase and vancomycin resistance genes as they are most clinically relevant. Eleven out of thirteen centres reported the presence of AMR genes. For AMR reporting in the KP dataset, the $bla_{\text{CTX-M}}$ genes were in concordance among all centres for 30 out of the 40 isolates (Supplementary Table S4). For KP07 and KP09 Centre 9, and for KP23 Centre 11 did not report a $bla_{\text{CTX-M}}$ gene. For KP34, only seven out of eleven centres managed to detect a $bla_{\text{CTX-M-14}}$. The presence of $bla_{\text{CTX-M-14}}$ in KP34 using KMA[30] was confirmed. In addition, manual inspection using ABRicate confirmed this gene was absent in the *de novo* assembly of the centres, which did not report $bla_{\text{CTX-M-14}}$. These centres filtered out contigs smaller than 1 kb from the *de novo* assembly (data not shown). For bla_{OXA} genes, all centres were in complete agreement except for strain KP09, for which Centre 9 did not report a $bla_{\text{OXA-1}}$ gene.

For *bla*$_{TEM}$ genes, mainly *bla*$_{TEM-1}$ was reported, and Centre 1 reported on *bla*$_{TEM-30}$ instead of *bla*$_{TEM-1}$. Centres reported a high heterogeneity on *bla*$_{SHV}$ gene variants, as only in six out of forty samples, a single variant was reported. Centre 5 was the only centre indicating the presence of bla$_{SHV-38}$, a beta-lactamase with carbapenemase activity for strain KP12 and KP30. Two centres (2 and 9) reported multiple *bla*$_{SHV}$ genes per strain for most of the *K. pneumoniae* strains in this study. This could be reproduced using the web service of ResFinder, for which multiple *bla*$_{SHV}$ genes were reported on the same genomic location.

For the VRE dataset, eleven out of thirteen centres reported on AMR genes. Here, although variation on reporting the v*anA* or *vanB* cassette, all centres agreed on the presence of vancomycin resistance gene variant A or B per strain (Supplementary Table S5). Seven out of eleven centres reported directly on the *vanHAX* or *vanHBX* cassette, and the remaining four centres reported on all separate *van* genes present on this cassette, including *vanS*, *vanR*, *vanY,* and *vanZ* genes.

Pipeline descriptions

Ten out of thirteen centres used an allele calling method for detecting outbreak clusters, of which eight centres used Ridom SeqSphere+ varying from version 4.1.9 to version 6.0.2, Centre 5 used the BioNumerics (version 7.6.3) software suite for outbreak analysis, and Centre 1 used Pathogenwatch (https://pathogen.watch). For allele calling, six and four centres used cgMLST, wgMLST respectively. The remaining three centres used an SNP approach (Centre 7, 11, and 13). The tools used for SNP-based outbreak analysis are either SKA or kSNP3 (Figure 1).

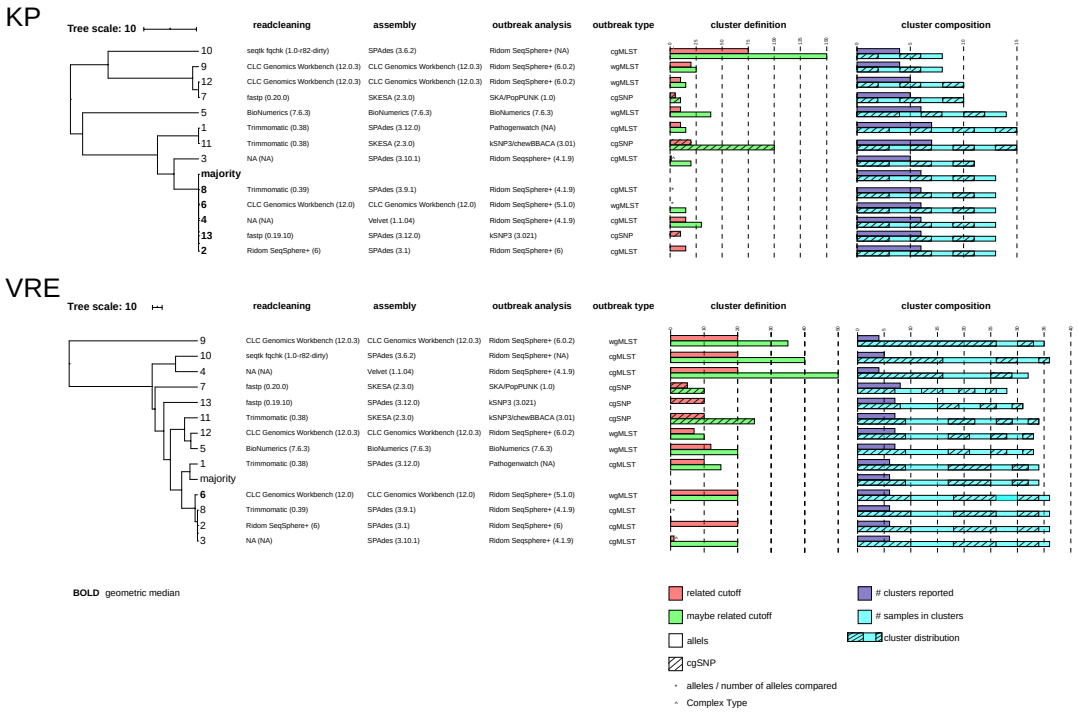


Figure 1. UPGMA tree-of-centres for both the KP as the VRE dataset. The trees indicate the relation of reported outbreak outcome of all 13 centres. Majority and geometric median calculations are added to the UPGMA trees. The data next to the UPGMA trees show the bioinformatic workflow used per centre divided in readcleaning, assembly, and outbreak analysis tools. Furthermore, cluster definitions applied per centre are plotted in barplots and the outcome of the centres is indicated in the barplots with cluster composition. Legends are integrated in the figure.

Reported sample-to-sample relations and outbreak clusters

All reported sample-to-sample relations were aggregated to assess if centres reported the same outbreak clusters. In Figure 2, the sum of all sample-to-sample relations are illustrated. For the KP dataset, we identified six independent clusters as defined by the majority of the centres. Contrarily, Centres 1 and 11 reported a link between KP19 and KP27. Furthermore, only Centre 5 reported sample KP24 as being part of cluster C1. Of all six majority clusters, only Cluster 6 (C6) was reported by all centres.

For the VRE dataset, six independent clusters were identified when using the majority vote. Firstly, Centre 9 reported sample-to-sample relations between clusters that other centres identified as separate clusters: C1, C2, and C3. Centre 4 and 10 reported a maybe relation between two clusters: C1 and C3. Multiple centres linked

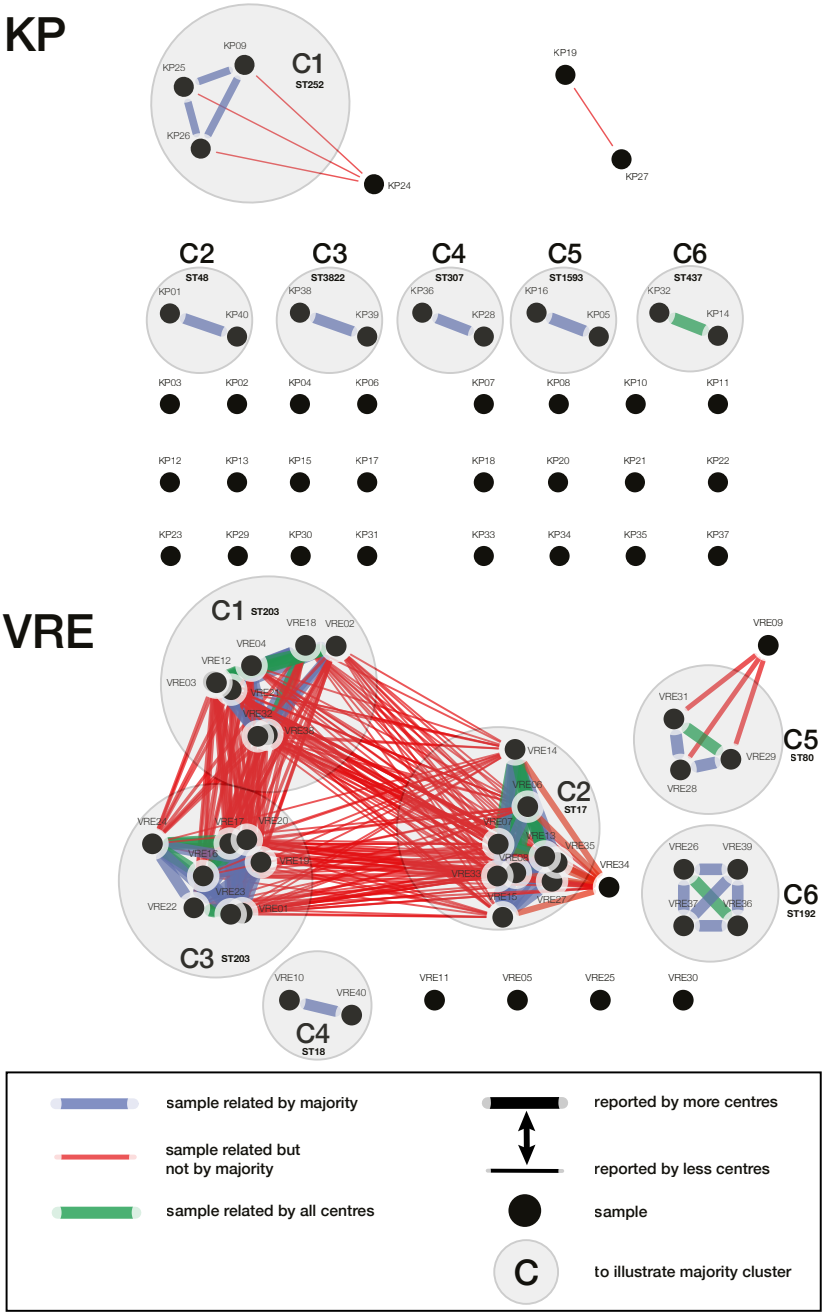


Figure 2. Sample-to-sample relations as reported by the thirteen participating centres. The figure is divided in the *K. pneumoniae* (KP) outcome and in the VRE outcome. All samples are named according the naming that was provided throughout the study. Legend of figure can be found in the right corner of the figure. The ST that was reported by the majority of the centres was added at each cluster.

sample VRE34 to C2, reported as related (Centre 2, 3, 4, and 8), may be related (Centre 6, 9, and 10) and not related (Centre 1, 5,7, 11, 12, and 13). All but Centre 4 reported C4. In C5, the majority did not report sample VRE09 as being part of this cluster. Nevertheless, it was reported by five centres. Lastly, C6 is well supported by twelve out of the thirteen centres for containing sample VRE26, VRE36, VRE37, and VRE39. Notwithstanding, Centre 7 only reported sample VRE26 with VRE36 as being linked.

KP UPGMA tree-of-centres outcome

A UPGMA tree-of-centres (Figure 1) was used to visualize the comparison of each centre's reported outcome, including pipeline description, cluster definition, and cluster composition. Furthermore, trees identified as being the geometric median are noted in bold (Figure 1). For KP, three groups of centres reported identical outbreak cluster content, Centre 7 and 12, Centre 1 and 11, and Centre 2, 4, 6, 8, and 13. The latter group of centres reported identical clusters as the majority vote. Centres 3, 5, 9, and 10 reported unique cluster compositions from any other centre. Centre 5 reported the most dissimilar cluster compositions to any other centre in this study.

VRE UPGMA tree-of-centres outcome

The VRE tree-of-centres shows more dissimilarity (median 64, range 0-283) compared to the KP tree-of-centres (median 39, range 0 to 68). Centre 2, 3, and 8 reported identical outbreak clusters and content. All other centres reported unique outbreak cluster compositions. The VRE tree-of-centres clearly shows a large branch for Centre 9, suggesting that Centre 9 reported a very different outbreak cluster composition. Centre 9 reported on only 4 clusters, which included 35 out of the 40 strains. Additionally, this centre reported on the biggest individual cluster and included 26 strains, which was a composite of C1 (ST203), C2 (ST17), and C3 (ST203) (Figure 2). However, the majority of the centres identified three separate cluster: C1, C3 (ST203) and C2 (ST17) (Figure 2, Supplementary Table S3).

Overall outbreak analysis performance

The majority vote and geometric median were calculated to evaluate centres' outbreak analyses outcome. The KP UPGMA tree-of-centres reported an identical majority vote and the geometric median (Figure 1). However, there is a difference between the majority vote and the geometric median in the VRE dataset, highlighting the vast diversity in reported clusters.

Centre 2 and 8 reported identical clusters for both the KP dataset as well as the VRE dataset. Centre 6 reported comparable clusters to Centre 2 and 8. Centre 3 was also reporting similar to Centre 2, 6, and 8 in both datasets.

Another observation is the type of cluster definitions and its wide distribution among centres. For instance, when using cg-/wgMLST schemes, this varied from 7 to 150 alleles difference (Figure 1).

Eight centres used Ridom SeqSphere+, but still, these centres reported different outbreak clusters. For the KP and VRE dataset, only four and three out of eight centres reported identical outcomes, respectively. Moreover, centres that used different bioinformatic workflows still were able to report identical outbreak clusters.

SNP-cutoff sweep

We standardized the cluster cut-offs to a range of 0 to 150 SNPs and used a single outbreak analysis tool (SKA) to remove bias that could be introduced by the various different cut-offs used by each centre. Hence, the results would give us insights into the influence of pre-processing on the outcome of each centres clusters composition. In Figure 3, the results of this analysis are visualized for all centres except for Centre 5, who submitted faulty formatted fasta files that could not be analysed. The blue bar indicates the mean Kendal-Colijn distances calculated for all centres. The red bar indicated the distances between Centre 9 and 12. A Kendall-Colijn distance of 0 would indicate no difference between the cluster composition between centres. Figure 3A and Figure 3C show that the blue bar plots start with 0 mean Kendal Colijn distance due to the absence of any clusters (data not shown). The lowest cut-off to result in a full agreement of cluster composition among all centres is 68 cgSNPs in the KP dataset. Overall, the cgSNP method (Figure 3C, 3D) results in lower mean Kendal-Colijn distances and shows a better agreement in cluster composition among centres compared to the wgSNP method (Figure 3A, 3B).

Centre 9 and 12 use identical tools and near identical settings (Figure 1), Kendal-Colijn distances are indicated by the red bars (Figure 3). The results in Figure 3 (indicated by the red bars) clearly show that the Kendall-Colijn distance is lower between Centre 9 and 12 compared to the distance between all centres (blue bars). The lowest SNP cut-off resulting in identical reporting between Centre 9 and 12 is 24 cgSNPs for the VRE dataset (Figure 3D) and 28 cgSNPs for the KP dataset (Figure 3C).

Impact on IPC measures

To study these differences in more detail and see the effect on potential IPC measures, cut-offs 5, 10, 15, and 20 cgSNPs were used to illustrate the differences between the sample-to-sample trees of Centre 9 and 12 (See Figure 4). Figure 4A illustrates the sample-to-sample trees with a cut-off of 5 SNPs. Centre 9 does not have samples clustered for the KP dataset, whereas Centre 12 already has 2 clusters of 2 samples each. For the VRE dataset, both centres have samples clustered. However, the composition of the clusters is not always identical. For instance, Centre 12 has a cluster of 5 samples, of which Centre 9 reported VRE33 not being part of a cluster and VRE13

being part of a different cluster. Furthermore, with 10, 15, and 20 SNPs (Figures 4B, 4C, and 4D), we also observe differences between Centre 9 and 12 in outbreak cluster composition for both the KP and the VRE datasets.

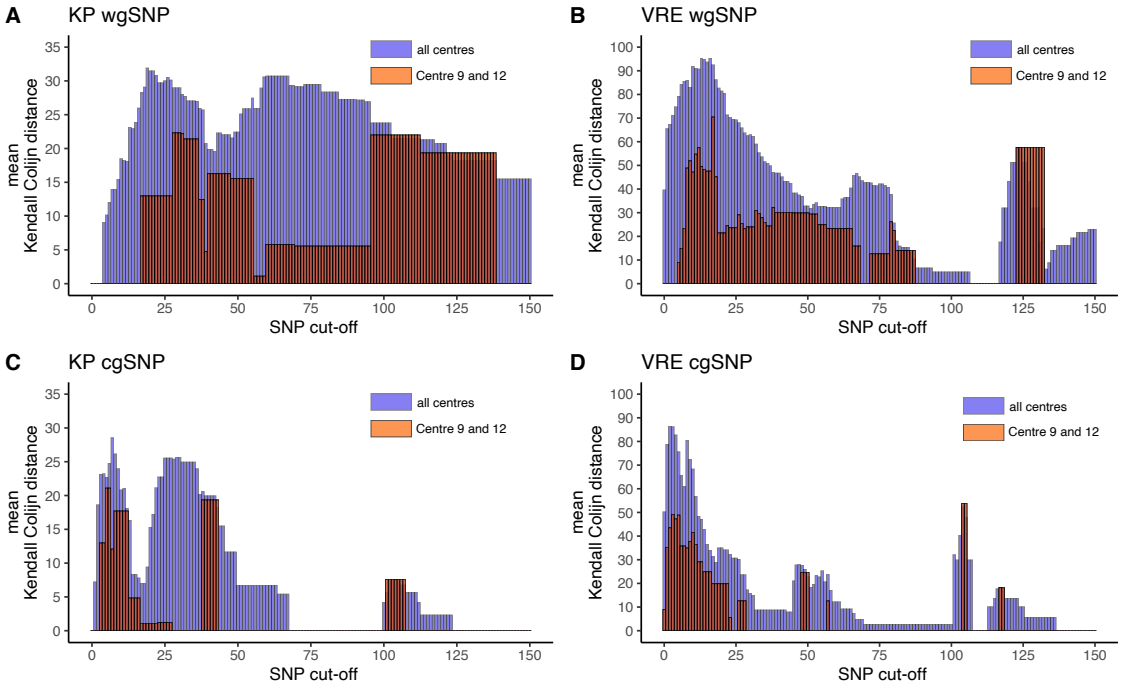


Figure 3. Sweep cut-off analysis results. The barplots in this figure illustrate the mean differences between the outbreak clusters reported among centres. For example; a distance of 0 means that centers reported identical outbreak clusters. The mean distance is calculated using the Kendall-Colijn distances metric. **A**) sweep cut-off analysis of the KP samples using the wgSNP method. **B**) sweep cut-off analysis of the VRE samples using the wgSNP method. **C**) sweep cut-off analysis of the KP samples using the cgSNP method. **D**) sweep cut-off analysis of the VRE samples using the cgSNP method.

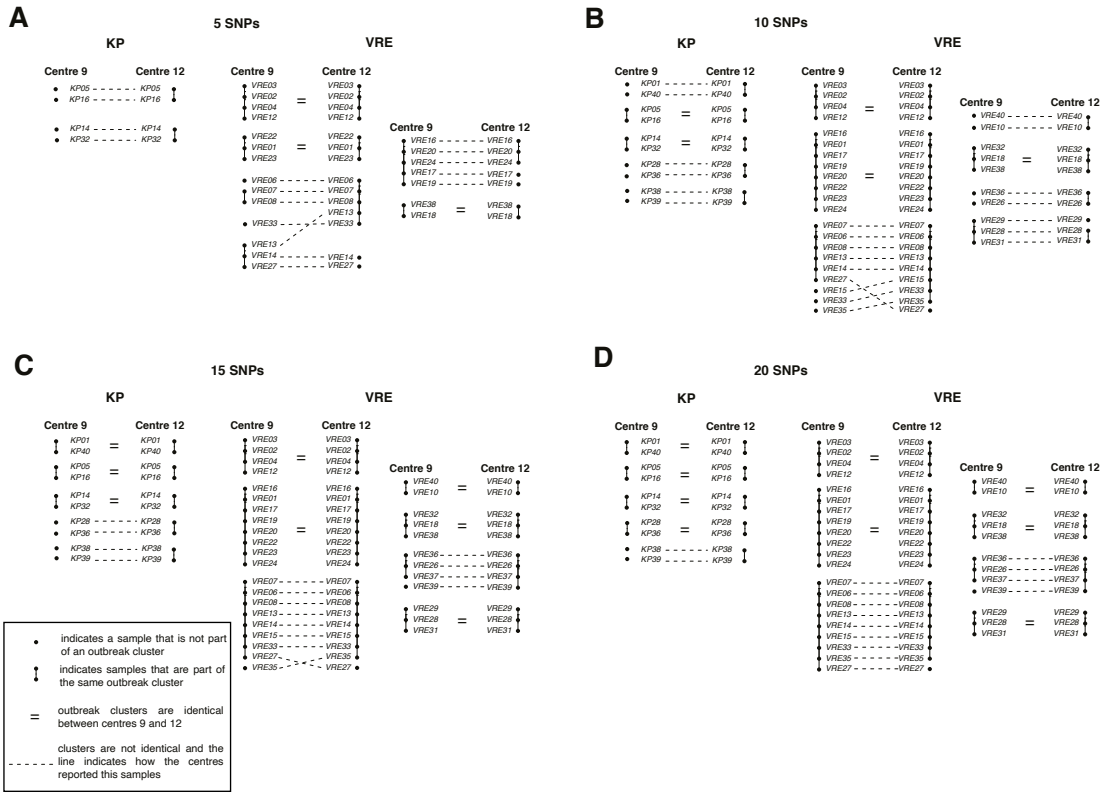


Figure 4. Illustration of differences in sample-to-sample relations between Centre 9 and Centre 12. This figure illustrates for a sweep cut-off of 5, 10, 15, and 20 SNPs using the cgSNP method the differences in outbreak cluster composition between Centre 9 and Centre 12. Both given for the KP as well as the VRE samples.

Discussion

This study aimed to assess the reproducibility of WGS-based bacterial outbreak analysis and interpretation in the medical microbiology laboratory. Thirteen Dutch hospitals and university medical centres participated and entered this study with the same WGS datasets and reported results for outbreak clusters, AMR genes, and ST. Hence, any form of variation or bias introduced during the sample preparation was mitigated.

Results presented here demonstrated an evident lack of reproducibility among centres, caused by differences in outbreak cluster definitions, bioinformatic workflow, and quality control. The four most important findings were: i) The large variety

in cluster definitions leading to a large diversity in reported outbreak clusters, which, in the current situation, makes it impossible to compare outbreak clusters across centres. ii) In light of the current situation, it is unachievable to obtain identical clusters when using a standardized cut-off because data processing introduces bias. iii) The failure of detecting specific loci, such as ESBL and housekeeping genes, due to mis-assembly and too stringent post-processing. iv) Imprecise data entry leads to erroneous conclusions. In a real-world scenario, all these issues will affect outbreak management, which impacts patient and healthcare worker safety.

To move the field of clinical bacterial typing and outbreak detection forward, we provide guidelines and recommendations based on our findings. These guidelines help to establish a workflow which has reproducible outcome, thereby minimizing the discrepancies between centres. Yet, we are aware this list is far from absolute.

- Tools: All tools used in the bioinformatic workflows should be deterministic, if possible, to guarantee fully reproducible results.
- Verification of species: Perform identification of species, to ensure proper sample handling.
- Contamination: Perform identification of sample composition using a metagenomic tool[39], as contamination will affect analyses.
- AMR detection and MLST typing: Perform gene detection preferably using a *de novo* assembly free method such as KMA[30]. This method can detect AMR and housekeeping genes using raw sequence reads as input and measure these targets' sequence depth.
- Automation of pipelines and reporting: The use of a bioinformatic management system will assist to create reproducible data analyses and facilitates standardize reporting. Furthermore, automation will limit manual intervention, which is known to be error prone.
- Harmonize workflows: Identical workflows ought to be used to be able to compare, share, and integrate data.

Outbreak cluster comparison

The differences in reported outbreak cluster composition among centres cannot be strictly appointed to the use of specific tools. No clear relation between reported cluster outcome and use of tool or methodology was observed (Figure 1). Three groups of centres (Centre 2,4,6, and 8; Centre 1 and 11; Centre 7 and 12) used different tools for outbreak analyses yet reported identical cluster compositions. On the contrary, not all centres using Ridom Seqsphere+ (eight out of thirteen centres) reported identical cluster composition. Based on these contradictory results, we cannot appoint the effect of a particular tool on cluster composition.

To exclude the possibility that all bias was introduced using different thresholds or different outbreak analysis tools, we used a single tool and a range of thresholds to determine the cluster composition of the assemblies generated in each centre. This analysis clearly illustrated that using a single outbreak analysis tool and defining standardized SNP cut-offs is not sufficient to obtain identical cluster compositions, since the impact of pre-processing already heavily impacts the cluster outcome (Figure 3). Even when comparing the two most closely related centres in terms of methodology and tools used, we still observe differences in outcome, leading to significant implications for outbreak management and IPC. Figure 4 highlights the differences in outcome in a sample-to-sample comparison. These findings support the need for a more standardized way of bacterial outbreak analysis to circumvent most of these short-comings.

In our final analysis, we focused on SNP analysis and determining outbreak clusters using SNP cut-offs. These cut-offs are often calculated ad hoc[15,40] and differ significantly among studies[41]. Combined with our findings, we can conclude that using these cut-offs when using non-identical bioinformatic outbreak analysis workflows is futile. Other analysis strategies have been proposed, for example, a method that uses a probabilistic method to infer transmissions to help solving these[42].

ST

All centres were in excellent concordance on the STs of the strains used in this study, however reporting-errors were detected for two VREs strains (VRE18 and VRE33), potentially impacting the final epidemiological assessment. All but one centre reported these two strains as ST17 and ST203, common nosocomial VRE[43–45]. One centre classified these two strains to an ST with an absent *pstS*, one of the seven genes in the MLST scheme for *E. faecium*. This would indicate the presence of rare types of VRE. However, no *pstS*-null *vanB* VRE has been reported, and only recently, the first non-typeable VRE isolates associated with a *pstS*-null genotype carrying a *vanA* cassette have been described in Australia, Korea, and Scotland[38,46,47]. The misinterpretation was introduced by mis-assembly or too stringent post-processing and may lead to different interpretations when reporting on routine surveillance or bacterial outbreaks.

AMR

Not all centres reported on the presence of specific beta-lactamase genes. Also, high variation was observed in the reported SHV genes. Many, but not all, of these bla_{SHV} genes result in an ESBL phenotype[48]. In addition, not all $bla_{CTX\text{-}M}$ genes were recovered by some centres. In a study investigating the reproducibility of AMR gene reporting, Doyle *et al*. reported similar discordance. However, the discordance in the reported gene variant was only minor in the genotypic resistance prediction[26].

Contradictory, in our study, both discordance in the gene variant reporting and false absence of ESBL genes was observed. Although we did not request genotypic resistance prediction reporting, failure to detect ESBL genes will influence resistance prediction. This can be of major impact as international guidelines advise contact isolation for patients carrying ESBL *Enterobacteriaceae*[49]. This problem is minor in practice, as strains are commonly phenotypically characterised for their AMR profile in a clinical microbiology laboratory. Analysis of how the false absence of the ESBL genes occurred demonstrated that centres that missed ESBL genes ($bla_{\text{ctx-m-14}}$) removed all small contigs of up to 1kb during post-processing. Resistance genes are often located on transposons or plasmids, which are difficult to assemble using short-read sequencing data. These hard to assemble regions can then be assembled into small contigs, sometimes of <1kb, which would be removed by stringent post-processing. Normally, small contigs are removed as they are often associated with contamination. To overcome the failure of detecting AMR genes, one could use an assembly free method such as KMA[30] or ARIBA[50] or simply retain these small contigs.

Data entry

This study evaluates the reporting of AMR, ST, and outbreak clusters, performed by molecular trained staff, to IPC teams, thereby mimicking a crucial procedure in outbreak management. However, in this study, we found multiple incidences of inaccurate or incorrect reporting of results, such as, i) swapping of samples KP12 with KP13 and KP23 with KP24 by Centre 3, and, ii) incomplete reporting of sample-to-sample relations, which mainly occurred by Centre 9 in the VRE dataset (data not shown). These flaws in data entry can have significant consequences for IPC. It may result in extra costs and could potentially miss or identify new faulty outbreaks, leading to further transmission and follows into the closure of hospital wards, lack of patient safety, and even loss of human lives[51]. When implementing WGS procedures, medical microbiology laboratories should carefully follow international norms and guidelines relating to data management[52].

Limitations

We are aware that this study is focused on the dry-lab part of outbreak analysis, thereby not taking into account the wet-lab. Assessing the combination of wet- and dry-lab will result in even larger discrepancies than observed in our study. To date, little effort has been conducted to assess the reproducibility of outbreak analyses in a clinical context. Wet-lab reproducibility has been previously evaluated but all used a single bioinformatic analysis method[24,53]. Notwithstanding, the current situation is that centres in The Netherlands that adopted WGS-based outbreak analyses use a plethora of bioinformatic workflows. As a result, centres may communicate outcomes to each other without knowing if these results are interchangeable and may not be reproducible. Moreover, communicating outbreaks to national reference laborato-

ries for surveillance and monitoring purposes is essential to mitigate nationwide outbreaks and prevent further spread.

Conclusion

To conclude, our study demonstrates limited reproducibility among centres applying WGS for bacterial outbreaks and AMR detection in the Netherlands. This will inevitably negatively impact IPC, healthcare workers', and patients' health and safety. Therefore, we advise the need for more collaboration among centres to better assess outbreaks and AMR detection through optimization and harmonization of bioinformatic tools. This would include extensive proficiency testing, open-source data sharing, and formulation of guidelines[54]. Eventually, leading to harmonization of protocols and guidelines to minimize centre to centre variability. Furthermore, we provided guidelines for bioinformatic workflow setup, which would address most of the issues detected in this study.

Acknowledgements

All members of the SIG Bioinformatics in Medical Microbiology NL Consortium for supporting this study initiative. In particular the following who actively contributed to this study: W.A. van der Reijden (Streeklab Haarlem); S.D. Pas, A. Vrolijk and S. Oome (Microvida); Y. Bisselink, F. Bosma and S. Rosema (UMCG); L.M. Schouls, A. Hendrickx, M. van Santen, S. Witteveen (RIVM); S. Hermann and W. Bekers (UMCU); B.C.L. van der Putten (AIGHD); J.J. Verweij and J.J.J.M. Stohr (ETZ); J.M. Fonville and P. van Alphen (PAMM); E.C.J. Claas, M.E.M. Kraakman and S. Nooij (LUMC); A. Burggraaf and P.W. Smit (Maasstad Hospital).

Author contributions

S.P.M, H.F.L.W. and L.B.A. conceived and supervised the study. J. P. M. C. and C.J. designed and performed the meta-analysis, data analysis, created figures, and wrote manuscript. J.W.A.R. performed data interpretation and assisted in writing. Members of the SIG Bioinformatics in Medical Microbiology NL Consortium co-designed the study and participated in the study. All authors read and approved the final manuscript.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Funding information

The authors received no specific grant from any funding agency.

References

1. **European Centre for Disease Prevention and Control**. *AER for 2017: Healthcare-associated infections acquired in intensive care units*. 2017.
2. **Barrasa-Villar JI, Aibar-Remón C, Prieto-Andrés P, Mareca-Doñate R, Moliner-Lahoz J**. Impact on Morbidity, Mortality, and Length of Stay of Hospital-Acquired Infections by Resistant Microorganisms. *Clin Infect Dis*. Epub ahead of print 2017. DOI: 10.1093/cid/cix411.
3. **Founou RC, Founou LL, Essack SY**. Clinical and economic impact of antibiotic resistance in developing countries: A systematic review and meta-analysis. *PLoS One* 2017;12:e0189621.
4. **Bannerman TL, Hancock GA, Tenover FC, Miller JM**. Pulsed-field gel electrophoresis as a replacement for bacteriophage typing of Staphylococcus aureus. *J Clin Microbiol* 1995;33:551–555.
5. **Urwin R, Maiden MCJ**. Multi-locus sequence typing: A tool for global epidemiology. *Trends in Microbiology* 2003;11:479–487.
6. **Vos P, Hogers R, Bleeker M, Reijans M, Lee T Van De, *et al.*** AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res* 1995;23:4407–4414.
7. **van Belkum A, Tassios PT, Dijkshoorn L, Haeggman S, Cookson B, *et al.*** Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clin Microbiol Infect* 2007;13:1–46.
8. **Bletz S, Mellmann A, Rothgänger J, Harmsen D**. Ensuring backwards compatibility: Traditional genotyping efforts in the era of whole genome sequencing. *Clin Microbiol Infect* 2015;21:347.e1-347.e4.
9. **Graham RMA, Doyle CJ, Jennison A V.** Real-time investigation of a Legionella pneumophila outbreak using whole genome sequencing. *Epidemiol Infect* 2014;142:2347–2351.
10. **Quainoo S, Coolen JPM, van Hijum SAFT, Huynen MA, Melchers WJG, *et al.*** Whole-genome sequencing of bacterial pathogens: The future of nosocomial outbreak analysis. *Clinical Microbiology Reviews* 2017;30:1015–1063.
11. **Schürch AC, van Schaik W**. Challenges and opportunities for whole-genome sequencing–based surveillance of antibiotic resistance. *Ann N Y Acad Sci* 2017;1388:108–120.
12. **Gilchrist CA, Turner SD, Riley MF, Petri WA, Hewlett EL**. Whole-genome sequencing in outbreak analysis. *Clin Microbiol Rev* 2015;28:541–563.
13. **Roetzer A, Diel R, Kohl TA, Rückert C, Nübel U, *et al.*** Whole Genome Sequencing versus Traditional Genotyping for Investigation of a Mycobacterium tuberculosis Outbreak: A Longitudinal Molecular Epidemiological Study. *PLoS Med* 2013;10:e1001387.
14. **Gray RR, Tatem AJ, Johnson JA, Alekseyenko A V., Pybus OG, *et al.*** Testing spatiotemporal hypothesis of bacterial evolution using methicillin-resistant staphylococcus aureus ST239 genome-wide data within a bayesian framework. *Molecular Biology and Evolution* 2011;28:1593–1603.
15. **Harris SR, Cartwright EJP, Török ME, Holden MTG, Brown NM, *et al.*** Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: A descriptive study. *Lancet Infect Dis* 2013;13:130–136.
16. **Snitkin ES, Zelazny AM, Thomas PJ, Stock F, Henderson DK, *et al.*** Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. *Sci Transl Med* 2012;4:148ra116-148ra116.
17. **Bastiaens GJH, Cremers AJH, Coolen JPM, Nillesen MT, Boeree MJ, *et al.*** Nosocomial outbreak of multi-resistant Streptococcus pneumoniae serotype 15A in a centre for chronic pulmonary diseases. *Antimicrob Resist Infect Control*;7. Epub ahead of print 2018. DOI: 10.1186/s13756-018-0457-3.
18. **Hughes A, Ballard S, Sullivan S, Marshall C**. An outbreak of vanA vancomycin-resistant Enterococcus faecium in a hospital with endemic vanB VRE. *Infect Dis Heal* 2019;24:82–91.
19. **Cremers AJH, Coolen JPM, Bleeker-Rovers CP, van der Geest-Blankert ADJ, Haverkate D, *et al.*** Surveillance-embedded genomic outbreak resolution of methicillin-susceptible Staphylococcus aureus in a neonatal intensive care unit. *Sci Rep* 2020;10:1–10.
20. **Protonotariou E, Poulou A, Politi L, Sgouropoulos I, Metallidis S, *et al.*** Hospital outbreak due to a Klebsiella pneumoniae ST147 clonal strain co-producing KPC-2 and VIM-1 carbapenemases in a tertiary teaching hospital in Northern Greece. *Int J Antimicrob Agents* 2018;52:331–337.
21. **van Ingen J, Kohl TA, Kranzer K, Hasse B, Keller PM, *et al.*** Global outbreak of severe Mycobacterium chimaera disease after cardiac surgery: a molecular epidemiological study. *Lancet Infect Dis* 2017;17:1033–1041.
22. **Hopman J, Meijer C, Kenters N, Coolen JPM, Ghamati MR, *et al.*** Risk Assessment After a Severe Hospital-Acquired Infection Associated With Carbapenemase-Producing Pseudomonas aeruginosa. *JAMA Netw open* 2019;2:e187665.
23. **Adrian E, Blanc Dominique S, Gilbert G, Keller Peter M, Vladimir L, *et al.*** Improving the quality

and workflow of bacterial genome sequencing and analysis: Paving the way for a Switzerland-wide molecular epidemiological surveillance platform. *Swiss Medical Weekly*;148. Epub ahead of print December 2018. DOI: 10.4414/smw.2018.14693.

24. **Mellmann A, Andersen PS, Bletz S, Friedrich AW, Kohl TA, *et al.*** High interlaboratory reproducibility and accuracy of next-generation-sequencing-based bacterial genotyping in a ring trial. *J Clin Microbiol* 2017;55:908–913.
25. **Jamin C, de Koster S, van Koeveringe S, de Coninck D, Mensaert K, *et al.*** Harmonization of whole genome sequencing for outbreak surveillance of Enterobacteriaceae and Enterococci. *bioRxiv* 2020;2020.11.20.392399.
26. **Doyle RM, O'Sullivan DM, Aller SD, Bruchmann S, Clark T, *et al.*** Discordant bioinformatic predictions of antimicrobial resistance from whole-genome sequencing data of bacterial isolates: an inter-laboratory study. *Microb Genomics*;6. Epub ahead of print February 2020. DOI: 10.1099/mgen.0.000335.
27. **Kluytmans-van den Bergh MFQ, Bruijning-Verhagen PCJ, Vandenbroucke-Grauls CMJE, de Brauwer EIGB, Buiting AGM, *et al.*** Contact precautions in single-bed or multiple-bed rooms for patients with extended-spectrum β-lactamase-producing Enterobacteriaceae in Dutch hospitals: a cluster-randomised, crossover, non-inferiority study. *Lancet Infect Dis* 2019;19:1069–1079.
28. **Seemann T**. ABRicate. https://github.com/tseemann/abricate.
29. **Seemann T**. mlst. https://github.com/tseemann/mlst.
30. **Clausen PTLC, Aarestrup FM, Lund O**. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinformatics*. Epub ahead of print 2018. DOI: 10.1186/s12859-018-2336-6.
31. **Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, *et al.*** Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–2504.
32. **Kendall M, Colijn C**. Mapping Phylogenetic Trees to Reveal Distinct Patterns of Evolution. *Mol Biol Evol* 2016;33:2735–2743.
33. **Jombart T, Kendall M, Almagro-Garcia J, Colijn C**. treespace: Statistical exploration of landscapes of phylogenetic trees. *Mol Ecol Resour* 2017;17:1385–1392.
34. **Harris SR**. SKA: Split Kmer Analysis Toolkit for Bacterial Genomic Epidemiology. *bioRxiv* 2018;453142.
35. **Seemann T, Klötzl F, Page AJ**. snp-dists: Pairwise SNP distance matrix from a FASTA sequence alignment. *2018-10-27*.
36. **Csardi G, Nepusz T**. The igraph software package for complex network research. *InterJournal Complex Syst*.
37. **Rodríguez-Medina N, Barrios-Camacho H, Duran-Bedolla J, Garza-Ramos U**. Klebsiella variicola: an emerging pathogen in humans. *Emerging Microbes and Infections* 2019;8:973–988.
38. **Lemonidis K, Salih TS, Dancer SJ, Hunter IS, Tucker NP**. Emergence of an Australian-like pstS-null vancomycin resistant Enterococcus faecium clone in Scotland. *PLoS One*;14. Epub ahead of print June 2019. DOI: 10.1371/journal.pone.0218185.
39. **Ye SH, Siddle KJ, Park DJ, Sabeti PC**. Benchmarking Metagenomics Tools for Taxonomic Classification. *Cell* 2019;178:779–794.
40. **Goyal M, Javerliat F, Palmieri M, Mirande C, Van Wamel W, *et al.*** Genomic evolution of staphylococcus aureus during artificial and natural colonization of the human nose. *Front Microbiol* 2019;10:1525.
41. **Miro E, Rossen JWA, Chlebowicz MA, Harmsen D, Brisse S, *et al.*** Core/Whole Genome Multilocus Sequence Typing and Core Genome SNP-Based Typing of OXA-48-Producing Klebsiella pneumoniae Clinical Isolates From Spain. *Front Microbiol* 2020;10:2961.
42. **Stimson J, Gardy J, Mathema B, Crudu V, Cohen T, *et al.*** Beyond the SNP Threshold: Identifying Outbreak Clusters Using Inferred Transmissions. *Mol Biol Evol* 2019;36:587–603.
43. **Lam MMC, Seemann T, Tobias NJ, Chen H, Haring V, *et al.*** Comparative analysis of the complete genome of an epidemic hospital sequence type 203 clone of vancomycin-resistant Enterococcus faecium. *BMC Genomics*. Epub ahead of print 2013. DOI: 10.1186/1471-2164-14-595.
44. **Kuo AJ, Shu JC, Liu TP, Lu JJ, Lee MH, *et al.*** Vancomycin-resistant Enterococcus faecium at a university hospital in Taiwan, 2002–2015: Fluctuation of genetic populations and emergence of a new structure type of the Tn1546-like element. *J Microbiol Immunol Infect*. Epub ahead of print 2018. DOI: 10.1016/j.jmii.2018.08.008.
45. **Johnson PDR, Ballard SA, Grabsch EA, Stinear TP, Seemann T, *et al.*** A Sustained Hospital Outbreak of Vancomycin-Resistant Enterococcus faecium Bacteremia due to Emergence of vanB E. faecium Sequence Type 203. *J Infect Dis*. Epub ahead of print 2010. DOI: 10.1086/656319.

46. **Kim HM, Chung DR, Cho SY, Huh K, Kang CI, *et al.*** Emergence of vancomycin-resistant Enterococcus faecium ST1421 lacking the pstS gene in Korea. *Eur J Clin Microbiol Infect Dis*. Epub ahead of print 2020. DOI: 10.1007/s10096-020-03853-4.
47. **Leong KWC, Kalukottege R, Cooley LA, Anderson TL, Wells A, *et al.*** State-Wide Genomic and Epidemiological Analyses of Vancomycin-Resistant Enterococcus faecium in Tasmania's Public Hospitals. *Front Microbiol*. Epub ahead of print 2020. DOI: 10.3389/fmicb.2019.02940.
48. **Liakopoulos A, Mevius D, Ceccarelli D**. A review of SHV extended-spectrum β-lactamases: Neglected yet ubiquitous. *Frontiers in Microbiology* 2016;7:1374.
49. **Tacconelli E, Cataldo MA, Dancer SJ, De Angelis G, Falcone M, *et al.*** ESCMID guidelines for the management of the infection control measures to reduce transmission of multidrug-resistant Gram-negative bacteria in hospitalized patients. *Clin Microbiol Infect*. Epub ahead of print 2014. DOI: 10.1111/1469-0691.12427.
50. **Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, *et al.*** ARIBA: Rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genomics*. Epub ahead of print 2017. DOI: 10.1099/mgen.0.000131.
51. **WHO**. *Global Antimicrobial Resistance and Use Surveillance System (GLASS) Report*. 2020.
52. **Turner P, Fox-Lewis A, Shrestha P, Dance DAB, Wangrangsimakul T, *et al.*** Microbiology Investigation Criteria for Reporting Objectively (MICRO): A framework for the reporting and interpretation of clinical microbiology data. *BMC Med*. Epub ahead of print 2019. DOI: 10.1186/s12916-019-1301-1.
53. **Nouws S, Bogaerts B, Verhaegen B, Denayer S, Piérard D, *et al.*** Impact of DNA extraction on whole genome sequencing analysis for characterization and relatedness of Shiga toxin-producing Escherichia coli isolates. *Sci Rep*. Epub ahead of print 2020. DOI: 10.1038/s41598-020-71207-3.
54. **Savelkoul PHM, Koopmans MPG, Schouls L, van Rhee-Luderer R, Ossewaarde JM, *et al.*** *Richtlijn Moleculaire typering in het kader van infectiepreventie*. 2018.

Supplementary Materials

Supplementary Table 1 metadata of datasets used for this study

Supplementary Table 2 sequence types for *Klebsiella pneumoniae* reported by each centre

	Center												
	1	2	3	4	5	6	7	8	9	10	11	12	13
KP01	48	48	48	48	48	48	48	48	48	48	48	48	48
KP02	359	359	359	359	359	359	359	359	359		359	359	359
KP03	873	873	873	873	873	873	873	873	873	873	873	873	873
KP04	1836	1836	1836	1836	1836	1836	1836	1836	1836	1836	1836	1836	1836
KP05	1593	1593	1593	1593	1593	1593	1593	1593	1593	1593	1593	1593	1593
KP06	392	392	392	392	392	392	392	392	392	392	392	392	392
KP07	37	37	37	37	37	37	37	37	37	37	37	37	37
KP08	261	261	261	261	261	261	261	261	261	261	261	261	261
KP09	252	252	252	252	252	252	252	252	252	252	252	252	252
KP10	29	29	29	29	29	29	29	29	29	29	29	29	29
KP11	395	395	395	395	395	395	395	395	395	395	395	395	395
KP12	1193	1193	485	1193	1193	1193	1193	1193	1193	1193	1193	1193	1193
KP13	485	485	1193	485	485	485	485	485	485	485	485	485	485
KP14	437	437	437	437		437	437	437	437	437	437	437	437
KP15	15	15	15	15	15	15	15	15	15	15	15	15	15
KP16	1593	1593	1593	1593	1593	1593	1593	1593	1593	1593	1593	1593	1593
KP17	17	17	17	17	17	17	17	17	17	17	17	17	17
KP18	904	904	904	904	904	904	904	904	904	904	904	904	904
KP19	101	101	101	101	101	101	101	101	101	101	101	101	101
KP20	1661	1661	1661	1661		1661	1661	1661	1661	1661	1661	1661	1661
KP21	37	37	37	37		37	37	37	37	37	37	37	37
KP22	37	37	37	37	37	37	37	37	37	37	37	37	37
KP23	641	641	111	641	641	641	641	641	641	641		641	641
KP24	111	111	641	111	111	111	111	111	111	111	111	111	111
KP25	252	252	252	252	252	252	252	252	252	252	252	252	252
KP26	252	252	252	252	252	252	252	252	252	252	252	252	252
KP27	101	101	101	101	101	101	101	101	101	101	101	101	101
KP28	307	307	307	307	307	307	307	307	307	307	307	307	307
KP29	45	45	45	45	45	45	45	45	45	45	45	45	45
KP30	105	105	105	105	105	105	105	105	105	105	105	105	105
KP31	1777	1777	1777	1777	1777	1777	1777	1777	1777	1777	1777	1777	1777
KP32	437	437	437	437		437	437	437	437	437	437	437	437
KP33	33	New ST, nearest =ST33							new		33		
KP34	273	273	273	273	273	273	273	273	273	273	392	273	273
KP35	869	869	869	869		869	869	869	869	869	869	869	869
KP36	307	307	307	307	307	307	307	307	307	307	307	307	307
KP37	1626	1626	1626	1626	1626	1626	1626	1626	1626	1626	1626	1626	1626
KP38	3822	3822	3822	3822		3822	3822	3822	3822	3822	3822	3822	3822
KP39	3822	3822	3822	3822		3822	3822	3822	3822	3822	3822	3822	3822
KP40	48	48	48	48	48	48	48	48	48	48	48	48	48

majority
discrepancy
no ST reported

Supplementary Table 3 sequence types for Enterococcus faecium reported by each centre

	Centre												
SampleName	1	2	3	4	5	6	7	8	9	10	11	12	13
VRE01	203	203	203	203	203	203	203	203	203	203	203	203	203
VRE02	203	203	203	203	203	203	203	203	203	203	203	203	203
VRE03	203	203	203	203	203	203	203	203	203	203	203	203	203
VRE04	203	203	203	203	203	203	203	203	203	203	203	203	203
VRE05	117	117	117	117	117	117	117	117	117	117	117	117	117
VRE06	17	17	17	17	17	17	17	17	17	17	17	17	17
VRE07	17	17	17	17	17	17	17	17	17	17	17	17	17
VRE08	17	17	17	17	17	17	17	17	17	17	17	17	17
VRE09	80	80	80	80	80	80	80	80	80	80	80	80	80
VRE10	18	18	18	18	18	18	18	18	18	18	18	18	18
VRE11	192	192	192	192	192	192	192	192	192	192	192	192	192
VRE12	203	203	203	203	203	203	203	203	203	203	203	203	203
VRE13	17	17	17	17	17	17	17	17	17	17	17	17	17
VRE14	17	17	17	17	17	17	17	17	17	17	17	17	17
VRE15	17	17	17	17	17	17	17	17	17	17	17	17	17
VRE16	203	203	203	203	203	203	203	203	203	203	203	203	203
VRE17	203	203	203	203	203	203	203	203	203	203	203	203	203
VRE18	203	203	203	203	203	203	203	203	1489	203	203	203	203
VRE19	203	203	203	203	203	203	203	203	203	203	203	203	203
VRE20	203	203	203	203	203	203	203	203	203	203	203	203	203
VRE21	203	203	203	203	203	203	203	203	203	203	203	203	203
VRE22	203	203	203	203	203	203	203	203	203	203	203	203	203
VRE23	203	203	203	203	203	203	203	203	203	203	203	203	203
VRE24	203	203	203	203	203	203	203	203	203	203	203	203	203
VRE25	203	203	203	203	203	203	203	203	203	203	203	203	203
VRE26	192	192	192	192	192	192	192	192	192	192	192	192	192
VRE27	17	17	17	17	17	17	17	17	17	17	17	17	17
VRE28	80	80	80	80	80	80	80	80	80	80	80	80	80
VRE29	80	80	80	80	80	80	80	80	80	80	80	80	80
VRE30	192	192	192	192	192	192	192	192	192	192	192	192	192
VRE31	80	80	80	80	80	80	80	80	80	80	80	80	80
VRE32	203	203	203	203	203	203	203	203	203	203	203	203	203
VRE33	17	17	17	17	17	17	17	17	1421	17	17	17	17
VRE34	17	17	17	17	17	17	17	17	17	17	17	17	17
VRE35	17	17	17	17	17	17	17	17	17	17	17	17	17
VRE36	192	192	192	192	192	192	192	192	192	192	192	192	192
VRE37	192	192	192	192	192	192	192	192	192	192	192	192	192
VRE38	203	203	203	203	203	203	203	203	203	203	203	203	203
VRE39	192	192	192	192	192	192	192	192	192	192	192	192	192
VRE40	18	18	18	18	18	18	18	18	18	18	18	18	18

majority
discrepancy
no ST reported

4

DATASETS FOR BENCHMARKING ANTIMICROBIAL RESISTANCE GENES IN BACTERIAL METAGENOMIC AND WHOLE GENOME SEQUENCING

Amogelang R. Raphenya[1,2,3] James Robertson[4], **Casper Jamin[5],** Leonardo de Oliveira Martins[6], Finlay Maguire[7,8,9], Andrew G. McArthur[1,2,3] & John P. Hays[10]

1. David Braley Centre for Antibiotic Discovery, McMaster University, Hamilton, Ontario, L8S 4K1, Canada.
2. Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Ontario, L8S 4K1, Canada.
3. Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario, L8S 4K1, Canada.
4. National Microbiology Laboratory, Public Health Agency of Canada, Guelph, Ontario, N1G 3W4, Canada.
5. Department of Medical Microbiology, care and Public Health Research institute (CAPHRi), Maastricht University Medical Center, P. Debyelaan 25, 6229HX, Maastricht, the Netherlands.
6. Quadram institute Bioscience, norwich Research Park, Norwich, NR4 7UQ, UK.
7. Department of community Health & epidemiology, Dalhousie University, Halifax, Nova Scotia, B3H 4R2, Canada.
8. faculty of computer Science, Dalhousie University, Halifax, nova Scotia, B3H 4R2, Canada.
9. Shared Hospital Laboratory, Sunnybrook Health Sciences Centre, Toronto, Ontario, M4N 3M5, Canada.
10. Department of Medical Microbiology & infectious Diseases, erasmus University Medical centre Rotterdam (Erasmus MC), Doctor Molewaterplein 40, 3015 GD, Rotterdam, the Netherlands.

Published: Raphenya, A. R., Robertson, J., **Jamin, C**., de Oliveira Martins, L., Maguire, F., McArthur, A. G., & Hays, J. P. (2022). Datasets for benchmarking antimicrobial resistance genes in bacterial metagenomic and whole genome sequencing. Scientific data, 9(1), 1-6.

Background & Summary

Whole genome sequencing (WGS) is a technique used to analyse the genomes of both prokaryotic and eukaryotic organisms. This includes a range of approaches including WGS of individual isolates (either via culture or single-cell methods) and the related simultaneous sequencing of all organisms present in a given sample (i.e., metagenomics)[1]. There are also a range of different sequencing technologies available such as technologies that generate 'short-read' or 'long-read' sequences[2]. Within the field of microbiology, sequencing is a valuable tool for mapping the epidemiology of bacterial isolates associated with clinical outbreaks of disease[3], as well as for the identification of potentially pathogenic strains of bacteria that could be present in both food and environ- mental samples[4]. It is increasingly common to use sequencing to identify the type and range of antimicrobial resistance (AMR) genes present in bacterial isolates in order to make predictions regarding the actual bacterial phenotype of particular isolates[5,6]. These data have the potential to guide antibiotic treatment decisions and patient therapy in clinical cases of disease[7]. However, many different bioinformatic software and pipelines exist to predict AMR genes in genomic and metagenomic sequencing data. These include methods designed to directly analyse unassembled short and long-reads as well as those involving the assembly of these reads into contiguous bacterial chromosomes, partial chromosomes (contigs) and/or mobile genetic elements, such as plasmids[8–10]. The ability to systematically compare and benchmark the range of WGS algorithms and pipelines available on a common dataset would provide increased confidence in the validity of interpreting the results of WGS genotyping, AMR carriage, and the inferred bacterial AMR phenotype[11–13]. Such benchmarking activities would be promoted by the availability of common gold standard reference datasets containing raw sequencing reads, contigs, chromosomes, and plasmid data[14] and including software associated with the assembly of both short and long-read sequence results[15]. Such a gold standard reference set of bacterial WGS data (focussing on short read sequence data and including simulated metagenomic data) was generated during the Microbial Bioinformatics Hackathon and Workshop 2021, which took place virtually between the 11th and 13th October, 2021. The event was jointly organized by the Public Health Alliance for Genomic Epidemiology (PHA4GE), the Joint Programming Initiative on Antimicrobial Resistance (JPIAMR), and the Cloud Infrastructure for Big Data Microbial Bioinformatics (CLIMB-BIG-DATA) initiative[16].

Methods

A selection of benchmarking genomes was made by prioritizing ESKAPE pathogens (i.e., *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* spp.) in addition to *Salmonella spp.* We selected only complete genomes from the NCBI Database Repository for Genome Access[17], where the primary sequence data was available and the Illumina data deposited included >40X coverage and >100 bp sequence read length.

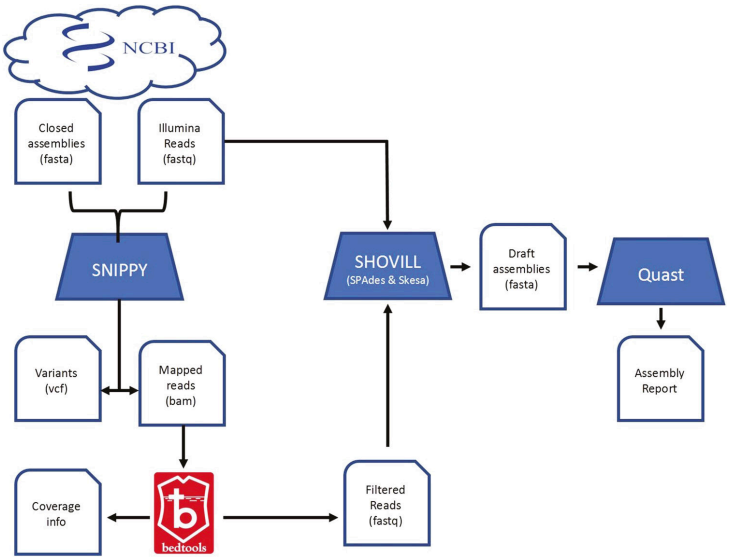


Figure 11. Diagram illustrating the sequence of steps and software involved in generating 'gold standard' bacterial whole genome sequence datasets for benchmarking bacterial assembly and prediction software.

Candidate genomes were processed using the workflow depicted in Fig. 1, with the genomes filtered according to the criteria described below. Initially, Illumina read sets were downloaded from NCBI and assembled using shovill v. 1.1.0[18] using both SPades[19] and Skesa[20]. Assembly metrics were determined using Quast v. 5.0.2[21] and assemblies with N50 <50Kb and >100 contigs were excluded. Illumina reads were mapped against their corresponding NCBI genome using SNIPPY v. 4.3.6[22] using the default parameters (minimum coverage depth = 10, minimum VCF quality = 100, minimum fraction = auto). Regions of 0 read coverage were identified using bedtools v. 2.29.2[23] and genomes with >200Kb of no Illumina read coverage were excluded. Additionally, any samples where there were >10 SNPs detected by SNIPPY between the Illumina reads and its corresponding assembly were excluded. The

mapped reads from the BAM were sorted so that read names appeared sequentially before extracting the reads using bedtools v. 2.29.2 bamtofastq functionality. If the extracted read coverage depth was <40X it was excluded from further analysis. Reads were then assembled in the same manner as the unfiltered reads and samples were excluded if their assembly metrics did not meet the criteria above. AMR genes were predicted from each assembly using the Comprehensive Antibiotic Resistance Database (CARD)'s Resistance Gene Identifier (RGI) software v.5.2.0 and CARD reference data v.3.1.4[24].

To generate a simulated metagenomic benchmarking dataset, a reproducible nextflow[25] simulation workflow was used. The generated gold-standard WGS assemblies were randomly amplified following a log-normal distribution ($\mu = 1$ $\sigma = 2$) to represent observed metagenomic species distributions[26]. Additional CARD (v3.1.4) AMR reference genes were randomly inserted into the contigs to ensure representation of the full canonical CARD database in the metagenome. ART v2.5.8[27] was then used to simulate 2.49 million 250 bp paired-end reads from these sequences using the Illumina MiSeqV3 error profile. Finally, using pysam (v0.16.0.1)[27,28] and bedtools (v2.30.0)[23] labels were generated for each read with the RGI (v5.2.0) annotated AMR gene from which that read was simulated.

We selected RGI as it performs at par with other AMR tools evaluated using the hAMRonization workflow[29]. The hAMRonization workflow uses 12 different AMR tools to predict AMR genes in genomic data and produces a standard report to compare results across tools. Five of these 12 tools work with genomic reads, while the other 7 use assembled genomes. Analysis of 94 from 174 selected genomes was performed via the hAMRonization workflow using the 5 tools associated with assembled genome analysis. The RGI results produced were similar to the other 4 tools tested i.e., abricate, csstar, resfinder, and srax. The results are presented as a radar plot in Fig. 2 and available at Zenodo[30].

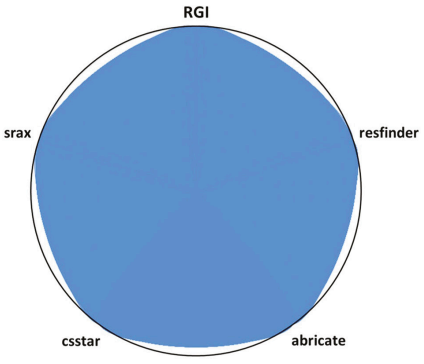


Figure 12. Radar plot showing 94 samples analyzed using hAMRonization workflow. There are 579 genes comparing presence or absence for all the 5 tools tested.

Data Records

The datasets are suitable for different AMR detection pipelines, as they provide assemblies using two different widely used assemblers in addition to mapped reads from the primary data used to generate the assembly for 174 bacterial genomes representing 22 distinct species (Table 1). To enable benchmarking of metagenomic AMR detection pipelines, these datasets also provide simulated metagenomic reads and a "perfect" metagenomic assembly derived from these 174 assemblies. Since it is possible for records to be updated in NCBI, we have included reads in the dataset to ensure that they can be consistently used. Due to the size of the data, we have split the dataset into assemblies, 6 batches of genomic reads, and a separate metagenomic dataset (including assemblies, reads, and label information). The assemblies (which include closed, draft versions for raw and filtered datasets) are located at Zenodo[31].The mapped raw reads (BAM files) are located at Zenodo:

Mapped Read Sets – 1[32]
Mapped Read Sets – 2[33]
Mapped Read Sets – 3[34]
Mapped Read Sets – 4[35]
Mapped Read Sets – 5[36]
Mapped Read Sets – 6[37]

The simulated metagenomic data (reads, assemblies, labels, simulation configuration) are located at Zenodo[38], with corresponding simulation workflow available at Zenodo[39]. The corresponding metadata for all isolates can be found can be found at Zenodo[30]. The Resistance Gene Identifier predictions can be found at Zenodo[30]. Note that each file name is the complete assemblies' accession number.

Table 1. Taxanomic composition of the benchmarking dataset

Organism	Sample Count	Organism	Sample Count
Acinetobacter baumannii	5	*Escherichia coli*	18
Aeromonas veronii	1	*Klebsiella aerogenes*	3
Citrobacter freundii	4	*Klebsiella oxytoca*	4
Enterobacter asburiae	2	*Klebsiella pneumoniae*	56
Enterobacter bugandensis	1	*Kluyvera intermedia*	1
Enterobacter cancerogenus	1	*Providencia stuartii*	1
Enterobacter cloacae	3	*Pseudomonas aeruginosa*	6
Enterobacter hormaechei	10	*Salmonella enterica*	22
Enterobacter roggenkampii	12	*Staphylococcus aureus*	30
Enterococcus faecium	2	*Staphylococcus lugdunensis*	1
Enterococcus sp.	1		

Technical Validation

The baseline data for the simulations were 100% completed genomes of ESKAPE pathogens, with accompanying FASTQ reads, all of which passed the National Center for Biotechnology Information curation process. The assembly and simulation software used to create benchmark metagenomic data sets have been previously validated in their own publications. As outlined in the Data Processing section, any assemblies or simulated reads not passing quality metrics were excluded.

Usage Notes

Not used.

Code availability

Custom code (hAMRonization v1.0.3) was used to compare different AMR tools to predict AMR genes in genomic data and produce a standard report to compare results across tools (Fig. 2.). This code is available at Github[29].

References

1. Boolchandani, M., D'Souza, A. W. & Dantas, G. Sequencing-based methods and resources to study antimicrobial resistance. Nat Rev Genet 20, 356–370 (2019).
2. Karst, S. M. et al. High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. Nat. Methods 18, 165–169 (2021).
3. Simar, S. R., Hanson, B. M. & Arias, C. A. Techniques in bacterial strain typing: past, present, and future. Curr. Opin. Infect. Dis. 34, 339–345 (2021).
4. Habets, A. et al. Genetic characterization of Shigatoxigenic and enteropathogenic Escherichia coli O80:H2 from diarrhoeic and septicaemic calves and relatedness to human Shigatoxigenic E. coli O80:H2. J. Appl. Microbiol. 130, 258–264 (2021).
5. Cooper, A. L. et al. Systematic evaluation of whole genome sequence-based predictions of Salmonella serotype and antimicrobial resistance. Front. Microbiol. 11, 549 (2020).
6. Dahl, L. G., Joensen, K. G., Osterlund, M. T., Kiil, K. & Nielsen, E. M. Prediction of antimicrobial resistance in clinical Campylobacter jejuni isolates from whole-genome sequencing data. Eur. J. Clin. Microbiol. Infect. Dis. 40, 673–682 (2021).
7. Zhou, H. et al. Clinical impact of metagenomic next-generation sequencing of bronchoalveolar lavage in the diagnosis and management of pneumonia: a multicenter prospective observational study. J. Mol. Diagn. 23, 1259–1268 (2021).
8. Harris, P. N. A. & Alexander, M. W. Beyond the core genome: tracking plasmids in outbreaks of multidrug-resistant bacteria. Clin. Infect. Dis. 72, 421–422 (2021).
9. David, S. et al. Integrated chromosomal and plasmid sequence analyses reveal diverse modes of carbapenemase gene spread among Klebsiella pneumoniae. Proc. Natl. Acad. Sci. USA 117, 25043–25054 (2020).
10. Strepis, N. et al. Genetic analysis of mcr-1-carrying plasmids from gram-negative bacteria in a dutch tertiary care hospital: evidence for intrapatient and interspecies transmission events. Front. Microbiol. 12, 727435 (2021).
11. Mahfouz, N., Ferreira, I., Beisken, S., von Haeseler, A. & Posch, A. E. Large-scale assessment of antimicrobial resistance marker databases for genetic

phenotype prediction: a systematic review. J. Antimicrob. Chemother. 75, 3099–3108 (2020).
12. Doyle, R.M., et al. Discordant bioinformatic predictions of antimicrobial resistance from whole-genome sequencing data of bacterial isolates: an inter-laboratory study. Microb. Genom. 6 (2020).
13. Jaillard, M., Palmieri, M., van Belkum, A. & Mahe, P. Interpreting k-mer-based signatures for antibiotic resistance prediction. Gigascience 9 (2020).
14. Petrillo, M. et al. A roadmap for the generation of benchmarking resources for antimicrobial resistance detection using next generation sequencing [version 1; peer review: 2 approved with reservations]. F1000Research 10, 80 (2021).
15. Chen, Z., Erickson, D.L. & Meng, J.H. Benchmarking hybrid assembly approaches for genomic analyses of bacterial pathogens using Illumina and Oxford Nanopore sequencing. BMC Genomics 21 (2020).
16. JPIAMR Secretariat. Microbial bioinformatics hackathon and workshop - virtual event, 11–15 October 2021. https://www.jpiamr. eu/app/uploads/2021/11/Microbial-Bioinformatics-Hackathon-and-Workshop-2021-report.pdf (2021).
17. National Center for Biotechnology Information (NCBI). Microbial genomes. https://www.ncbi.nlm.nih.gov/genome/microbes/ (2021).
18. Seemann, T. Shovill. Github https://github.com/tseemann/shovill (2020).
19. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A. & Korobeynikov, A. Using SPAdes de novo assembler Curr. Protoc. Bioinform. 70, e102 (2020).
20. Souvorov, A., Agarwala, R. & Lipman, D.J. SKESA: strategic k-mer extension for scrupulous assemblies. Genome Biol. 19 (2018).
21. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075 (2013).
22. Seemann, T. Snippy. Github https://github.com/tseemann/snippy (2020).
23. Quinlan, A. R. BEDTools: the Swiss-army tool for genome feature analysis. Curr. Protoc. Bioinform. 47, 11–12 (2014).
24. Alcock, B. P. et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids, Res. 48, D517–D525 (2020).
25. Di Tommaso, P. et al. Nextflow enables reproducible computational workflows. Nat. Biotechnol. 35, 316–319 (2017).
26. Sczyrba, A. et al. Critical assessment of metagenome interpretation-a benchmark of metagenomics software. Nature Methods 14, 1063–1071 (2017).
27. Huang, W. C., Li, L. P., Myers, J. R. & Marth, G. T. ART: a next-generation sequencing read simulator. Bioinformatics 28, 593–594 (2012).
28. Pysam-developers. Pysam. Vol. 2021.
29. hAMRonization. Public Health Alliance for Genomic Epidemiology (pha4ge). https://github.com/pha4ge/hAMRonization (2020).
30. de Oliveira Martins, L., Jamin, C., Raphenya, A. R. & Maguire, F. AMR-Hackathon-2021/benchmarking_datasets: v1.1.0. Zenodo https://doi.org/10.5281/zenodo.6543963 (2021).
31. Robertson, J., Hays, J. P., Jamin, C., de Oliveira Martins, L. & Raphenya, A. R. AMR Benchmarking dataset - Assemblies. Zenodo https://doi.org/10.5281/zenodo.5604579 (2021).
32. Robertson, J., Hays, J. P., Jamin, C., de Oliveira Martins, L. & Raphenya, A. R. AMR Benchmarking dataset - Mapped ReadSets - 1. Zenodo https://doi.org/10.5281/zenodo.5647909 (2021).
33. Robertson, J., Hays, J. P., Jamin, C., de Oliveira Martins, L. & Raphenya, A. R. AMR Benchmarking dataset - Mapped ReadSets - 2. Zenodo https://doi.org/10.5281/zenodo.5715459 (2021).
34. Robertson, J., Hays, J. P., Jamin, C., de Oliveira Martins, L. & Raphenya, A. R. AMR Benchmarking dataset - Mapped ReadSets - 3. Zenodo https://doi.org/10.5281/zenodo.5718463 (2021).
35. Robertson, J., Hays, J. P., Jamin, C., de Oliveira Martins, L. & Raphenya, A. R. AMR Benchmarking dataset - Mapped ReadSets - 4. Zenodo https://doi.org/10.5281/zenodo.5719315 (2021).
36. Robertson, J., Hays, J. P., Jamin, C., de Oliveira Martins, L. & Raphenya, A. R. AMR Benchmarking dataset - Mapped ReadSets - 5. Zenodo https://doi.org/10.5281/zenodo.5720889 (2021).
37. Robertson, J., Hays, J. P., Jamin, C., de Oliveira Martins, L. & Raphenya, A. R. AMR Benchmarking dataset - Mapped ReadSets - 6. Zenodo https://doi.org/10.5281/zenodo.5725680 (2021).
38. Maguire, F. AMR Benchmarking dataset - Metagenomics. Zenodo https://doi.org/10.5281/zenodo.6543357 (2021).
39. Maguire, F. fmaguire/AMR_Metagenome_Simulator: v1.0.0. Zenodo https://doi.org/10.5281/zenodo.6509951 (2021).

Acknowledgements

This work was made possible and supported by a collaboration between the Public Health Alliance for Genomic Epidemiology (PHA4GE – https://pha4ge.org), the Joint Programming Initiative on Antimicrobial Resistance (JPIAMR – https://www.jpiamr.eu/) and the MRC Cloud Infrastructure for Microbial Bioinformatics (MRC CLIMB-BD – https://tinyurl.com/climb-movie). We would also like to thank Boas van der Putten (University of Amsterdam) for initial contributions to the work performed in this publication.

Author contributions

Data selection: A.R.R., J.R., C.J., L.de.O.M., J.P.H. Data processing: A.R.R., J.R., C.J., L.de.O.M., J.P.H. Manuscript Writing: A.R.R., J.R., L.de.O.M., C.J., J.P.H., F.M. Manuscript Editing: A.R.R., J.R., L.de.O.M., C.J., A.G.M., J.H.E.N., J.P.H., F.M.

Competing interests

The authors declare no competing interests. L.de.O.M was funded by the Quadram Institute Bioscience BBSRC funded Core Capability Grant (project number BB/CCG1860/1). A.R.R and A.G.M. were supported by a grant from the Canadian Institutes of Health Research (PJT-156214) and A.G.M. was additionally supported by a David Braley Chair in Computational Biology. F.M. was funded by the Dalhousie University and Sunnybrook Health Sciences Centre. C.J. was funded by the Care and Public Health Research Institute (CAPHRI), Maastricht University Medical Center. J.P.H was funded by a Network Plus 2020 grant from the Joint Programming Initiative on Antimicrobial Resistance (JPIAMR - Seq. 4AMR - ZonMW 549010001) and the Erasmus University Medical Centre Rotterdam (Erasmus MC). No additional funding was required for the work described in this manuscript

5

MOBILE COLISTIN RESISTANCE GENE 9 IS NOT CLINICALLY RELEVANT

Casper Jamin[1, #], Erik Beuken[1], Christian von Wintersdorff[1], Marjolein F.Q. Kluytmans - van den Bergh2[2, 3, 4], Jan A.J. W. Kluytmans[4], Christian J.P.A. Hoebe[1, 5, 6], Paul H.M. Savelkoul[1, 7], Lieke B. van Alphen[1], on behalf of the i-4-1-Health Study Group

1. Department of Medical Microbiology, Care and Public Health Research Institute (CAPHRI), Maastricht University Medical Center+, Maastricht, The Netherlands.
2. Department of Infection Control, Amphia Hospital, Breda, the Netherlands
3. Amphia Academy Infectious Disease Foundation, Amphia Hospital, Breda, the Netherlands
4. Julius Center for Health Sciences and Primary Care, university Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands
5. Department of Sexual Health, Infectious Diseases and Environment, South Limburg Public Health Service, Heerlen, The Netherlands.
6. Department of Social Medicine, Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, The Netherlands.
7. Department of Medical Microbiology and Infection Control, Amsterdam UMC, location VUmc, Amsterdam, The Netherlands.

Manuscript under preparation

Abstract

Since its discovery in 2015, mobile colistin resistance (MCR) has rapidly spread across the globe. Here, we unraveled the clinical prevalence and relevance of *mcr-9*, reportedly causing inducible colistin resistance. First, we observed *mcr-9* by PCR in 6.3% (21/333) of patients after enriching rectal carriage swabs in tryptone soy broth. Next, we screened 2698 Whole Genome sequencing (WGS) datasets of 3rd generation cephalosporin-resistant or ciprofloxacin-resistant *Enterobacteriaceae* from patients in hospitals, residents of long-term care facilities, children in daycare centers, broilers, and weaned pigs in The Netherlands for *mcr-9. Mcr-9-like* genes were most often found in *Enterobacter* sp. (96/167, 57%) and *Citrobacter sp.* (26/62, 42%). The majority (91%) carried an *mcr-9* like gene, with an extra TGG tryptophan codon before the stop codon. Only one *E. coli* and *Enterobacter* had an exact *mcr-9.1* gene with the *QseBC* two-component system, required for inducible colistin resistance. These two isolates and several *mcr-9*-like isolates showed no inducible colistin resistance. Attempts were made to functionally clone *mcr-9* in *E. coli* DH10β, but no clones expressing *mcr-9* were recovered. Even after several efforts and strategies, only reverse orientation clones were obtained. This observation strongly suggests that the expression of *mcr-9* in *E. coli*, even at very low levels, may exhibit a toxic effect for this host.

Therefore, we argue that, although *mcr-9-like* is quite prevalent, the effect on colistin resistance seems of little consequence, and screening for this gene is not helpful for clinical practice.

Introduction

Treatment options for multidrug-resistant pathogens have been dwindling, and colistin remains one of the few last-resort antibiotics to treat infections with these bacteria. Resistance towards colistin was thought to be restricted to chromosomal mutations in, for instance, the *pmrAB* genes[1]. Since the first report of mobilized colistin resistance (MCR) in 2015[2], up to ten *mcr*-genes have been identified[3–11]. *Mcr-9* was first described in *Salmonella enterica* Serovar *Typhimurium*. Although the authors that first describe it[10,12] claim that *mcr-9* gene leads to inducible colistin resistance, it remains unclear whether this gene is truly associated with colistin resistance, as others have been unable to find a link between *mcr-9* and colistin resistance[13,14]. Here, we investigated the relevance and prevalence of *mcr-9.1* by screening patients for carriage of *mcr-9.1*, screening of isolates obtained from human rectal carriage in hospitals, daycare- and longterm healthcare facilities and at livestock farms.

Material and Methods

Patient screening

For routine hospital screening of multidrug-resistant bacteria at Maastricht university medical centre, perianal swabs are taken from hospitalized patients and are enriched for *Enterococcus* and *Enterobacterales* by culturing these swabs in tryptone soya broth (TSB, Tritium) for 18-24 hours at 35-37°C. After incubation, DNA was extracted on a MagnaPure96 platform (Roche). Quantitative PCR (qPCR) was performed on a QuantStudio 5 Real-Time PCR System (ThermoFisher). For PCR, 1µl eluted DNA, 0.75 µl forward primer, 0.75 µl reverse primer (300 nM), 0.5 µl probe (100nM), 7 µl MilliQ™ and 10µl TaqPath qPCR master Mix (ThermoFisher) was used in the PCR reaction. Primer and probe sequences are available in Table 1. The following PCR cycle protocol was used: activation was done for 2 min. 95 °C, followed by 42 cycles of 15 sec. at 95 °C and 40 sec. at 60 °C. DNA extracts from a selection of 96-wells-plates containing DNA extracts of 333 enriched broths, from 2019 were selected to screen for the *mcr-9* gene.

Whole genome sequencing and resistance gene identification

Isolates from BioProject PRJEB45369 were subjected to whole-genome sequencing as previously described[15]. *Mcr-9* and the two-component system *qseBC* were identified using KMA[16] on default settings. Only hits with more than 99% sequence identity and corresponding length were further investigated by *de novo* assembly using SKESA[17]. The entire workflow and reference sequences are available at Zenodo (v1.0.1, 10.5281/zenodo.4696676). For identification of *mcr-9* in *Enterobacteriaceae*, two strain collections of 1806 and 890 that were subjected to paired-end WGS (SoM: BioProject PRJEB15226 and i-4-1-health: PRJEB45369), which represents a collection of ESBL-producing *Enterobacteriaceae* isolates from patients in Dutch hospitals between 2011 and 2014 (SoM study)[18] and a collection of ESBL-producing, ciprofloxacin or carbapenem resistant *Enterobacteriaceae* isolates from patients in Dutch hospitals, residents in long-term care facilities, children in daycare centers, weaned pigs and broilers in 2017 and 2018 (i-4-1-Health study)[19].

Cloning and expression of *mcr-9* in *E. coli*

Primers used to amplify the *mcr-9* gene (without native promoter) or the complete *mcr-9-QseBC* cassette (with *mcr-9* promoter) are listed in Table 1. After amplification, the resulting *mcr-9* amplicon (2 kb) was ligated into pGEM®-T Easy (Amp^R). Next, the amplicon of the *mcr-9* cassette (5.3 kb) was digested with *NotI* and *PmeI*, purified and ligated into vector pZE21-NP (Kan^R), digested with *NotI* and *PmeI*. The ligation mixes were used to transform *E. coli* DH10β cells. Plating was done on LB-agar plates with

or without 0.5µg/ml colistin. After overnight incubation, only plates without colistin contained colonies.

Plasmid DNA isolation was performed, and clones containing an insert were confirmed by Sanger sequencing, resulting in plasmids p1281 (pGEM-T Easy plus *mcr-9* gene, in the reverse direction to P^{LacZ}) and p1275 (pZE21-NP containing the *mcr-9* cassette). In the cloning experiment of the *mcr-9* gene, no clones with the *mcr-9* gene in the direct direction regarding P^{LacZ} were found. Additionally, sticky overhang cloning strategies (in both pUC119 and pZE21-NP) were addressed for directed cloning of the *mcr-9* gene, without any results.

Table 1. Primer and probes sequences used in this study

qPCR mcr-9 Forward	AAGCCTAGTGATAACCCGAAAC
qPCR mcr-9 reverse	TGATATGGAAGGCGACAAGC
qPCR mcr-9 probe	FAM-AACGTGCCATGACGAGGTGATGCT-BHQ1
MCR9cassF	TACCGGTTTAAACAGCTGTTCGGGGGTTCAC
MCR9cassR	ATCAGCGGCCGCCAGCGCGAATATATCCAGTGG
MCR9genF	TTGTAGATATCATCAATGTTTTTACTGGTTTAC
MCR9genR	GCCTGGATCCATCTCCAGCACTTTACAGTC

Induction of colistin resistance in *mcr-9* carrying isolates

Isolates from BioProject PRJEB45369 with *mcr-9* and *mcr-9*-like were used to test for inducible colistin resistance. To measure inducible colistin resistance, isolates were inoculated (0.5 MacFarland) in microtiter plates with serial diluted colistin (64, 32, 16, 8, 4, 2, 1, 0,5, 0.25, 0,1125, 0 mg/µl). Dilutions in the maximum concentration where growth was still observed in the microtiter plates were used for subculturing into a new microtiter plate to see if growth at higher concentration of colistin was attained. An *E. coli* DH10β with and without *mcr-3* was used as a positive and a negative control, respectively.

Results & discussion

In total, 21 out of 333 patients (6.3%) had detectable levels of *mcr-9* rectal carriage detected by qPCR. So far, no other studies have determined the rectal carriage of *mcr-9* in humans.

Two previously generated WGS datasets[18,19] of antibiotic-resistant *Enterobacteriaceae* were screened to better understand the prevalence of *mcr-9* in these multi-drug-resistant *Enterobacteriaceae* and see how often the required two-component system *qseBC* for inducible colistin resistance is present in conjunction with *mcr-9*.

Of the 2696 WGS datasets, 97/167 *Enterobacter* (58%), 26/62 *Citrobacter sp.* (42%), 5/28 *Klebsiella oxytoca* (18%), 15/357 *Klebsiella pneumoniae* (4.2%) and 23/1992 *E. coli* (1.1%) isolates carried *mcr-9* like genes (more than 99% length and sequence identity, Figure 1). No *mcr-9* was detected in isolates from weaned pigs (n = 28), broilers (n = 203), residents from long-term care facilities (n = 151) or children (n = 60) in daycare centers (I-4-1-health study data, Bioproject PRJEB45369, 442/890, data not shown). However, this may be due to low numbers of *Enterobacter* (n = 2), *Citrobacter* (n = 2) and *Klebsiella* (n = 18) isolates that were retrieved in these populations, which do not seem to carry *mcr-9* frequently. It should be noted that for the I-4-1-health study, only ESBL-producing, carbapenem or ciprofloxacin resistant *E. coli* from broiler and weaned pigs farms were collected for this study. The high frequency of *Enterobacter* sp. carrying *mcr-9*-like and *mcr-9* isolated from patients in hospitals is in line with other work[20]. In total, 152/167 (=91%) isolates have an *mcr-9.1* like gene. This gene has an extra tryptophan codon (TGG), right before the stop-codon (TAA). Five isolates had the *QseBC* two-component system required for inducible colistin resistance according to literature[10,12], of which three also had an exact *mcr-9.1* gene. Only 15 (0.6%) had an exact *mcr-9.1* gene. A single occurrence of *mcr-3* was the only other *mcr* gene present in this dataset (ERR1617955, data not shown). Without an exact *mcr-9* and *QseBC*, it is unlikely that these isolates will be resistant to colistin.

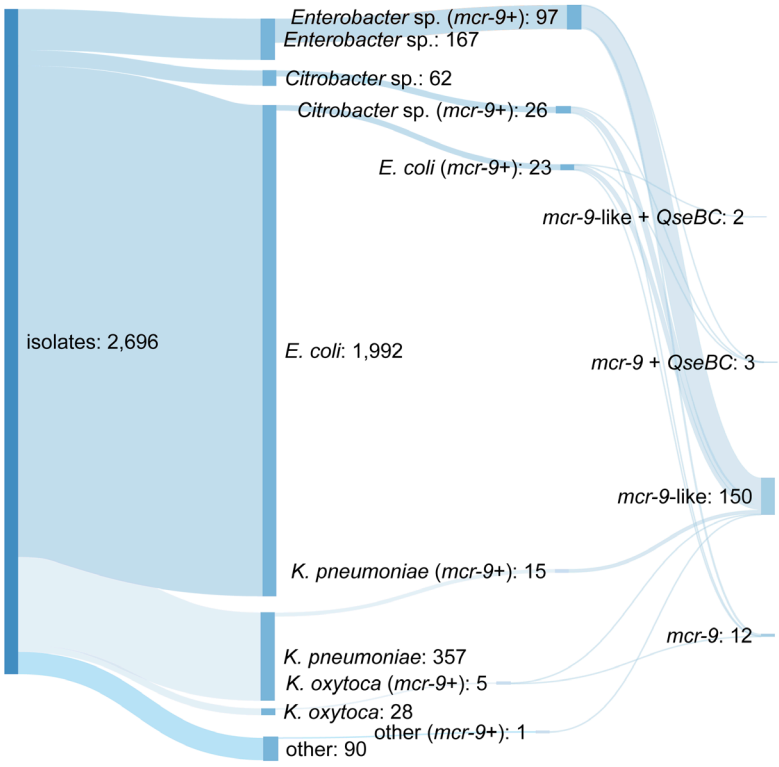


Figure 1. Sankey diagram showing the distribution of mcr-9 carrying *Enterobacteriaceae*, split up for each species. Combinations of mcr-9, mcr-9 like, either with or without QseBC are shown as fluxes coming from each species.

Clinical isolates screening

Two *mcr-9 + qseBC*, two *mcr-9*-like *+ qseBC*, two *mcr-9* and two *mcr-9*-like isolates were screened for inducible colistin resistance. All strains were susceptible to colistin (<2mg/µl), and showed no change in colistin resistance after induction.

Cloning and expression of MCR-9

Attempts were made to clone either *mcr-9* (with or without native *mcr-9* promoter) or the whole *mcr-9 – qseBC* cassette (with or without the *mcr-9* promoter) into *E. coli* plasmids containing different (inducible) promoters. Both blunt-ended (in pGEMT-easy and pUC119, digested with HincII) and sticky overhang cloning strategies (in both pUC119 and pZE21-NP) were addressed for random or directed ligation of the fragments and transformations of *E. coli* K12 DH10β were carried out with or without colistin (0.5 µg/ml). To avoid over-expression, IPTG was omitted in the LB-agar plates (no blue-white screening). No colonies were recovered from the plates containing colistin. Unexpectedly, only blunt-ended ligation strategies resulted in clones, but all the inserts showed a reverse orientation compared to the plasmid promoter (P^{lacZ}). Moreover, after trying to force the ligation of the fragments in the correct orientation compared to the P^{tet} promotor of pZE21-NP, using different sticky-end overhangs, the few clones that were obtained turned out to be aberrant ligation products (e.g. deletions in the fragments or vector).

These observations strongly suggest that the MCR-9 protein is toxic for the *E. coli* host when expressed by E. coli promoters P^{lacZ} or P^{tet}.

Conclusion

This study shows the ubiquitous presence of the newly described *mcr-9* in hospitalized patients gut microbiota. We demonstrated that the *QseBC* two-component system is infrequently encoded along with *mcr-9*, and that, in our study, this did not contribute to colistin resistance. Therefore clinicians, medical microbiologists and other hospital staff should not be alarmed when *mcr-9* carriage is observed and therefore screening for this gene should not be performed.

References

1. Bourrel, A. S. *et al.* Colistin resistance in Parisian inpatient faecal Escherichia coli as the result of two distinct evolutionary pathways. *J. Antimicrob. Chemother.* (2019) doi:10.1093/jac/dkz090.
2. Liu, Y.-Y. *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.* **16**, 161–168 (2017).
3. Xavier, B. B. *et al.* Identification of a novel plasmid-mediated colistinresistance gene, mcr-2, in Escherichia coli, Belgium, june 2016. *Eurosurveillance* (2016) doi:10.2807/1560-7917.ES.2016.21.27.30280.
4. Yin, W. *et al.* Novel plasmid-mediated colistin resistance gene mcr-3 in Escherichia coli. *MBio* (2017) doi:10.1128/mBio.00543-17.
5. Carretto, E. *et al.* Detection of mcr-4 positive Salmonella enterica serovar Typhimurium in clinical isolates of human origin, Italy, october to November 2016. *Eurosurveillance* (2018) doi:10.2807/1560-7917.ES.2018.23.2.17-00821.
6. Borowiak, M. *et al.* Identification of a novel transposon-associated phosphoethanolamine transferase gene, mcr-5, conferring colistin resistance in d-tartrate fermenting Salmonella enterica subsp. enterica serovar Paratyphi B. *J. Antimicrob. Chemother.* (2017) doi:10.1093/jac/dkx327.
7. AbuOun, M. *et al.* mcr-1 and mcr-2 (mcr-6.1) variant genes identified in Moraxella species isolated from pigs in Great Britain from 2014 to 2015. *J. Antimicrob. Chemother.* (2018) doi:10.1093/jac/dky272.
8. Yang, Y. Q., Li, Y. X., Lei, C. W., Zhang, A. Y. & Wang, H. N. Novel plasmid-mediated colistin resistance gene mcr-7.1 in Klebsiella pneumoniae. *J. Antimicrob. Chemother.* (2018) doi:10.1093/jac/dky111.
9. Wang, X. *et al.* Emergence of a novel mobile colistin resistance gene, mcr-8, in NDM-producing Klebsiella pneumoniae article. *Emerg. Microbes Infect.* (2018) doi:10.1038/s41426-018-0124-z.
10. Carroll, L. M. *et al.* Identification of novel mobilized colistin resistance gene mcr-9 in a multidrug-resistant, colistin-susceptible salmonella enterica serotype typhimurium isolate. *MBio* **10**, (2019).
11. Wang, C. *et al.* Identification of novel mobile colistin resistance gene mcr-10. *Emerg. Microbes Infect.* (2020) doi:10.1080/22221751.2020.1732231.
12. Kieffer, N. *et al.* Mcr-9, an Inducible Gene Encoding an Acquired Phosphoethanolamine Transferase in Escherichia coli, and Its Origin. *Antimicrob. Agents Chemother.* (2019) doi:10.1128/AAC.00965-19.
13. Börjesson, S. *et al.* A link between the newly described colistin resistance gene mcr-9 and clinical Enterobacteriaceae isolates carrying blaSHV-12 from horses in Sweden. *J. Glob. Antimicrob. Resist.* **20**, 285–289 (2020).
14. Tyson, G. H. *et al.* The mcr-9 Gene of Salmonella and Escherichia coli Is Not Associated with Colistin Resistance in the United States. doi:10.1128/AAC.00573-20.
15. Jamin, C. *et al.* Harmonization of whole genome sequencing for outbreak surveillance of enterobacteriaceae and enterococci. *bioRxiv* (2020) doi:10.1101/2020.11.20.392399.
16. Clausen, P. T. L. C., Aarestrup, F. M. & Lund, O. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinformatics* (2018) doi:10.1186/s12859-018-2336-6.
17. Souvorov, A., Agarwala, R. & Lipman, D. J. SKESA: Strategic k-mer extension for scrupulous assemblies. *Genome Biol.* (2018) doi:10.1186/s13059-018-1540-z.
18. Kluytmans-Van Den Bergh, M. F. Q. *et al.* Whole-genome multilocus sequence typing of extended-spectrum-beta-lactamase-producing enterobacteriaceae. *J. Clin. Microbiol.* (2016) doi:10.1128/JCM.01648-16.
19. Bergh, M. K. Den *et al.* Microbiological methods to detect intestinal carriage of highly-resistant microorganisms (HRMO) in humans and livestock in the i-4-1-Health Dutch- Belgian cross-border project. *Preprints.org* 1–16 (2019) doi:10.20944/preprints201912.0216.v1.
20. Khedher, M. Ben *et al.* Massive analysis of 64,628 bacterial genomes to decipher water reservoir and origin of mobile colistin resistance genes: is there another role for these enzymes? *Sci. Rep.* **10**, 1–10 (2020).

Acknowledgements

We are grateful to the collaborators in the participating laboratories for their contribution to the collection of whole genome sequence data. These collaborators are part of the SoM Study Group and i-4-1-Health Study Group:

SoM Study group:
Peter van Keulen, Jan Kluytmans, Marjolein Kluytmans-van den Bergh, Kees Verduin, Carlo Verhulst, and Ina Willemsen (Amphia Hospital, Breda, the Netherlands); Els de Brauwer (Atrium MC, Heerlen, the Netherlands); Margreet Vos (Erasmus MC, Rotterdam, the Netherlands); Erika van Elzakker (Haga Hospital, The Hague, the Netherlands); Gijs Ruijs (Isala Clinics, Zwolle, the Netherlands); Guy Oudhuis and Paul Savelkoul (Maastricht UMC, Maastricht, the Netherlands); Bram Diederen (Red Cross Hospital, Beverwijk, the Netherlands); Anton Buiting (St. Elisabeth/TweeSteden Hospital, Tilburg, the Netherlands); Alexander Friedrich, Greetje Kampinga, Erwin Raangs, Sigrid Rosema, and John Rossen (UMC Groningen, Groningen, the Netherlands); Joost Hopman and Andreas Voss (UMC St Radboud, Nijmegen, the Netherlands); Marc Bonten, Martin Bootsma, Patricia Bruijning-Verhagen, Jan Kluytmans, Marjolein Kluytmans-van den Bergh, Annet Troelstra, and Rob Willems (UMC Utrecht, Utrecht, the Netherlands); Christina Vandenbroucke-Grauls (VU University Medical Center, Amsterdam, the Netherlands); Nashwan al Naiemi (Ziekenhuisgroep Twente, Almelo/Hengelo, the Netherlands).

I-4-1-Health Study Group:
Lieke van Alphen (Maastricht University Medical Center+, Maastricht, the Netherlands), Nicole van den Braak (Avans University of Applied Sciences, Breda, the Netherlands), Caroline Broucke (Agency for proofreading ;)Care and Health, Brussels, Belgium), Anton Buiting (Elisabeth-TweeSteden Hospital, Tilburg, the Netherlands), Liselotte Coorevits (Ghent University Hospital, Ghent, Belgium), Sara Dequeker (Agency for Care and Health, Brussels, Belgium and Sciensano, Brussels, Belgium), Jeroen Dewulf (Ghent University, Ghent, Belgium), Wouter Dhaeze (Agency for Care and Health, Brussels, Belgium), Bram Diederen (ZorgSaam Hospital, Terneuzen, the Netherlands), Helen Ewalts (Regional Public Health Service Hart voor Brabant, Tilburg, the Netherlands), Herman Goossens (University of Antwerp, Antwerpen, Belgium and Antwerp University Hospital, Antwerp, Belgium), Inge Gyssens (Hasselt University, Hasselt, Belgium), Casper den Heijer (Regional Public Health Service Zuid- Limburg, Heerlen, the Netherlands), Christian Hoebe (Maastricht University Medical Center+, Maastricht, the Netherlands and Regional Public Health Service Zuid-Limburg, Heerlen, the Netherlands), Casper Jamin (Maastricht University Medical Center+, Maastricht, the Netherlands), Patricia Jansingh (Regional

Public Health Service Limburg Noord, Venlo, the Netherlands), Jan Kluytmans (Amphia Hospital, Breda, the Netherlands and University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands), Marjolein Kluytmans-van den Bergh (Amphia Hospital, Breda, the Netherlands and University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands), Stefanie van Koeveringe (Antwerp University Hospital, Antwerp, Belgium), Sien De Koster (University of Antwerp, Antwerp, Belgium), Christine Lammens (University of Antwerp, Antwerp, Belgium), Isabel Leroux-Roels (Ghent University Hospital, Ghent, Belgium), Hanna Masson (Agency for Care and Health, Brussel, Belgium), Ellen Nieuwkoop (Elisabeth-TweeSteden Hospital, Tilburg, the Netherlands), Anita van Oosten (Admiraal De Ruyter Hospital, Goes, the Netherlands), Natascha Perales Selva (Antwerp University Hospital, Antwerp, Belgium), Merel Postma (Ghent University, Ghent, Belgium), Stijn Raven (Regional Public Health Service West-Brabant, Breda, the Netherlands), Veroniek Saegeman (University Hospitals Leuven, Leuven, Belgium), Paul Savelkoul (Maastricht University Medical Center+, Maastricht, the Netherlands), Annette Schuermans (University Hospitals Leuven, Leuven, Belgium), Nathalie Sleeckx (Experimental Poultry Centre, Geel, Belgium), Arjan Stegeman (Utrecht University, Utrecht, the Netherlands), Tijs Tobias (Utrecht University, Utrecht, the Netherlands), Paulien Tolsma (Regional Public Health Service Brabant Zuid-Oost, Eindhoven, the Netherlands), Jacobien Veenemans (Admiraal De Ruyter Hospital, Goes, the Netherlands), Dewi van der Vegt (PAMM Laboratory for Pathology and Medical Microbiology, Veldhoven, the Netherlands), Martine Verelst (University Hospitals Leuven, Leuven, Belgium), Carlo Verhulst (Amphia Hospital, Breda, the Netherlands), Pascal De Waegemaeker (Ghent University Hospital, Ghent, Belgium), Veronica Weterings (Amphia Hospital, Breda, the Netherlands), Clementine Wijkmans (Regional Public Health Service Hart voor Brabant, Tilburg, the Netherlands), Patricia Willemse-Smits (Elkerliek Hospital, Helmond, the Netherlands), Ina Willemsen (Amphia Hospital, Breda, the Netherlands).

Funding

The i-4-1-Health project was financed by the Interreg V Flanders-The Netherlands program, the cross-border cooperation program with financial support from the European Regional Development Fund (ERDF). Additional financial support was received from the Dutch Ministry of Health, Welfare and Sport, the Dutch Ministry of Economic Affairs, the Province of Noord-Brabant, the Belgian Department of Agriculture and Fisheries, the Province of Antwerp and the Province of East-Flanders." The authors are free to publish the results from the project without interference from the funding bodies. Selective and non-selective agar plates, ETEST® strips and VITEK® 2 AST cards were provided by bioMérieux (Marcy l'Etoile); FecalSwabs® and tryptic soy broths were provided by Copan Italy (Brescia, Italy). The authors are free to publish the results from the project without interference from bioMérieux or Copan Italy.

The SoM study was financed by the Netherlands Organisation for Health Research and Development (ZonMw, project 205100010).

Conflicts of interest

The author(s) declare that there are no conflicts of interest

6

GENETIC ANALYSIS OF PLASMID-ENCODED *MCR-1* RESISTANCE IN *ENTEROBACTERIACEAE* DERIVED FROM POULTRY MEAT IN THE NETHERLANDS

Casper Jamin[1], Bazante K. Sanders[2,3], Miaomiao Zhou[2], Adalberto Costessi[4], Danny Duijsings[4], Jan A.J.W. Kluytmans[5,6], Lieke B. van Alphen[1] and Eefje J.A. Schrauwen[2,3]*

1. Department of Medical Microbiology, Care and Public Health Research Institute (CAPHRI), Maastricht University Medical Center, Maastricht, The Netherlands;
2. Academy for Technology of Health and Environment, Avans University of Applied Science, Breda, The Netherlands;
3. Research Group Analysis Techniques in Life Sciences, Avans University of Applied Sciences, Breda, The Netherlands;
4. BaseClear B.V., Leiden, The Netherlands;
5. Department of Infection Control, Amphia Hospital, Breda, The Netherlands;
6. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

Published:
Jamin, C., Sanders, B. K., Zhou, M., Costessi, A., Duijsings, D., Kluytmans, J. A., ... & Schrauwen, E. J. (2021). Genetic analysis of plasmid-encoded *mcr-1* resistance in *Enterobacteriaceae* derived from poultry meat in the Netherlands. JAC-antimicrobial resistance, 3(4), dlab156.

Abstract

Background

Colistin is classified as the highest priority and critically important antimicrobial for human medicine by WHO as it is the last resort agent for treatment of carbapenem-resistant *Enterobacteriaceae* in humans. Additional research is necessary to elucidate the genetic structure of *mcr-1* resistance genes, commonly found on plasmids, using WGS.

Objectives: To map and compare the genetic characteristics of 35 *mcr-1*-mediated colistin-resistant *Enterobacteriaceae* isolated from chicken meat to highlight the genetic variation of the *mcr-1*-containing plasmids.

Methods

Sequencing was performed using Illumina HiSeq2500, Novaseq6000 and ONT's GridION. GridION data was locally basecalled and demultiplexed using ONT's Albacore 2.3.4 followed by Porechop 2.3. Quality filtering was performed using Filtlong 2.0. Hybrid Assembly was performed using Unicycler 4.7. Plasmids were compared with reference sequences in plasmid-RefSeq and pATLAS.

Results

A total of 35 *mcr-1* positive *Enterobacteriaceae* were investigated, which resulted in 34 qualitatively robust hybrid assemblies of 2 *Klebsiella pneumoniae* and 32 *Escherichia coli*. *mcr-1.1* was present in 33/34 isolates. One isolate contained an *mcr-1.1*-like resistance gene, due to a deletion of one codon. Two *mcr-1.1* genes were located on the chromosome, while the majority of the *mcr-1* genes were found on IncX4 type plasmids (n = 19). Almost all plasmids identified in this study were highly similar to plasmids found in human-derived strains.

Conclusions

The *mcr-1.1*-containing plasmids from retail chicken show high sequence similarity to human *mcr-1.1* plasmids, suggesting that this may be a contributor to the presence of colistin resistance in humans.

Introduction

In 2015, a plasmid-mediated colistin resistance gene was reported in China.[1] From that moment on, many more mobile colistin resistance (*mcr*) genes and variants have been detected all over the globe.[2] This discovery represents a mechanism for an easy transferable resistance mechanism to colistin, which is seen as a last-resort antibiotic to treat carbapenem-resistant *Enterobacteriaceae*.[3] In Europe, colistin is used to treat infections caused by *Enterobacteriaceae* in sheep, cows, pigs, goats and chicken.[4] Therefore, the detection of *mcr-1*-harbouring *Enterobacteriaceae* isolates in chicken meat was self-evident.[5,6] In order to understand the molecular epidemiology and resistance mechanism of *mcr* genes, WGS approaches should be used. Characteristically, high-throughput sequencing platforms (e.g. Illumina) are used in order to sequence the full bacterial genome.[7] However, short reads from these high-throughput sequencers can make it challenging to reconstruct plasmids and therefore they are inaccurate for studying antibiotic resistance epidemiology.[8] Single-molecule sequencing platforms such as the Oxford Nanopore Technologies (ONT) MinION, GridION and PromethION are able to sequence long fragments of DNA. Subsequently, with the use of a hybrid assembly, increased information content can be generated since the genome completeness is increased and the location of resistance genes in the genome can be determined.[9] In this study, short- and long-read sequencing platforms were used in order to study the *mcr-1*-containing *Enterobacteriaceae* isolated from retail chicken meat.[5,6] We used a hybrid-assembly approach to extract the plasmid sequences that contain *mcr-1* and studied the plasmid relationship compared with publicly available *mcr-1* plasmid sequences.

Methods

Sample collection

In total, 35 confirmed *mcr-1*-holding *Enterobacteriaceae* were subjected to Illumina short read and ONT sequencing. The isolates derived from previous studies,[5,6] with the exception of EC-MCR34. All samples derived from three prevalence surveys in Dutch retail chicken meat performed in 2009, 2014 and 2015, which were initially performed to study the presence of ESBL- producing *Enterobacteriaceae*.[10,11] The isolates in this study were genotypically *mcr-1* PCR positive and phenotypical colistin resistant.[5]

Illumina sequencing

The 35 samples were sequenced using paired-end Illumina HiSeq2500. The library prep for 35 samples was performed using the Nextera XT DNA library prep kit and

the Nextera XT Index Kit v2 (Illumina, Eindhoven, The Netherlands), according to the manufacturer's instructions. Libraries were subsequently purified using Agencourt AMPure XP beads (Beckman Coulter, Woerden, The Netherlands) and quantified using the Quant-iT dsDNA HS-kit (Thermo Fisher, Bleiswijk, The Netherlands) and using a Fragment Analyzer (Agilent, Amstelveen, The Netherlands.) Samples were then loaded on a HiSeq2500 system and run for 251 cycles (PE125) using HiSeq Rapid SBS Kit v2 chemistry. Due to low quality, EC-MCR10 and EC-MCR21 were re-sequenced using the Illumina NovaSeq 6000. The library prep for these two samples was performed using the Nextera XT DNA library prep kit and the IDT for Illumina Nextera DNA Unique Dual Indexes (Illumina), according to the manufacturer's instructions. Libraries were subsequently purified using Agencourt AMPure XP beads (Beckman Coulter) and quantified using the Quant-iT dsDNA HS-kit (Thermo Fisher) and using a Fragment Analyzer (Agilent). Samples were then loaded on an S1 flow cell on the NovaSeq6000 system and run for 301 cycles (PE150). Fastq read sequence files were generated using bcl2fastq2 version 2.18. Initial quality assessment was based on data passing the Illumina Chastity filtering. Subsequently, reads containing PhiX control signal were removed using an in-house filtering protocol. In addition, reads containing (partial) adapters were clipped (up to a minimum read length of 50 bp). The second quality assessment was based on the remaining reads using the FASTQC quality control tool version 0.11.5.

ONT sequencing

All 35 samples were sequenced using the ONT GridION (Oxford Nanopore Technologies, Oxford, UK). Libraries were prepared using shearing by needle shearing (KP-MCR01–02 and EC-MCR03–31) or using the Covaris G-tube (EC-MCR32–35). The library was prepared using the ONT 1D ligation sequencing kit (SQK-LSK109) with the native barcoding kit (EXP-NBD103). Samples KP-MCR01–02 and EC-MCR03–29 were loaded on FLO-MIN107 R9.5.1 flow cells and the remaining on a FLO-MIN106 R9.4.1 flow cell.

Sequence data availability

All data is available from the National Center for Biotechnology Information (NCBI) under BioProject number PRJEB44175. Raw short-read Illumina and long-read ONT sequencing data and metadata for all 35 isolates used in this study are available from the NCBI Sequence Read Archive database under accession numbers ERR5727763 to ERR5727797 (short read) and ERR5726838 to ERR5726872 (long read).

Assembly

GridION data were locally basecalled and demultiplexed using ONT's Albacore 2.3.4 followed by Porechop 2.3 to demultiplex the unclassified reads. Quality filtering was

performed using Filtlong 2.0 using the following settings: (i) maximum size of 500 Mbp; (ii) keep 90% percentage of the best reads of the data; and (iii) minimum size of 1000 bp. The long-read quality was evaluated using FastQC and NanoPlot v1.13.0 and the short-read qual- ity using FastQC. Hybrid assembly was performed using Unicycler 4.7 using default settings and a minimum length of 1000bp and subsequently assessed using QUAST 5.0.12 Genetic characterization of the hybrid assemblies was performed using the online service of goseqit.com. The coverage of the ONT sequence reads was calculated by mapping the long reads back to the assembly using minimap2 (v2.13) and SAMtools (v1.9) using the in-house scripts. Sequence annotation was done using Bakta (v1.1).[13]

Plasmid analysis

The *mcr-1* plasmid sequences were manually identified and extracted from the assembly graphs (.gfa files) using Bandage.[14] The *mcr-1* gene sequence (AKF16168.1) was used to locate the *mcr-1* -containing plasmids. *mcr-1* gene-containing plasmids from RefSeq plasmid database and pATLAS (accessed April 8, 2020) were retrieved.[15] Any duplicates entries were removed prior subsequent analyses. In total 69 publicly available plasmids and *mcr-1*-containing plasmids from this study were used. Plasmid sequences were clustered using Plasmidsimilarity (v0.3.0, https://github.com/Casperjamin/Plasmidsimilarity). In short, dissimilarity among plasmids was calculated using the Jaccard index, using the complete k-mer composition (all subsequences in a sequence of length k) of each plasmid sequence, using k length of 31bp. Antimicrobial resistance (AMR) genes, virulence genes and plasmid origin of replications were identified with Abricate (v1.0.1, default settings) using the NCBI, virulence factor database and PlasmidFinder database respectively (retrieved on 10 September 2019).[16,17]

Results and discussion

Table 1. Overview of *mcr-1*-positive isolates with corresponding Inc type, size and other genetic characteristics. (a) Identity or alignment length is not 100%. (b) Substitution in second base pair of first starting codon. (c) Resistance gene detected twice.

Sample	Species	ST	Mcr type	Inc type on *mcr1.1* plasmid	Other AMR genes	Transposase gene located near *mcr1.1*	Contig no.	Contig size	Study reference
KP-MCR01	*K. pneumoniae*	ST107	1.1	IncX4	—	—	6	33 303	5
KP-MCR02	*K. pneumoniae*	ST1944	1.1	IncHI2, IncHI2A[a]	*aph(3')-Ia*[a], *sul3*, *aadA1*[a], *dfrA12*	IS*30*-like element ISApl1 family transposase	2	211 949	5
EC-MCR03	*E. coli*	ST10	1.1	IncX4	—	—	7	33 303	5
EC-MCR04	*E. coli*	ST8262	1.1[b]	IncX4	—	—	5	33 310	5
EC-MCR05	*E. coli*	ST8262	1.1	IncX4	—	—	6	33 310	5
EC-MCR06	*E. coli*	ST1564	1.1	IncX4	—	—	6	33 303	5
EC-MCR07	*E. coli*	ST752	1.1	IncB/O/K/Z	*sul2*	IS*30*-like element ISApl1 family transposase	5	93 122	5
EC-MCR08	*E. coli*	ST10	1.1	IncX4	—	—	5	23 832	5
EC-MCR09	*E. coli*	ST162	1.1	IncX4	—	—	5	35 016	5
EC-MCR11	*E. coli*	ST1842	1.1	IncX4	—	—	3	33 310	5
EC-MCR12	*E. coli*	ST10	1.1	IncX4	—	—	6	33 303	5
EC-MCR13	*E. coli*	ST641	1.1	IncX4	—	—	7	33 303	5
EC-MCR14	*E. coli*	ST155	1.1	IncHI2, IncHI2A[a]	*aadA2*, *cmlA1*[a], *aadA1*[a], *sul3*	IS*30*-like element ISApl1 family transposase	2	243 755	5
EC-MCR15	*E. coli*	ST10	1.1	IncX4	—	—	4	34 755	5
EC-MCR16	*E. coli*	ST997	1.1	IncHI2, IncHI2A[a]	*tet*(A), *sul1*, *aadA1*[a], *dfrA1*[a], *aph(6)-Id*, *aph(3'')-Ib*[a]	IS*30*-like element ISApl1 family transposase	2	214 156	5
EC-MCR17	*E. coli*	ST57	1.1	IncHI2, IncHI2A[a]	*aadA1*[a], *sul3*, *aph(3')-Ia*[a]	IS*30*-like element ISApl1 family transposase	2	211 552	5
EC-MCR18	*E. coli*	ST997	1.1	IncX4	—	—	5	33 310	5
EC-MCR19	*E. coli*	ST997	1.1	IncX4	—	—	5	33 310	5
EC-MCR20	*E. coli*	ST624	1.1	IncX4	—	—	5	33 310	5
EC-MCR21	*E. coli*	ST624	1.1	IncX4	—	—	6	33 310	5
EC-MCR22	*E. coli*	ST10	1.1	IncHI2, IncHI2A[a]	bla_{TEM-1}, *tet*(A), *sul1*, *aadA1*[a], *dfrA1*, *lnu*(F), *aph(3')-Ia*	IS*30*-like element ISApl1 family transposase	2	234 218	5
EC-MCR23	*E. coli*	ST93	1.1	none	—	IS*30*-like element ISApl1 family transposase	1	chromosomal	5
EC-MCR24	*E. coli*	ST48	1.1	IncX4	—	—	6	34 639	5
EC-MCR25	*E. coli*	ST624	1.1	IncX4	—	—	3	33 310	5
EC-MCR26	*E. coli*	ST997	1.1	IncHI2, IncHI2A[a], IncQI	*tet*(A), *sul1*, *aadA1*[a], *dfrA1*[a], *aph(6)-Id*[c], *aph(3'')-Ib*[a,c], *sul2*[a], *aph(3')-Ia*, *aac(3)-IIe*, $bla_{TEM-150}$[a]	IS*30*-like element ISApl1 family transposase	2	267 214	5
EC-MCR27	*E. coli*	ST1011	1.1	IncX4	—	—	8	33 310	5
EC-MCR28	*E. coli*	ST354	1.1	IncHI2, IncHI2A[a], IncQI, Col(MG828)	*tet*(A)[a], *sul1*, *aadA1*[a,c], *dfrA1*[a], *aph(6)-Id*[c], *aph(3'')-Ib*[a,c], *sul2*[a], *aadA2*, *cmlA1*[a], *sul3*, *aac(3)-IIe*, $bla_{TEM-150}$[a]	IS*30*-like element ISApl1 family transposase	3	252 468	5
EC-MCR29	*E. coli*	ST624	1.1	IncHI2, IncHI2A[a]	*cmlA1*, *aadA1*[a,c], *sul3*, *aph(3')-Ia*, bla_{TEM-1}[a], *tet*(A), *aadA2*, *aac(3)-VIa*	IS*30*-like element ISApl1 family transposase	2	261 285	5
EC-MCR30	*E. coli*	ST624	1.1	IncHI2, IncHI2A[a]	*cmlA1*, *aadA1*[c], *sul3*, *aph(3')-Ia*, bla_{TEM-1}, *tet*(A), *aadA2*, *aac(3)-VIa*	IS*30*-like element ISApl1 family transposase	2	261 102	5
EC-MCR31	*E. coli*	ST624	1.1	IncHI2, IncHI2A[a]	*aadA2*, *cmlA1*, *aadA1*[c], *sul3*, *aph(3')-Ia*, bla_{TEM-1}, *tet*(A), *aac(3)-VIa*	IS*30*-like element ISApl1 family transposase	2	260 457	5
EC-MCR32	*E. coli*	ST624	1.1	IncHI2, IncHI2A[a]	*cmlA1*[a], *aadA1*[a,c], *sul3*, *aph(3')-Ia*, bla_{TEM-1}[a], *tet*(A), *aadA2*, *aac(3)-VIa*	IS*30*-like element ISApl1 family transposase	2	261 285	5
EC-MCR33	*E. coli*	ST1564	1.1	IncX4	—	—	4	33 303	5
EC-MCR34	*E. coli*	ST117	1.1	none	—	IS*30*-like element ISApl1 family transposase	1	chromosomal	
EC-MCR35	*E. coli*	ST2079	1.1	IncHI2, IncHI2A[a]	*tet*(A), *sul1*, *aadA1*[a,c], *dfrA1*[a], *aph(6)-Id*, *aph(3'')-Ib*[a], *sul3*, *cmlA1*[a], *aadA2*, *catA1*[a]	IS*30*-like element ISApl1 family transposase	2	248 481	6

[a]Identity or alignment length is not 100%.
[b]Substitution in second base pair of first starting codon.
[c]Resistance gene detected twice.

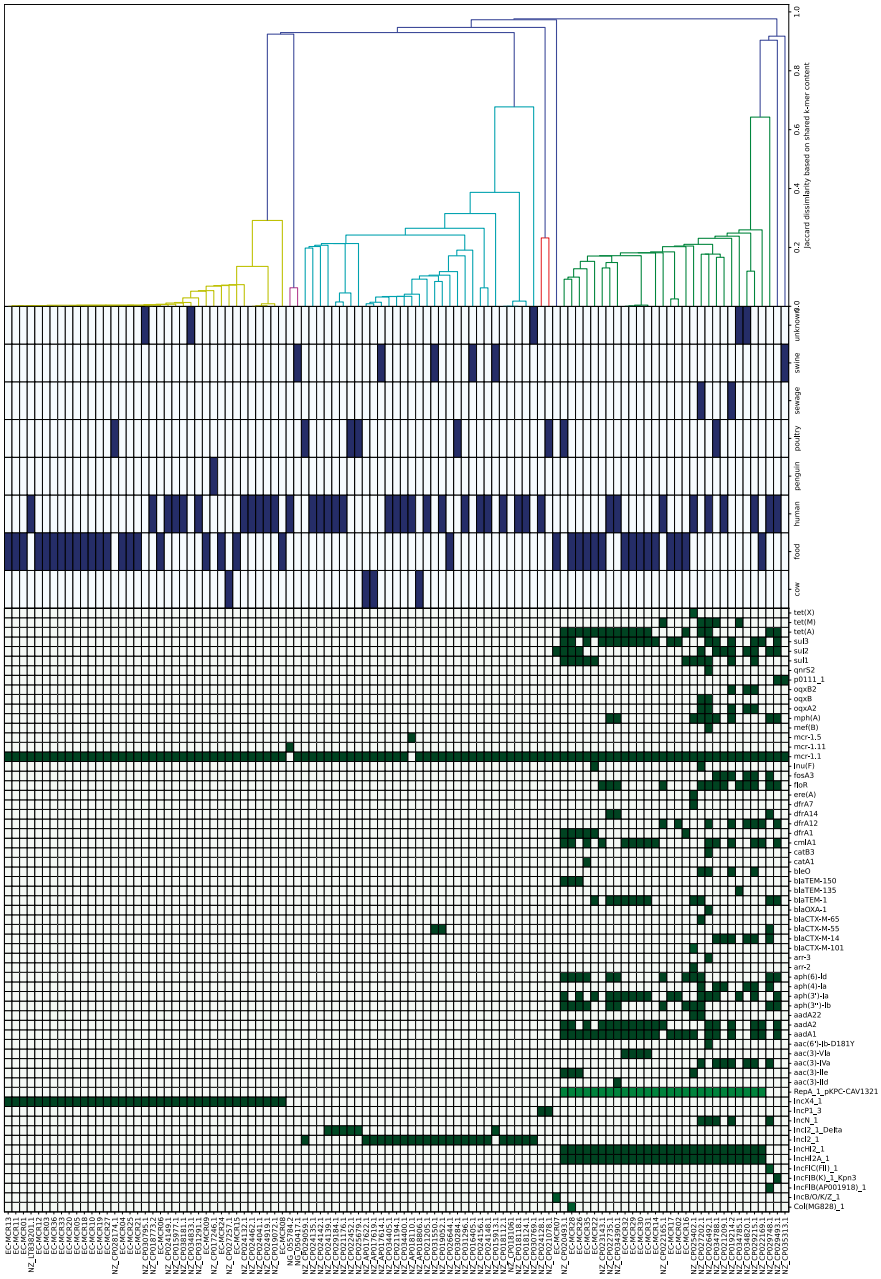


Figure 1. Heatmap and dendrogram showing all plasmids analysed in this study. The dendrogram represents the similarity among plasmid sequences based on the Jaccard dissimilarity index of 31-mers of each plasmid. Coloured cells in the heatmap indicate either the presence of this gene or the origin of replication of this plasmid.

A total of 35 *mcr-1*-positive *Enterobacteriaceae* were investigated, which resulted in 34 qualitatively robust hybrid assemblies of 2 *Klebsiella pneumoniae* and 32 *Escherichia coli* isolates (Table S1, available as Supplementary data at JAC-AMR Online). The hybrid assembly substantially improved the reconstruction of the microbial genome (data not shown). The *mcr-1.1* gene was present in 33/34 isolates (Table 1). The most common STs for E. coli were ST624 (n = 7), ST10 (n = 5) and ST997 (n = 4). The two *K. pneumoniae* isolates belonged to ST107 and ST1944. One isolate contained an *mcr-1.1*-like resistance gene, due to a mutation in the start codon, but still remained resistant to colistin.[5] The second codon in *mcr-1.1* is ATG and will likely replace the first codon as start codon, leading to a truncated but functional gene. Two *mcr-1.1* genes were located on the chromosome, while the majority of the *mcr-1* genes were found on IncX4 type plasmids (n = 19, Table 1), which is a common plasmid type harbouring *mcr-1* found in Europe.[18,19]

Except for the IncX4, *mcr-1.1* plasmids, all *mcr-1.1* genes, plasmid or chromosomal, were flanked by IS30 transposases (Table 1). All the IncX4 *mcr-1* plasmids shared, on average, 0.93 (standard deviation 0.08) of their k-mer content and did not contain any additional resistance genes. As a result, these plasmids were highly similar in size (average 33 kb, range 23 kb to 35 kb, Table 1). Additionally, the IncX4 plasmids found in this study were also highly similar to plasmids present in public databases, which originated from clinical isolates (Figure 1). Furthermore, the bacterial hosts of these IncX4 plasmids showed various STs (Table 1), indicating the widespread nature of this plasmid, most likely driven by conjugation. All IncX4 *mcr-1.1*-containing plasmids carried a *virB* type IV secretion system, required for conjugation (data not shown). The AWGS0007 *mcr-1.1* plasmid (IncB/O/K/Z) encoded an IncI-1-type conjugal transfer protein *trbA*. All other plasmids carried specific incompatibility group-associated conjugation machinery (data not shown). The IncHI2 /IncHIA2 plasmids showed high k-mer similarity among each other (mean 0.72, standard deviation 0.12), but less than the IncX4 plasmids. These were generally much larger in size, ranging from 151 kb to 267 kb and additionally encoded a heterogenous set of AMR genes. It should be noted that no plasmids with an IncI origin of replication (ORI) containing *mcr-1.1* were encountered in the strains in this study (Table 1, Figure 1) and only one IncI plasmid outside this study (NZ.CP02554.1) was derived from food origin. None of the *mcr-1* carrying plasmids in our study carried other genes encoding ESBL resistance. Three isolates contained resistance genes ($bla_{\text{TEM-52c}}$, $bla_{\text{SHV-12}}$ and $bla_{\text{CTX-M-1}}$, respectively), however, these resistance genes were not present on the *mcr-1.1* plasmid. Additionally, one *mcr-1.1* plasmid encoded virulence factors, as it contained five genes of the aerobactin gene cluster (NZ.CP029748.1).

One novel *mcr-1*-containing plasmid was found (EC-MCR07) with a size of 93 kb, which also encoded the sulphonamide resistance gene *sul2*. This plasmid shared barely any sequence similarity as, on average, only a fraction of 0.02 (standard devi-

ation 0.069) of all k-mers were shared with the other plasmids. This was the only plasmid with an IncB/O/K/Z ORI.

The two strains with a chromosomal *mcr-1.1* gene (EC-MCR23 and EC-MCR34) had no other known resistance genes within the same chromosomal region (within 50 kb, data not shown), indicating the mobilization of colistin resistance as a sole passenger of its mobile genetic element IS30. We observed multiple different isolates from retail meat with similar plasmids, which might be caused by the spread of these plasmids within the farms or by individual introduction since these are common plasmids. In addition, similar isolates with identical plasmids are found, which could indicate a batch effect.

Conclusions

In this study we aimed to elucidate the plasmid backbones from *mcr-1*-containing plasmids obtained from retail chicken in the Netherlands. In the strains collected here, *mcr-1* resided often but not always in various plasmids, indicating the high mobility nature of this gene in *E. coli* as a host. Most plasmid backbones found in this study were also found in human clinical isolates. This indicates the possibility of retail meat to be a significant contributor to the dissemination of mobile colistin resistance in the Netherlands.

Acknowledgements

We thank Carlo Verhulst for technical assistance and Pepijn Huizinga and Marjolein Kluytmans-van den Bergh for sample collection.

Funding

This work was funded by ZonMw for Enabling Technologies Hotels (Funding: 40-43500-98-15). Development of Plasmidsimilarity was made possible by a grant from the Dutch workgroup for molecular diagnostics of infectious diseases (WMDI).

References

1. Liu YY, 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. Lancet Infect Dis 2016; 16: 161–8.
2. Sun J, Zhang H, Liu YH et al. Towards understanding MCR-like colistin resistance. Trends Microbiol 2018; 26: 794–808.
3. Poulakou G, Bassetti M, Righi E et al. Current and future treatment options for infections caused by multidrug-resistant Gram-negative pathogens. Future Microbiol 2014; 9: 1053–69.
4. Catry B, Cavaleri M, Baptiste K et al. Use of colistin-containing products within the European Union and European Economic Area (EU/EEA): development of resistance in animals and possible impact on human and animal health. Int J Antimicrob Agents 2015; 46: 297–306.
5. Schrauwen EJA, Huizinga P, van Spreuwel N et al. High prevalence of the mcr-1 gene in retail chicken meat in the Netherlands in 2015. Antimicrob Resist Infect Control 2017; 6: 83.
6. Kluytmans-van den Bergh MF, Huizinga P, Bonten MJ et al. Presence of mcr-1-positive Enterobacteriaceae in retail chicken meat but not in humans in the Netherlands since 2009. Euro Surveill 2016; 21: pii=30149.
7. Perez-Losada M, Arenas M, Castro-Nallar E. Microbial sequence typing in the genomic era. Infect Genet Evol 2018; 63: 346–59.
8. Arredondo-Alonso S, Willems RJ, van Schaik W et al. On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. Microb Genom 2017; 3: e000128.
9. George S, Pankhurst L, Hubbard A et al. Resolving plasmid structures in Enterobacteriaceae using the MinION nanopore sequencer: assessment of MinION and MinION/Illumina hybrid data assembly approaches. Microb Genom 2017; 3: e000118.
10. Kluytmans JA, Overdevest IT, Willemsen I et al. Extended-spectrum β-lactamase-producing Escherichia coli from retail chicken meat and humans: comparison of strains, plasmids, resistance genes, and virulence factors. Clin Infect Dis 2013; 56: 478–87.
11. Overdevest I, Willemsen I, Rijnsburger M et al. Extended-spectrum β-lactamase genes of Escherichia coli in chicken meat and humans, the Netherlands. Emerg Infect Dis 2011; 17: 1216–22.
12. Wick RR, Judd LM, Gorrie CL et al. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 2017; 13: e1005595.
13. Swengers O, Jelonk L, Dieckmann M et al. Bakta: rapid & standardized annotation of bacterial genomes via alignment-free sequence identification. bioRxiv 2021; doi:10.1101/2021.09.02.458689
14. Wick RR, Schultz MB, Zobel J et al. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 2015; 31: 3350–2.
15. Brooks L, Kaze M, Sistrom M. A curated, comprehensive database of plasmid sequences. Microbiol Resour Announc 2019; 8: e01325-18.
16. Carattoli A, Hasman H. PlasmidFinder and in silico pMLST: identification and typing of plasmid replicons in whole-genome sequencing (WGS). Methods Mol Biol 2020; 2075: 285–94.
17. Chen L, Zheng D, Liu B et al. VFDB 2016: hierarchical and refined dataset for big data analysis–10 years on. Nucleic Acids Res 2016; 44: D694–7.
18. Matamoros S, van Hattem JM, Arcilla MS et al. Global phylogenetic analysis of Escherichia coli and plasmids carrying the mcr-1 gene indicates bacterial diversity but plasmid restriction. Sci Rep 2017; 7: 15364.
19. Schuele L, Fleres G, Strutzberg-Minder K et al. Detection of a small IncX4 plasmid carrying the mcr-1.1 gene in a pig oral fluid sample by shotgun metagenomic sequencing. J Glob Antimicrob Resist 2021; 24: 205–6.

Supplementary data

Table S1. Quality of assemblies made by Unicycler

Sample	Coverage long read[A]	#Contigs	Total #Bases	Max. contig size	N50	%GC[C]	%Expected Genome size[D]
KP-MCR01	50	6	5625920	5183587	5183587	56.89	100,53
KP-MCR02	53	2	5553141	5341192	5341192	56.96	99,23
EC-MCR03	55	9	5479521	3411178	3411178	50.38	106,56
EC-MCR04	95	9	5000708	4645108	4645108	50.64	97,25
EC-MCR05	60	9	5003661	4643817	4643817	50.58	97,31
EC-MCR06	65	8	5348917	4973288	4973288	50.80	104,02
EC-MCR07	48	12	5650364	5165739	5165739	50.51	109,88
EC-MCR08	49	14	5424293	4949435	4949435	50.22	105,49
EC-MCR09	82	5	5237772	3373312	3373312	50.56	101,86
EC-MCR10[E]	36	108	7209457	1395950	698469	50.33	140,20
EC-MCR11	67	6	5381953	5152918	5152918	50.43	104,66
EC-MCR12	94	12	5138921	4760363	4760363	50.69	99,94
EC-MCR13	91	8	5279382	4790097	4790097	50.41	102,67
EC-MCR14	92	7	5227795	4626779	4626779	50.37	101,67
EC-MCR15	81	7	5077637	4777520	4777520	50.46	98,75
EC-MCR16	87	6	5535540	5213285	5213285	50.19	107,65
EC-MCR17	59	12	6059500	5349456	5349456	50.38	117,84
EC-MCR18	54	6	5298046	4997031	4997031	50.27	103,03
EC-MCR19	58	7	5292726	4996226	4996226	50.27	102,93
EC-MCR20	50	8	5464992	5193752	5193752	50.37	106,28
EC-MCR21	87	14	5460321	2897977	2897977	50.41	106,19
EC-MCR22	39	6	5570856	5058493	5058493	50.19	108,34
EC-MCR23	57	6	5299674	4890562	4890562	50.63	103,06
EC-MCR24	31	23	5503906	3448470	3448470	50.75	107,04
EC-MCR25	55	5	5385637	5213056	5213056	50.43	104,74
EC-MCR26	42	9	5612654	5230921	5230921	50.17	109,15
EC-MCR27	4	19	5299588	2109155	1141562	50.43	103,06
EC-MCR28	50	12	5623601	3173669	3173669	50.25	109,36
EC-MCR29	29	6	5633347	5221882	5221882	50.31	109,55
EC-MCR30	86	6	5634232	5222950	5222950	50.31	109,57
EC-MCR31	87	6	5637198	5226561	5226561	50.30	109,63
EC-MCR32	87	6	5633335	5221870	5221870	50.31	109,55
EC-MCR33	91	12	5343966	4877641	4877641	50.86	103,93
EC-MCR34	91	10	5375857	5111885	5111885	50.71	104,55
EC-MCR35	86	5	5570297	4999347	4999347	50.39	108,33

[A] The coverage long read was based on mapping long reads back to the assembly.
[B] The coverage short read was calculated in CLC on an assembly of only short reads to calculate the average matched coverage
[C] Mean %GC of *K. pneumoniae* 57.15% (NCBI) and mean %GC of *E. coli* 50.6%
[D] Mean genome size of *K. pneumoniae* 5.596 Mb (NCBI) was used and mean genome size of *E. coli* 5.14213 Mb (NCBI) was used to calculate expected percentage
[E] This sample was discarded from further analysis due to the high percentage of expected genome size

7

MOBILE COLISTIN RESISTANCE *MCR-4.3*- AND *MCR-4.6*-HARBORING PLASMIDS IN LIVESTOCK- AND HUMAN-RETRIEVED *ENTEROBACTERALES* IN THE NETHERLANDS

Casper Jamin[1,#], Michael S.M. Brouwer[2], Kees T. Veldman[2], Sandra Witteveen[3], Fabian Landman[3], Edou Heddema[4], Paul H.M. Savelkoul[1,5], Lieke van Alphen[1#] and Antoni P.A. Hendrickx[3#], on behalf of the Dutch CPE Surveillance Study Group[†].

1. Department of Medical Microbiology, Infectious Diseases & infection Prevention Care and Public Health Research Institute (CAPHRI), Maastricht University Medical Center+, Maastricht, The Netherlands.
2. Wageningen Bioveterinary Research, Lelystad, The Netherlands.
3. Center for Infectious Disease Control (CIb), National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands.
4. Department of Medical Microbiology and Infection Control, Zuyderland Medical Center, Sittard-Geleen/Heerlen, the Netherlands.
5. Department of Medical Microbiology and Infection Control, Amsterdam UMC, location VUmc, Amsterdam, The Netherlands.

† Members are listed in the Acknowledgements section.
# These authors share senior authorship

Manuscript in preparation

Correspondence

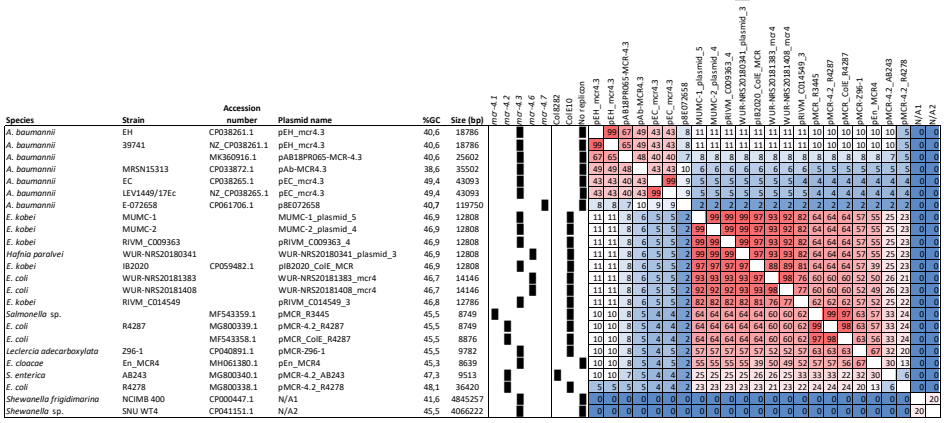
Since the first report of the mobile colistin resistance gene *mcr-1*, ten *mcr* variants were identified among antibiotic-resistant *Enterobacterales* which were mostly located on plasmids[1,2]. The *mcr* genes encode phosphoethanolamine transferases capable to modify membrane-associated lipopolysaccharide thereby ultimately leading to colistin resistance. Colistin is considered a last-resort antibiotic for treatment of human *Enterobacterales* infections but is also used in veterinary medicine.

Until now, the presence of *mcr-4* has not been described before in The Netherlands. The major objective of this study was to analyse *mcr-4* encoding plasmids from *Enterobacterales* obtained from humans and livestock in The Netherlands. To address this, a search for *mcr-4*-containing *Enterobacterales* isolates was performed in two surveillance databases containing bacterial next-generation sequencing (NGS) data. The carbapenemase-producing *Enterobacterales* (CPE) surveillance collection from the National Institute for Public Health and the Environment (RIVM) and the veterinary typing surveillance collection from the Wageningen Bioveterinary Research (WBVR). In addition, two clinical isolates from the Zuyderland Medical Center (ZMC) were analysed. Isolates from ZMC were analysed at the bacterial typing laboratory in the Maastricht University Medical Centre (MUMC). Isolates were subjected to Illumina short-read and Nanopore long-read sequencing by different in-house methods.

Analysis from the national CPE surveillance collection of the RIVM revealed two *Enterobacter* sp. with multi-locus sequence typing (MLST) ST54 out of 3,008 *Enterobacterales* (0.07%) that carried *mcr-4.3* in the period 2012 until 2020. Average nucleotide identity (ANI) calculation using http://enve-omics.ce.gatech.edu/ani/index of the *Enterobacter* sp. ST54 with *Enterobacter kobei* strain UCI 23 (Genbank accession NZ_KI973153.1) revealed an ANI of 99.09%, demonstrating that the *mcr-4.3*-carrying isolates are *E. kobei*. The collection of the WBVR revealed 3 isolates (1 *Hafnia paralvei,* 2 *E. coli*) that carried *mcr-4.6*, and the collection of the ZMC contained also two clinical isolates *E. kobei* ST54 with *mcr-4.3*. Hybrid assembly of short-read and long-read sequencing data was performed using Unicycler[3] (v.0.4.8, default settings) to reconstruct *mcr-4* plasmids, which were compared to internationally reported *mcr-4* plasmid sequences from NCBI using BioNumerics. Resistance to colistin was determined by broth microdilution in all seven isolates that harbor *mcr-4* alleles. Seven highly similar (76-99%) and small (12.8-14.1 kb) *mcr-4* plasmids with a ColE10 replicon were identified with either a *mcr-4.3* allele in four human *E. kobei* isolates (MIC colistin <1 mg/L), or a *mcr-4.6* allele in three livestock-associated isolates (MIC colistin 4-8 mg/L) collected between 2012-2020 (Fig. 1A). More specifically, the MUMC-1_plasmid_5, MUMC-2_plasmid_4, and RIVM-C009363_mcr4 were 99% identical plasmids of 12.8kb with a %G+C of 46.92. The WUR-341_plasmid_3, WUR-NRS20181383_mcr4

and WUR-NRS_20181408_mcr4 had slightly different % G+C and a *mcr-4.6* allele possibly conferring resistance to colistin. Four human *E. kobei* isolates (MUMC-01, MUMC-02, RIVM_C014549, RIVM_C009363) carried both *mcr-4.3* and *mcr-9* of which *mcr-9* is known not to encode colistin resistance in *Enterobacter*[4]. The RIVM_C009363 isolate contained in addition the *ant(2")-Ia*, bla_{ACT-9}-like, $bla_{CTX-M-9}$, *qnrA1*, *sul1* and *tetA* resistance genes and lacks a carbapenemase gene.

A



Species	Strain	Accession number	Plasmid name	%GC	Size (bp)
A. baumannii	EH	CP038261.1	pEH_mcr4.3	40,6	18786
A. baumannii	39741	NZ_CP038261.1	pEH_mcr4.3	40,6	18786
A. baumannii		MK360916.1	pAB18PR065-MCR-4.3	40,6	25602
A. baumannii	MRSN15313	CP033872.1	pAb-MCR4.3	38,6	35502
A. baumannii	EC	CP038265.1	pEC_mcr4.3	49,4	43093
A. baumannii	LEV1449/17Ec	NZ_CP038265.1	pEC_mcr4.3	49,4	43093
A. baumannii	E-072658	CP061706.1	pBE072658	40,7	119750
E. kobei	MUMC-1		MUMC-1_plasmid_5	46,9	12808
E. kobei	MUMC-2		MUMC-2_plasmid_4	46,9	12808
E. kobei	RIVM_C009363		pRIVM_C009363_4	46,9	12808
Hafnia paralvei	WUR-NRS20180341		WUR-NRS20180341_plasmid_3	46,9	12808
E. kobei	IB2020	CP059482.1	pIB2020_ColE_MCR	46,9	12808
E. coli	WUR-NRS20181383		WUR-NRS20181383_mcr4	46,7	14146
E. coli	WUR-NRS20181408		WUR-NRS20181408_mcr4	46,7	14146
E. kobei	RIVM_C014549		pRIVM_C014549_3	46,8	12786
Salmonella sp.		MF543359.1	pMCR_R3445	45,5	8749
E. coli	R4287	MG800339.1	pMCR-4.2_R4287	45,5	8749
E. coli		MF543358.1	pMCR_ColE_R4287	45,5	8876
Leclercia adecarboxylata	Z96-1	CP040891.1	pMCR-Z96-1	45,5	9782
E. cloacae	En_MCR4	MH061380.1	pEn_MCR4	45,3	8639
S. enterica	AB243	MG800340.1	pMCR-4.2_AB243	47,3	9513
E. coli	R4278	MG800338.1	pMCR-4.2_R4278	48,1	36420
Shewanella frigidimarina	NCIMB 400	CP000447.1	N/A1	41,6	4845257
Shewanella sp.	SNU WT4	CP041151.1	N/A2	45,5	4066222

B

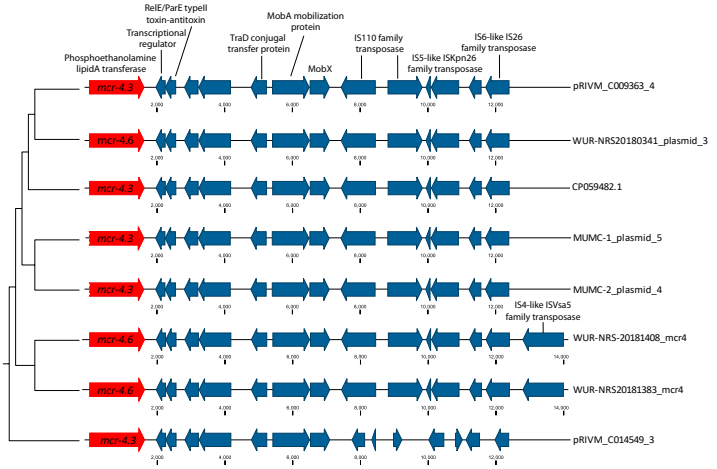


Figure 1A. Comparison of mcr-4 plasmids from the Netherlands in context with NCBI-retrieved plasmids. % G+C content and plasmid size in bp is indicated. Presence of the mcr-4-allele and replicons is indicated by black squares. Heatmap indicated percentage identity, in which high similarity was indicated in dark red (80-99%) and low similarity (<10%) in blue. **Figure 1B.** Alignment of mcr-4 plasmids from the Netherlands. Plasmids were rotated and the mcr-4 allele was set as starting point for comparison.

The RIVM_C014549 isolate was multi-resistant, containing the *aac(6')-Ib-cr*, *aac(6')-Ib3*, *aadA1*, *aph(3'')-Ib*, *aph(3')-XV*, *aph(6)-Id*, bla_{ACT-9}, bla_{SHV-12}, bla_{VIM-1}, *catB2*, *dfrA14*, *fosA*, *mph(A)*, *qnrS1*, and *sul1/2* genes. The human *E. kobei* isolates from the ZMC were also from MLST ST54 but were not clonal to the isolates from RIVM (191 to 270 SNPs different, inferred using Split Kmer Analysis, using default settings), and contained the *mcr-4.3*, *aph(6)-Id**, *aph(3'')-Ib**, *sul2**, *catA1**, $bla_{CTX-M-15}$*, *dfrA14*, *oqxA10*, *oqxB4*, bla_{OXA-1}, *aac(6')-Ib-cr*, *aac(3)-IIe*, *aadA1**, *mcr-9.1*, *qnrB1*, bla_{ACT-64}, *tet(A)* (* absent in MUMC-1) resistance genes. The livestock *E. coli* isolates were from ST216, and contained the bla_{TEM-1}, bla_{EC-15}, *dfrA1*, *aadA1*, *sul1*, *tet(B)* resistance genes and the *H. paralvei* contained the *mph(B), sul1, aadA1, dfrA1, mcr-9.1, tet(B)* and bla_{ACC-1b}. None of the seven isolates carrying *mcr-4* contained chromosomal mutations in the *pmrAB* genes known to be associated with colistin resistance, as inferred using ResFinder v4.1[5]. The *mcr-4* plasmids identified among *Enterobacterales* in the Netherlands were distinct from the *mcr-4* plasmids in the NCBI database (Fig. 1A). The Dutch *mcr-4* plasmid architecture was comparable and comprised the phosphoethanolamine lipid A transferase gene *mcr-4*, followed by a type II toxin/antitoxin system, TraD conjugal transfer protein, MobA/X mobilization proteins and transposases (Fig. 1B). Dutch livestock-retrieved isolates with *mcr-4.6* plasmids differed from the human-derived plasmids by one SNP in the *mcr-4.3* gene, leading to a *mcr-4.6* allelic variant, possibly causing resistance to colistin[6,7]. The RIVM_C014549_plasmid_2 was different from the other *mcr-4* plasmids obtained in the Netherlands. In summary, the occurrence of *mcr-4* plasmids among *Enterobacterales* is low in the Netherlands. The *mcr-4.6* plasmid may cause decreased colistin susceptibility in livestock *E. coli* based on MIC values of these isolates, while *mcr-4.3* in *E. kobei* does not. A recent study also failed to detect colistin-resistance in *E. kobei* ST54 co-harboring *mcr-4.3* and *mcr-9*[8]. In contrast, *mcr-4.3* conferred colistin-resistance in *Acinetobacter baumannii and A. nosocomialis*[9,10] indicating species-specific functionality of this colistin resistance gene. The *mcr-4* plasmids from human and livestock were highly similar, suggesting a zoonotic potential and unnoticed horizontal dissemination of colistin resistance in the Netherlands.

Acknowledgements

We thank all the members of the Dutch CPE surveillance study Group and the Dutch medical microbiology laboratories for submitting *Enterobacterales* isolates to the RIVM for the national CPE surveillance program. Members of the Dutch CPE Surveillance Study Group:

- Maijer-Reuwer, ADRZ medisch centrum, Department of Medical Microbiology, Goes
- M.A. Leversteijn-van Hall, Alrijne Hospital, Department of Medical Microbiology, Leiden
- W. van den Bijllaardt, Amphia Hospital, Microvida Laboratory for Microbiology, Breda
- R. van Mansfeld, Amsterdam UMC - location AMC, Department of Medical Microbiology, Amsterdam
- K. van Dijk, Amsterdam UMC - location Vumc, Department of Medical Microbiology and Infection Control, Amsterdam
- Zwart, Atalmedial, Department of Medical Microbiology, Amsterdam
- B.M.W. Diederen, Bravis Hospital/ZorgSaam Hospital Zeeuws-Vlaanderen, Department of Medical Microbiology, Roosendaal/Terneuzen
- J.W. Dorigo-Zetsma, CBSL, Department of Medical Microbiology, Hilversum
- D.W. Notermans, Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven
- Ott, Certe, Department of Medical Microbiology, Groningen
- W. Ang, Comicro, Department of Medical Microbiology, Hoorn
- J. da Silvia, Deventer Hospital, Department of Medical Microbiology, Deventer
- A.L.M. Vlek, Diakonessenhuis, Department of Medical Microbiology and Immunology, Utrecht
- A.G.M. Buiting, Elisabeth-TweeSteden (ETZ) Hospital, Department of Medical Microbiology and Immunology, Tilburg
- L. Bode, Erasmus University Medical Center, Department of Medical Microbiology, Rotterdam
- S. Paltansing, Franciscus Gasthuis & Vlietland, Department of Medical Microbiology and Infection Control, Rotterdam
- A.J. van Griethuysen, Gelderse Vallei Hospital, Department of Medical Microbiology, Ede
- M. den Reijer, Gelre Hospitals, Department of Medical Microbiology and Infection prevention, Apeldoorn
- M.J.C.A. van Trijp, Groene Hart Hospital, Department of Medical Microbiology and Infection Prevention, Gouda
- M. Wong, Haga Hospital, Department of Medical Microbiology, 's-Gravenhage

- A.E. Muller, HMC Westeinde Hospital, Department of Medical Microbiology, 's-Gravenhage
- M.P.M. van der Linden, IJsselland hospital, Department of Medical Microbiology, Capelle a/d IJssel
- M. van Rijn, Ikazia Hospital, Department of Medical Microbiology, Rotterdam
- S.B. Debast, Isala Hospital, Laboratory of Medical Microbiology and Infectious Diseases, Zwolle
- K. Waar, Certe Medische Microbiologie Friesland | Noordoostpolder, Department of Medical Microbiology, Leeuwarden
- E. Kolwijck, Jeroen Bosch Hospital, Department of Medical Microbiology and Infection Control, 's-Hertogenbosch
- N. Alnaiemi, LabMicTA, Regional Laboratory of Microbiology Twente Achterhoek, Hengelo
- T. Schulin, Laurentius Hospital, Department of Medical Microbiology, Roermond
- S. Dinant, Maasstad Hospital, Department of Medical Microbiology, Rotterdam
- S.P. van Mens, Maastricht University Medical Centre, Department of Medical Microbiology, Maastricht
- D.C. Melles, Meander Medical Center, Department of Medical Microbiology, Amersfoort
- M.P.A. van Meer, Rijnstate Hospital, Laboratory for Medical Microbiology and Immunology, Velp
- J.W.T. Cohen Stuart, Noordwest Ziekenhuisgroep, Department of Medical Microbiology, Alkmaar
- P. Gruteke, OLVG Lab BV, Department of Medical Microbiology, Amsterdam
- I.T.M.A. Overdevest, PAMM, Department of Medical Microbiology, Veldhoven
- van Dam, Public Health Service, Public Health Laboratory, Amsterdam
- Maat, Radboud University Medical Center, Department of Medical Microbiology, Nijmegen
- Maraha, Albert Schweitzer Hospital, Department of Medical Microbiology, Dordrecht
- J.C. Sinnige, Regional Laboratory of Public Health, Department of Medical Microbiology, Haarlem
- E.E. Mattsson, Reinier de Graaf Groep, Department of Medical Microbiology, Delft
- N. van Maarsenveen, Saltro Diagnostic Centre, Department of Medical Microbiology, Utrecht
- E. de Jong, Slingeland Hospital, Department of Medical Microbiology, Doetinchem
- S.J. Vainio, St Antonius Hospital, Department of Medical Microbiology and Immunology, Nieuwegein

- E. Heikens, St Jansdal Hospital, Department of Medical Microbiology, Harderwijk
- R. Steingrover, St. Maarten Laboratory Services, Department of Medical Microbiology, Cay Hill (St. Maarten)
- Troelstra, University Medical Center Utrecht, Department of Medical Microbiology, Utrecht
- E. Bathoorn, University of Groningen, Department of Medical Microbiology, Groningen
- T.A.M. Trienekens, VieCuri Medical Center, Department of Medical Microbiology, Venlo
- D.W. van Dam, Zuyderland Medical Centre, Department of Medical Microbiology and Infection Control, Sittard-Geleen
- E.I.G.B. de Brauwer, Zuyderland Medical Centre, Department of Medical Microbiology and Infection Control, Heerlen

Ethical statement

Ethical approval was not needed for the study, since it is based on surveillance data only. Samples from which the isolates were cultured, were all taken as part of routine health care.

Funding

This study was carried out as part of the Dutch National CPE surveillance, as part of the regular activities of the RIVM, financed by the Dutch Ministry of Health, Welfare and Sport.

Transparency declaration

The authors have nothing to disclose.

References

1. Ling, Z. *et al.* Epidemiology of mobile colistin resistance genes mcr-1 to mcr-9. *J. Antimicrob. Chemother.* **75**, 3087–3095 (2020).
2. Wang, C. *et al.* Identification of novel mobile colistin resistance gene mcr-10. *Emerg. Microbes Infect.* (2020) doi:10.1080/22221751.2020.1732231.
3. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* **13**, (2017).
4. Tyson, G. H. *et al.* The mcr-9 Gene of Salmonella and Escherichia coli Is Not Associated with Colistin Resistance in the United States. doi:10.1128/AAC.00573-20.
5. Zankari, E. *et al.* PointFinder: A novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *J. Antimicrob. Chemother.* **72**, 2764–2768 (2017).
6. Rebelo, A. R. *et al.* 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**, (2018).
7. Zhang, H. *et al.* Action and mechanism of the colistin resistance enzyme MCR-4. *Commun. Biol.* **2**, (2019).
8. Kim, J. S. *et al.* Identification of a carbapenem-resistant Enterobacter kobei clinical strain co-harbouring mcr-4.3 and mcr-9 in Republic of Korea. *Journal of global antimicrobial resistance* vol. 26 (2021).
9. Martins-Sorenson, N. *et al.* A novel plasmid-encoded mcr-4.3 gene in a colistin-resistant Acinetobacter baumannii clinical strain. *J. Antimicrob. Chemother.* **75**, (2020).
10. Snyman, Y. *et al.* Characterisation of mcr-4.3 in a colistin-resistant Acinetobacter nosocomialis clinical isolate from Cape Town, South Africa. *J. Glob. Antimicrob. Resist.* **25**, (2021).

1. Ling, Z. et al. Epidemiology of mobile colistin resistance genes mcr-1 to mcr-9. J. Antimicrob. Chemother. 75, 3087–3095 (2020).
2. Wang, C. et al. Identification of novel mobile colistin resistance gene mcr-10. Emerg. Microbes Infect. (2020) doi:10.1080/22221751.2020.1732231.
3. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput. Biol. 13, (2017).
4. Tyson, G. H. et al. The mcr-9 Gene of Salmonella and Escherichia coli Is Not Associated with Colistin Resistance in the United States. doi:10.1128/AAC.00573-20.
5. Zankari, E. et al. PointFinder: A novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. J. Antimicrob. Chemother. 72, 2764–2768 (2017).
6. Rebelo, A. R. et al. 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, (2018).
7. Zhang, H. et al. Action and mechanism of the colistin resistance enzyme MCR-4. Commun. Biol. 2, (2019).
8. Kim, J. S. et al. Identification of a carbapenem-resistant Enterobacter kobei clinical strain co-harbouring mcr-4.3 and mcr-9 in Republic of Korea. Journal of global antimicrobial resistance vol. 26 (2021).
9. Martins-Sorenson, N. et al. A novel plasmid-encoded mcr-4.3 gene in a colistin-resistant Acinetobacter baumannii clinical strain. J. Antimicrob. Chemother. 75, (2020).
10. Snyman, Y. et al. Characterisation of mcr-4.3 in a colistin-resistant Acinetobacter nosocomialis clinical isolate from Cape Town, South Africa. J. Glob. Antimicrob. Resist. 25, (2021).

8

DISSEMINATION OF A MOBILE RESISTANCE GENE ISLAND IN NOSOCOMIAL *ENTEROBACTERALES*

Casper Jamin[1], Edou Heddema[2], Marjolein F.Q. Kluytmans - van den Bergh[3,4,5], Wil van der Zwet[1], Erik Beuken[1], Christian J.P.A. Hoebe[1,6,7], Paul Savelkoul[1,8], Lieke van Alphen[1]

1. Department of Medical Microbiology, Care and Public Health Research Institute (CAPHRI), Maastricht University Medical Center+, Maastricht, The Netherlands.
2. Department of Medical Microbiology, Zuyderland Medical Center Heerlen/Sittard-Geleen, The Netherlands
3. Department of Infection Control, Amphia Hospital, Breda, the Netherlands
4. Amphia Academy Infectious Disease Foundation, Amphia Hospital, Breda, the Netherlands
5. Julius Center for Health Sciences and Primary Care, university Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands
6. Department of Sexual Health, Infectious Diseases and Environment, South Limburg Public Health Service, Heerlen, The Netherlands.
7. Department of Social Medicine, Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, The Netherlands.
8. Department of Medical Microbiology and Infection Control, Amsterdam UMC, location VUmc, Amsterdam, The Netherlands.

Manuscript in preparation

Abstract

Antimicrobial resistance (AMR) has become a worldwide challenge. AMR genes are often mobilized by integrons, plasmids, transposons or integrative and conjugative elements. During 2019, we observed the appearance of 27 multidrug-resistant (MDR) *Enterobacterales* with highly similar phenotypic AMR in two Dutch hospitals. Bacterial typing using whole-genome sequencing (WGS) was performed, and clusters were identified. However, a set of non-clonal *K. pneumoniae* and *Enterobacter* spp. isolates harbored the same set of AMR genes: *dfrA14*, *qnrB1*, *aac(3)-IIa*, *aac(6')-Ib-cr*, *aadA1*, *aph(3'')-Ib*, *aph(6)-Id*, $bla_{CTX\text{-}M\text{-}15}$, $bla_{OXA\text{-}1}$, $bla_{TEM\text{-}1}$, *catA1*, *catB3*, *sul2* and *tet(A)*. Long-read sequencing of selected isolates revealed co-localization in a 46kb resistance gene island, either on plasmids or the chromosome. Conjugation experiments with isolates containing this element integrated in the chromosome or plasmids were only successful for the isolates containing plasmids. Furthermore, this resistance gene island harbored many insertion sequences (IS), with *IS26* as the dominant one. Reference mapping of this region with TETyper showed the presence of this resistance gene island in 20/32 *Enterobacter cloacae* and 2/10 *Klebsiella pneumoniae* isolates from two hospitals in the province of Limburg in 2019. Investigation of this resistance island in a historical Dutch collection of 1806 extended-spectrum beta-lactamase-producing *Enterobacterales* isolates showed this gene island to be present in 37 isolates in 2012 and 2013 in various Dutch hospitals including one hospital in the present study. This study shows wide dissemination of a long-term stable mobile resistance gene island in different species conferring resistance to many antibiotics in several hospitals in The Netherlands.

Introduction

Increasing antimicrobial resistance has become a worldwide problem[1]. Nosocomial infections by multidrug-resistant (MDR) bacteria have decreased treatment options and do not favor a positive outcome for the patient[2–4]. Curbing the spread of antibiotic-resistant micro-organisms is key in the hospital setting. The role of mobile genetic elements carrying resistance genes is often overlooked and merely further investigated when resistance mechanisms of last-resort antibiotics such as carbapenems are involved[5,6]. Resistance genes can jump within the genome by integrases or 'hitchhike' with transposons between plasmids and chromosome[7]. Lesser known mechanisms for moving large fragments of DNA (20-500kb) are by 'integrative and mobilizable elements' (IME) and 'integrative and conjugative elements' (ICE, previously known as conjugative transposons), which can both circularize and conjugate

utilizing type VI secretion systems encoded on the element (ICE)[8–10] or elsewhere in the chromosome (IME)[11].

For surveillance of the spread of AMR bacteria in our hospitals, whole-genome sequencing (WGS) is used for further analysis of cultured isolates. Throughout 2019, *Enterobacterales* isolates that were resistant to all first-line antibiotics were identified from two hospitals in Limburg, The Netherlands. We studied whether the nosocomial spread of resistant isolates had occurred and what role mobile genetic elements play in the possible dissemination of antimicrobial resistance genes. Subsequently, we evaluated whether this resistance gene island was already present in a historical dataset of extended-spectrum beta-lactamase (ESBL)-producing *Enterobacterales* from hospitals throughout The Netherlands from 2011 to 2014[12,13].

Material and Methods

Strain selection & whole-genome sequencing

WGS was performed for typing of suspected outbreak *Enterobacterales*[14] isolates from patients' urine, feces, rectal swabs and blood cultures, and from the hospital environment (Supplemental spreadsheet 1). Antimicrobial susceptibility testing was done using VITEK® 2 (Biomerieux, Marcy-l'Étoile, France). DNA isolation, library preparation, WGS and subsequent quality control was performed as previously described[14]. In short, isolates were grown in Mueller Hinton broth (Becton Dickinson, Sparks, MD, United States) for 16-24 hours at 37°C. DNA was isolated using the Total Nucleotide isolation kit (Lucigen, Middleton, WI, United States). Sequencing libraries were made using NexteraXT (Illumina, San Diego, CA, United States). WGS was performed on a MiSeq sequencer (Illumina) using a 2x250 paired-end sequencing on a V2 flowcell. Six isolates were selected for additional long-read sequencing. Two clonal *E. cloacae* isolates, EC-02 and EC-03, (Figure 1) were chosen to investigate the absence of six resistance genes in isolate EC-02 compared to EC-03. Two clonal *E. cloacae (*EC-05 and EC-12) and two clonal *K. pneumoniae* (KP-01 and KP-07) were selected to determine the variation in structure of a possible resistance gene island. Long-read sequencing was performed using the same DNA extraction method. Rapid barcoding kit SQK-RBK004 from Oxford Nanopore was used to generate long-read sequencing libraries, and libraries were sequenced on the MinION platform (Oxford Nanopore Technologies) for 48 hours using R9.4 flowcells. Basecalling was performed using Albacore (v2.1.7). Barcode trimming and demultiplexing was performed using Porechop (v0.2.4). All WGS data generated, is available under BioProject PRJEB46126.

Bio-informatic workflow

All bio-informatic tools were run on default settings unless otherwise specified. Sequencing reads were assembled using SKESA (v2.3.0)[15]. Resistance genes and origin of replications (ORIs) were identified using Abricate (v1.0.1)[16]. Sequence types were inferred using MLST tool[17], using pubMLST schemes[18]. Correlations among the presence and absence of resistance genes and ORIs were calculated using Kendall rank correlations implemented by Pandas (v1.0.3) in Python (v3.7). To mitigate spurious results by a low number of observations, genotypes that were present less than 5 times were omitted from these correlation calculations. Genetic distances among genomes were determined on a SNP basis on raw reads, using Split K-mer analysis (SKA, 'ska fastq', v1.0)[19]. Hierarchical clustering of isolates based on SNPs was performed using the linkage function (method = 'single'), and dendrograms were generation using Scipy.cluster.hierarchy (v1.4.1). Clonality between isolates was determined when two isolates differed by 20 SNPs or less and shared more than 90% of their split k-mer content, as implemented by default by SKA. These cut-offs were validated and deemed suitable for this study, using the identified clusters in the external quality assessment study on bio-informatic workflows outbreak typing with standardized datasets[20]. Similarity among plasmid sequences was determined using Plasmidsimilarity (v1.0.0), https://github.com/casperjamin/plasmidsimilarity). Similarities of plasmids were calculated using the Jaccard dissimilarity index between the k-mer content of two plasmids, meaning that the number of unique k-mers present in both sequences is divided by the total number of unique k-mers of both sequences to calculate the Jaccard index. Data visualization was done using Seaborn and Matplotlib. Coding sequences were predicted using bakta[21] (v1.0.3) using database version 2.0. Locally Collinear Blocks (LCBs) in the resistance gene islands were aligned and visualized using ProgressiveMauve[22].

Conjugation experiment

The donor isolates KP-01 (plasmid-borne resistance gene island) and EC-05 (chromosomal borne resistance gene island) and a NaN_3 resistant J53 *E. coli*[23] (acceptor) were each grown overnight at 37 °C with agitation in LB medium, complemented with chloramphenicol (Cm; 20 µg/ml) or sodium-azide (NaN_3; 100 µg/ml), for the donor or the acceptor strain, respectively. Bacteria were diluted 1:10 in LB medium supplied with Cm or NaN_3, and incubated for another three hours. Subsequently, bacteria were pelleted and washed twice with 500 µl PBS. The cells were resuspended in 300 µl of PBS. Conjugation was performed by mixing 100 µl of donor and acceptor strains, and direct plating of 100 µl and 10 µl of the mixed cells onto LB-agar plates, supplied with 100 µg/ml NaN_3 and 20µg/ml chloramphenicol. Plates were incubated overnight at 37°C. Eight colonies were picked based on the expected colony morphology of

the acceptor strain and subjected to WGS to determine if conjugation occurred and which plasmids were transferred. Antimicrobial susceptibility was determined by disk diffusion tests. ESBL phenotype was determined by a combination disk-diffusion test with cefotaxime, ceftazidime and cefepime with and without clavulanic acid.

Resistance gene island typing

The multi-drug resistance genetic island in isolates was identified using TETyper[24]. Only isolates for which more than 50% of the resistance gene island could be aligned with KMA[25] were used for TEtyper. Clustering of these genetic elements in these isolates was performed based on the number of homozygous SNPs and deletions in the resistance gene island using the resistance gene island of strain KP-01 as reference. The full workflow is available at "https://github.com/Casperjamin/transposontyping" (v1.0.0). The reference sequence for the resistance gene island is available at "https://github.com/Casperjamin/transposon_madness/blob/main/sequences/reference.fasta". Additionally, the 'SoM study' dataset of 1806 Dutch ESBL-producing *Enterobacterales* collected from hospitalized patients were screened for the presence of this resistance gene island[12]. Datasets were deduplicated to only include one isolate for each species for each patient. For SoM study isolates, only isolates aligning more than 90% to the resistance gene island with KMA were used for identification with TETyper and visualization in this manuscript. Antimicrobial resistance genes of interest (https://github.com/Casperjamin/transposon_madness/blob/main/sequences/resistancegenestransposon.fasta") were identified directly from raw reads in these isolates using KMA[25] and resfinder database[26]. Reference- and resistance gene island sequences are also available under the digital object identifier (DOI): 10.5281/zenodo.5220710

Data summary

All sequencing data was deposited on EBI-ENA under bioProject PRJEB46126. Reference- and resistance gene island sequences are also available under the digital object identifier (DOI): 10.5281/zenodo.5220710

Results

Clusters of MDR bacterial isolates

For routine surveillance and requests from the hospital infection prevention (HIP) teams from the two regional hospitals, we characterized, by short-read WGS, a total of 54 MDR Gram-negative bacterial isolates in 2019. This included 32 *Enterobacter*

sp., 10 *Klebsiella pneumoniae*, 7 *Klebsiella oxytoca*, 3 *Pseudomonas aeruginosa* and 2 *Acinetobacter baumannii*. 27 isolates were resistant to ciprofloxacin, gentamycin, cotrimoxazole and had an ESBL phenotype (supplemental spreadsheet 1). In one hospital, we encountered a cluster of *E. cloacae* which was present for four months on a single ward and included 30 isolates of which 22 were subjected to WGS. Details of this outbreak and its mitigation has been described elsewhere[27]. All *A. baumannii, P. aeruginosa* and *K. oxytoca* isolates were clonally related as they had 4 or less SNPs amongst each other (data not shown). *K. pneumoniae* isolates grouped into 3 possible outbreak clusters (Figure 1A): i) KP-05 and KP-06 (10 SNPs), ii) KP-09 and KP-10 (3 SNPs), and iii) KP-01, KP-07 (18 SNPs) and possibly KP-02 (139, 137 SNPs compared to KP-01 and KP-07 respectively). The 32 *Enterobacter* sp. consisted of 5 outbreak clusters (Figure 1B): i) EC-21 and EC-30 (20 SNPs), ii) EC-06, EC-07, and EC-22 (range 187 to 368), iii) EC-02 and EC-03 (10 SNPs), iv) EC-29, EC-32, EC-20, EC-04, EC-24, EC-28, EC-19, EC-25, EC-26, EC-27, EC-31, and EC-18 (median 10 SNPs, range 1 to 21 SNPs) , and cluster v) EC-05, EC-15, EC-17, and EC-12 (median 6 SNPs, range 1 to 11 SNPs) together with EC-01 and EC-10 that had on average 230 SNPs to the other 4 isolates in this cluster. Even though EC-02 and EC-03 were clonal, EC-02 was missing $bla_{CTX\text{-}M\text{-}15}$, *aadA1*, *aph(3'')-Ib*, *aph(6)-Id*, *catA1* and a *sul2* resistance genes (supplemental spreadsheet 2).

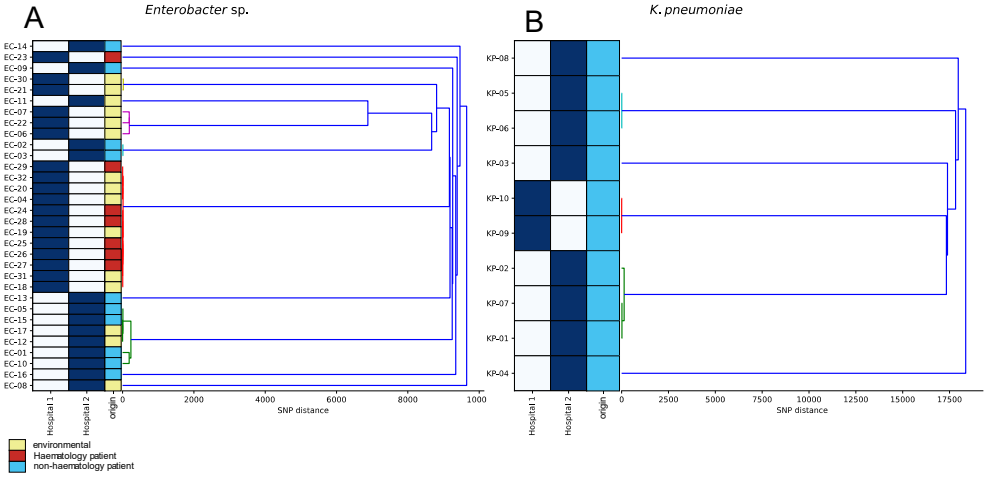


Figure 1. Phylogeny and origin of *Enterobacter* sp. **(A)** *K. pneumoniae* **(B)** isolates.
The different colored leaves in the dendrogram represent possible outbreak clusters.

Analysis of resistance gene clusters

During phenotypic cluster analysis, six *K. pneumoniae* and 20 *Enterobacter* sp. were resistant to ciprofloxacin, cotrimoxazole, gentamycin and were positive for the production of ESBL (Supplemental spreadsheet 1). Therefore, we investigated the possibility of the co-occurrence of a set of resistance genes on mobile and conjugative plasmids in these isolates.

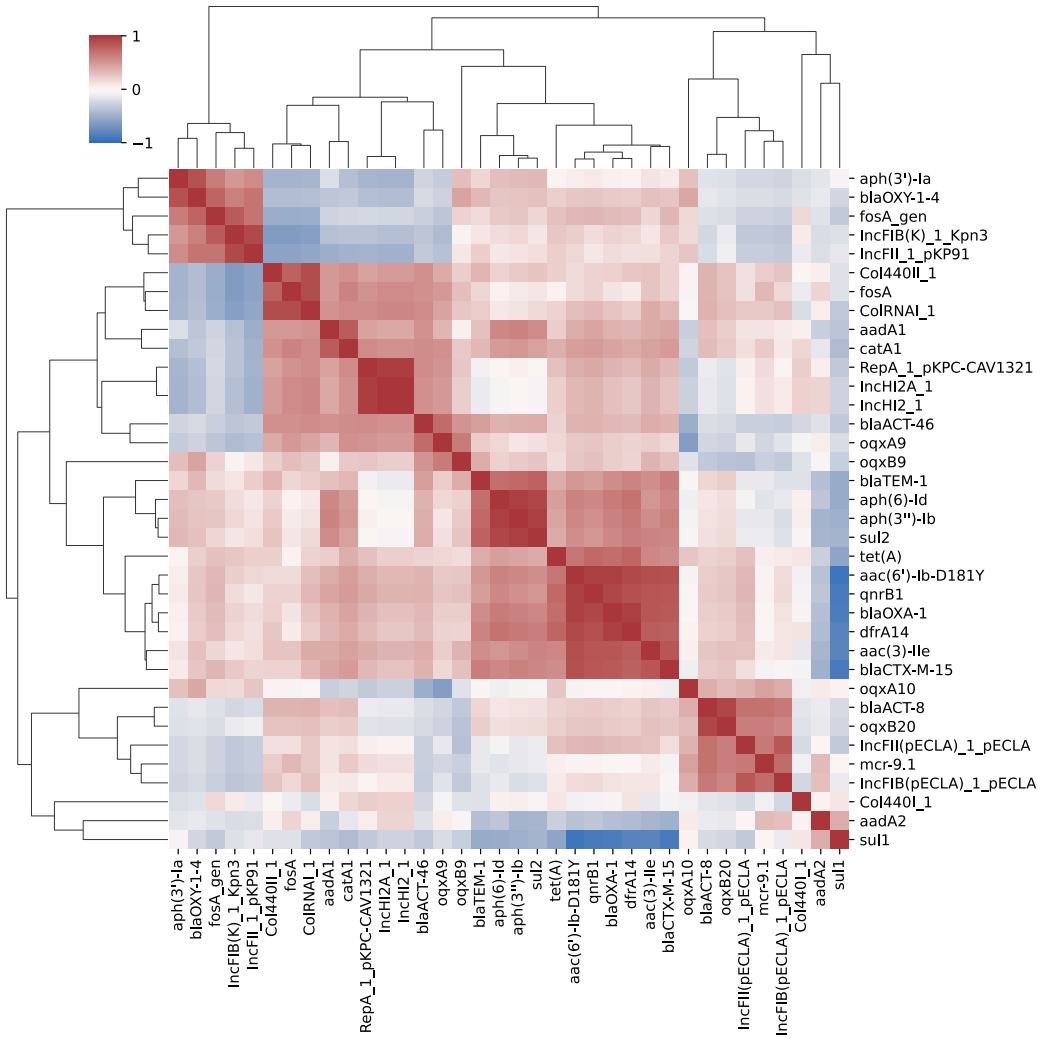


Figure 2. Heatmap of co-occurrence correlations of resistance genes and ORIs based on the Kendall Tau coefficient. Red cells (positive values) indicate a possible co-occurrence of these two genes. Blue cells (negative values) indicate a negative relation, in which one gene is present when the other one is not.

First, we compared the co-occurrence of resistance genes among isolates to identify possible gene clusters (Figure 2). Of main interest was *bla*$_{CTX-M-15}$, as this was the most abundant ESBL gene in the dataset (Present in 37/54, 69% of the isolates). When *bla*$_{CTX-M-15}$ was present, it co-occurred most often with *aac(3)-IIe* (34/37, 92%), *dfrA14* (36/37, 97%), *bla*$_{OXA-1}$ (36/37, 97%), *qnrB1* (35/37, 95%%), *aac(6)-Ib-D181Y* (36/37, 97%) , *tet(A)* (34/37, 92%), *sul2* (31/37, 84%), *aph(3'')-Ib* (32/37, 86%), *aph(6)-Id* (32/37, 86%), *and bla*$_{TEM-1}$ (29/37, 78%), which might indicate a resistance gene island. No clear association with an ORI was observed for this group. Furthermore, a co-occurring group consisting of *mcr-9, oqxA10, bla*$_{ACT-8}$, *oqxb20, IncFII(pECLA)_1_pECLA* and *IncFIB(pECLA)_1_pECLA* was identified.

Investigation of plasmids

To investigate the role of plasmids in the possible dissemination of the resistance genes, the location of the genes co-occurring with the ESBL gene *bla*$_{CTX-M-15}$ was determined in six isolates using long-read sequencing (see method section). All isolates carried the following resistance genes in a 46-48kb region: *dfrA14*, *qnrB1*, *aac(3)-IIa*, *aac(6')-Ib-cr*, *aadA1*, *aph(3'')-Ib*, *aph(6)-Id*, *bla*$_{CTX-M-15}$, *bla*$_{OXA-1}$, *bla*$_{TEM-1}$, *catA1*, *catB3*, *sul2* and *tet(A)*. Two isolates (EC-05, EC-12) carried the complete set of co-occurring genes on the chromosome. For EC-02, the resistance genes were localized on a linear contig of 751kb, but six resistance genes (*bla*$_{CTX-M-15}$, *aadA1*, *aph(3'')-Ib*, *aph(6)-Id*, *catA1* and a *sul2)* were lacking. Isolate EC-05 and EC-12 had these resistance genes encoded closely together in 48kb on a 3.0Mb and 4.9Mb contig respectively, indicating likely chromosomal carriage. Three isolates, KP01, KP07 and EC-03, carried large IncHI2_1, IncHI2A_1 plasmids (250-350kb) plasmids that encoded 14 resistance genes in a 46-48kb region. These genes decrease susceptibility against tetracyclines, fluoroquinolones, beta-lactams, trimethoprim, sulfonamides, aminoglycosides and chloramphenicol. These plasmids were highly similar as they shared 78% to 98% of their k-mer content (Figure 3). This plasmid also carried three gene clusters conferring metal resistance to mercury, copper, cobalt, nickel, tellurium and arsenic (Supplemental Figure 1).

Furthermore, EC-02 and EC-03 carried a ColE10 12kb plasmid encoding for a *mcr-4.3*, with no other known resistance genes co-localized (EC-02_plasmid_4 and EC-03_plasmid_5, figure 3).

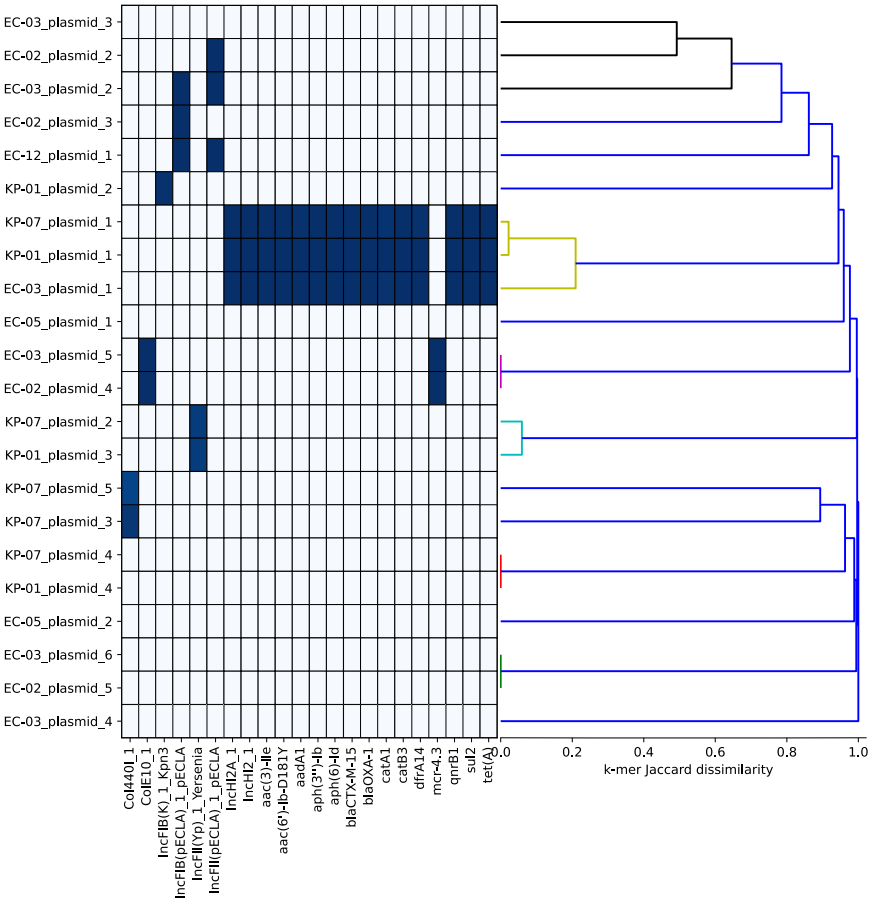


Figure 3. Heatmap (left) and dendrogram (right) showing the occurence of AMR genes and ORIs on plasmids. Dendrogram shows the similarity of among plasmids based on k-mer (k = 31) similarity.

Resistance gene island

The resistance gene islands showed high homology in the isolates that were long-read sequenced (Figure 4). In total, 4 locally collinear blocks (LCBs) of DNA sequences were identified of 20kb (red), 1.5kb (yellow), 5.5kb (green) and 19kb (blue) (Figure 4). EC-02 missed the (blue) 19kb region, that encodes for *bla*$_{CTX-M-15}$, *sul2, aph(6)-Id, aph(3'')-Ib, bla*$_{TEM-1}$, *aadA14 and catA1*.

Although these resistance genes were all encoded closely together, rearrangements and inversions were observed among the six isolates that were long-read sequenced. For KP01, the 5.5kb (green) region (*aac(6')-Ib-cr, bla*$_{OXA-1}$, *catB3, aac(3)-IIe*) seemed inverted between two inward coding *IS6* family *IS26* transposases of 723bp

(Figure 4). Further inspection of this region by comparing KP01 and KP07 shows that both isolates have the same *IS26-CCGGT- aac(6')-Ib-cr* motif upstream of this resistance gene. The rearrangement occurred between position 20450 and 27701 (7251bp, Figure 4), interrupting a hypothetical protein sequence before *IS26*. The regions from 20450 to 21218 and 26933 to 27701 are exact reverse complementary sequences, as the *IS26* gene was located here. In total, eight *IS26* genes, nine other transposase genes and three integrases were located on this gene island. Furthermore, a mobilization protein *mobC* and a tyrosine recombinase were located at position 11222 and 35662, respectively.

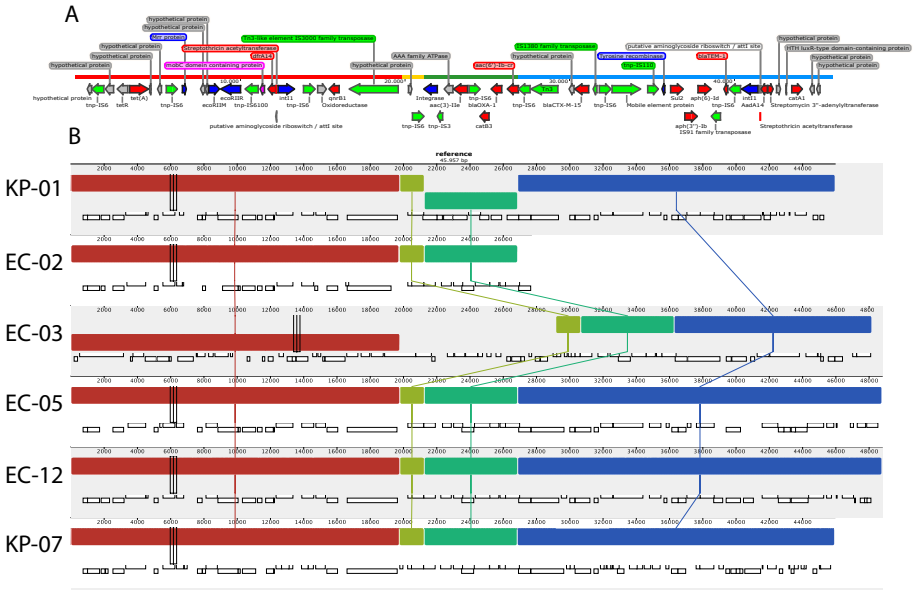


Figure 4. Panel A: layout of resistance gene island, annotated using bakta, and visualized using Snapgene. Panel B. ProgressiveMauve alignment of all isolates that were subjected to long-read sequencing. Colored bars indicate homologous sequences between two or more subject sequences. Colored bars underneath the main track for each sequence, indicates an inverse block.

Transfer of the resistance gene island

To demonstrate whether this plasmid or resistance gene island is mobile or conjugative as an integrative and mobilizable element (IME) or integrative and conjugative element (ICE), we performed conjugation experiments to an *E. coli* acceptor strain. Two donor isolates were used; i) chromosomally encoded resistance gene island EC-05, and ii) Plasmid encoded resistance gene island in isolate KP-01. EC-05 did encode for *virB* and *Tra* type IV secretion systems elsewhere in the genome,

which IMEs can potentially utilize for conjugation[11]. No conjugants were recovered for EC-05 as donor. For KP-01 as donor strain, several conjugants were observed on LB agar with chloramphenicol. Eight colonies were subjected to Illumina WGS. Six isolates turned out to be the original donor strain, while two isolates were the recipient strain (BioSample: ERS7182709 and ERS7182714). Both recipient isolates had contigs that covered the entire AMR KP-01_plasmid_1. Moreover, contigs that spanned the entire 129kb IncFIB(K)_1_Kpn3 plasmid from KP-01 (data not shown) were identified. Both conjugated plasmids encoded for type IV conjugation machinery (*traJ, traM, traY, traA*). Conjugates showed inhibition zones similar to the donor strain KP-01 for tetracycline, chloramphenicol, and gentamycin (Supplemental spreadsheet 3). For ciprofloxacin, the zone of inhibition decreased, although not to the levels of the donor strain. Furthermore, the two conjugates obtained an ESBL phenotype confirmed by the combination disk diffusion test, indicating the transfer of $bla_{\text{CTX-M-15}}$ (Supplemental spreadsheet 3).

Transposable element typing

Transposable element typing[24] was performed to determine if other isolates within this study carried the same resistance gene island, using KP-01 as a reference sequence (Figure 5). On average, 0 to 11553 SNPs (16 SNPs median) differences were observed among clonal isolates compared to 0 to 23953 SNPs (5095 median) among non-clonal isolates of the same species (Supplemental figure 2). One large cluster of 22 isolates (KP-07, EC-18, EC-31, EC-19, EC-27, EC-24, EC-25, EC-26, EC-28, EC-29, EC-20, EC-32, EC-04, EC-07, EC-17, EC-15, KP-01, EC-09, EC-03, EC-12, EC-05 and EC-10) of both *E. cloacae and K. pneumoniae* and had 0 to 185 SNPs (19 SNPs median, Figure 5). All of the *K. oxytoca* clustered together based on the number of SNPs by mapping the resistance gene island, as these isolates were all clonally-related.

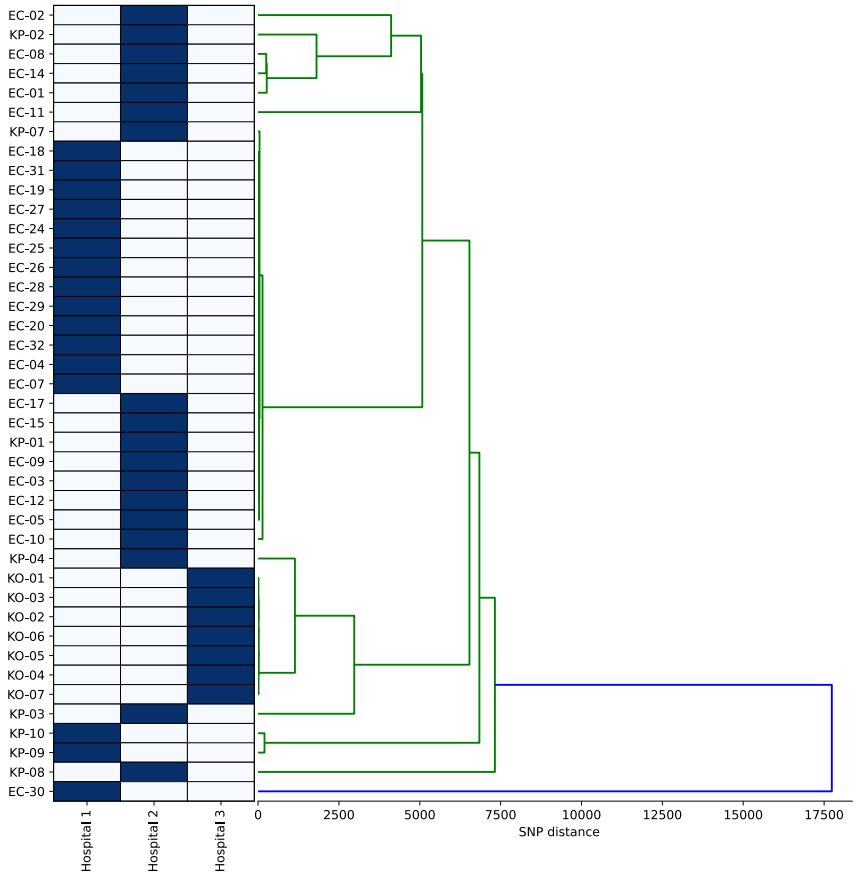


Figure 5. Dendrogram based on transposable element typer. Distances indicate the dissimilarity based on homozygeos SNPs and deletions, using KP-01 resistance gene island as a reference. Only isolates from 2019 were included.

Transposable element typing of Dutch ESBL *Enterobacterales*

WGS datasets of 1806 Dutch ESBL *Enterobacterales* from hospitalized patients collected from 2011 to 2014 were examined by TETyper[24] to determine if this resistance gene island was already present in The Netherlands in this period. This dataset consisted of 1356 *E. coli*, 270 *K. pneumoniae*, 121 *E. cloacae*. 18 *K. oxytoca* and 41 *Citrobacter sp.* (Supplemental Figure 3). In total, 6/1356 *E. coli,* 7/270 *K. pneumoniae*, 20/121 *E. cloacae*, 0/18 *K. oxytoca, 4/41 Citrobacter sp.* had at least 90% of the resistance gene island (Supplemental Figure 2). To identify whether a similar or identical resistance gene island was present in one of these historical isolates, SNPs distances were calculated compared to KP-01 (Figure 5). In total, 24 isolates were identified

with 5 or less SNPs comprising of *C. freundii*: 1 ST8 and 2 ST98, *Cronobacter:* 10 ST419 (ENA metadata indicated these isolates belonging to *Enterobacter* sp. complex), *Enterobacter* sp.: 2 ST24, 1 ST102, 1 ST121 and 1 ST168, *E. coli* 1 ST399, 1 ST401, 1 ST1394 *K. pneumoniae*: 3 ST 252. 14 isolates (10 *Cronobacter* ST419*,* 1 *E. cloacae* ST168, 2 *K. pneumoniae* and 1 *E. coli* ST1394) did not have any SNPs compared to KP-01. Interestingly, for all isolates carrying the resistance gene island in this study but also for the SoM dataset, the chloramphenicol gene *catB3*, only encoded the first ~70% of the gene (Figure 6). ERR1616179 to ERR1616442 did not encode for *aadA1* and *catA1*, which are co-located on the resistance gene island (Figure 4). None of the SoM isolates was clonally related to the isolates from this study (data not shown). Interestingly, 37/50 isolates from the SoM study originated from one of the participating hospitals in the present study (Supplemental spreadsheet 4).

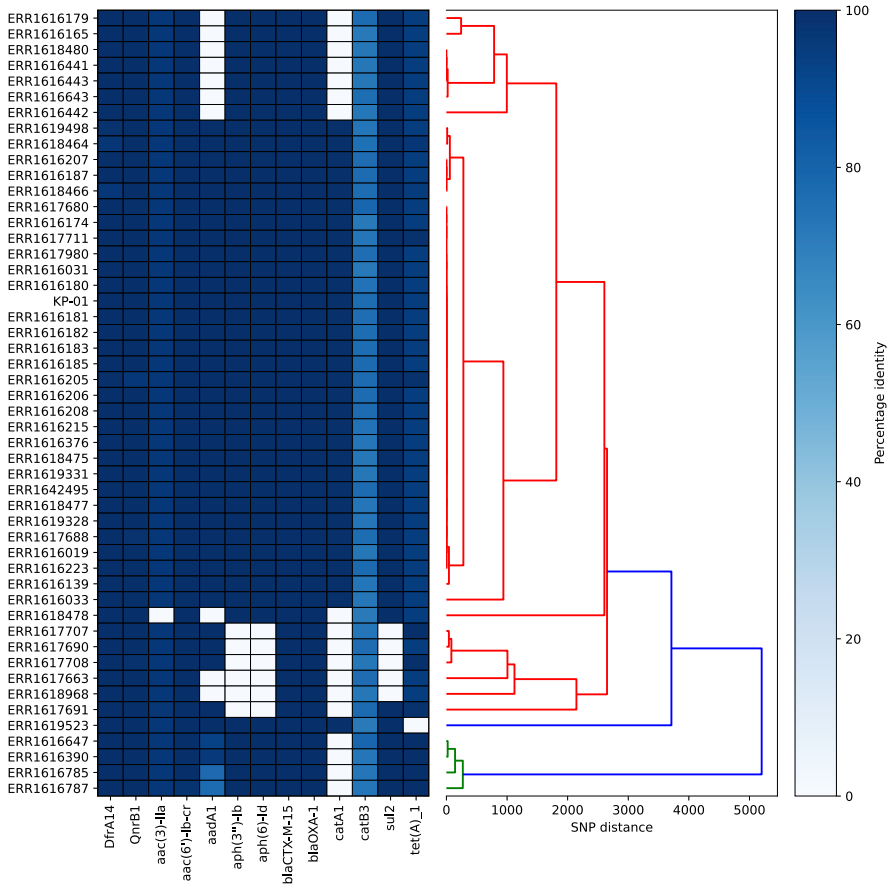


Figure 6. Left: Heatmap indicating the presence or absence of a particular resistance gene or ORI. Right: Dendrogram based on transposable element typer. Distance indicate the dissimilarity based on homozygous SNPs and deletions, using KP-01 resistance gene island as a reference.

Discussion & Conclusion

In total, 26 *E. cloacae* and *K. pneumoniae* isolates were identified with the same or highly similar resistance gene island originating from hospitals in the same region in the south of the Netherlands. Co-occurrence of a set of resistance genes indicated a possible resistance gene island, which was confirmed by long-read sequencing of six isolates. This resistance gene island encoded 14 different antibiotic resistance genes that cause a decreased susceptibility against 7 different classes of antibiotics: tetracyclines, fluoroquinolones, beta-lactams, trimethoprim, sulfonamides, aminoglycosides and chloramphenicol. Furthermore, several metal resistance genes were identified on the plasmids carrying the resistance gene island. All *E. cloacae* isolates from hospital 1 originated from the hematology department. Arsenic resistance was one of the metal resistance genes present and could select for these bacteria when hematology patients are treated with arsenic trioxide.

Many AMR genes on the resistance gene island were flanked by *IS26*. The *IS26*-mediated aggregation of transposable elements into composite structures is a previously well-described phenomenon[28–31]. These *IS26* are often identified together with AMR genes[32]. *IS26* preferably insert themselves into other *IS26* transposases and as such generate entire arrays of *IS26* with their cargo genes[32]. This phenomenon may lead to the co-selection of different AMR genes when these bacteria are exposed to one type of antibiotic, resulting in increasingly drug-resistant isolates. The assembly of the entire resistance gene island that encodes for seven different classes of antibiotics might have been a result of this co-selection. For the isolates described here, inversion of an entire block flanked by two inversely oriented *IS26: IS26-catB3-bla*$_{OXA-1}$*-aac(6')-Ib-cr-IS26* (green block, Figure 3) was observed. It is unclear whether this was caused by *IS26* mediated inversion or by homologous recombination between two copies of the AMR plasmid, using the two inverse *IS26* as homologous sites for recombination. Only intra-molecular inversion of DNA sequences at random insertion sites, using one *IS26* has been described so far[7,31]. No mechanism of inversion of DNA using two IS26 transposases has been described as of yet. Therefore, it seems more likely that this inversion event was caused by homologous recombination.

The 46kb resistance gene island was integrated into the chromosome in some isolates. The required machinery was present to allow for mobilization of large genetic islands by *mobC* or tyrosine recombinase, both present on the island (Figure 3). Movement like this resembles the movement of IME or ICE[9,11]. Although IMEs can conjugate to other bacteria by hijacking conjugation machinery elsewhere in the genome[11], no transfer of this resistance gene island out of the chromosome was achieved.

Most of the isolates of the SoM dataset that encoded for the resistance gene island originated from the same hospital as the isolates from 2019, which may indi-

cate that this resistance gene island has been present in this hospital for at least seven years in various isolates and species, indicating the transmission of this gene island within this hospital for a long time. The main two species carrying this resistance gene island were *E. cloacae* and *K. pneumoniae,* it was hardly detected in *E. coli* (Supplemental Figure 3). Conjugation experiments performed here demonstrated the possibility for *E. coli* to receive the AMR plasmid encoding the resistance gene island. It remains unclear why this apparent bacterial host specificity of this resistance gene island exists.

Limitations

Several limitations of this study should be noted. First, only a small subset of isolates was subjected to long-read sequencing. Therefore, it remains unclear for the other isolates if this resistance gene island is localized on the chromosome or on a plasmid. This also holds for the isolates from the SoM study. Next, no common source could be attributed to the origin of this resistance gene island, as it seemed that this element was already present in 2012 in at least one of the hospitals. Secondly, only isolates with the resistance gene island described here can be identified with an identical or almost identical structure and single nucleotide variants. Furthermore, the resistance gene island described in this study was identified by a strong co-occurrence of resistance genes; however, less frequently occurring gene islands may also disseminate within a hospital. Only AMR genes of high priority in hospitals, such as bla_{KPC} carbapenemases, might be more easily identified in different isolates but with a common source[6], compared to the more prevalent ESBL gene $bla_{CTX\text{-}M\text{-}15}$.

Concluding remarks

This study shows the different layers of complexity in the dissemination of antimicrobial resistance. First, small outbreak clusters were observed, all having highly similar antibiotic resistance phenotypes and genotypes. Secondly, resistance genes co-occurred closely together on the chromosome and on plasmids, indicating the spread and persistence of plasmids. Finally, typing the transposon on SNP level revealed the spread of this single mosaic resistance gene island in different isolates. The dissemination of such an island is probably driven by local antibiotic pressure and possibly by other drugs (here arsenic(III)oxide) that are otherwise lethal to the bacterial host, and thus enabling bacteria to survive the most commonly used antimicrobial agents in hospitals.

Acknowledgements

We are grateful for the SoM study group for the collection and generation of the WGS dataset of Dutch ESBL producing *Enterobacterales*.

Funding statement

The development of the Plasmidsimilarity tool was made possible by funding from the Dutch Workgroup for molecular diagnostics of infectious diseases (WMDI)

Conflicts of interest

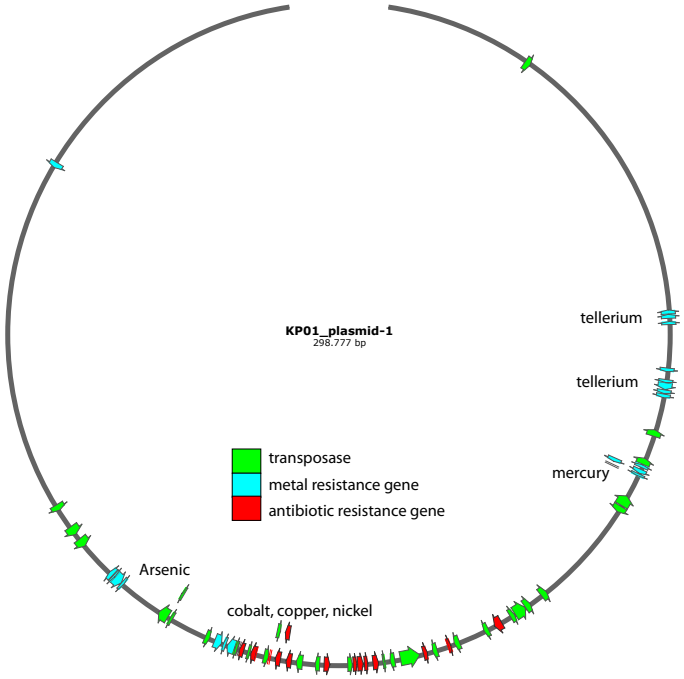
The authors declare that there are no conflicts of interest.

References

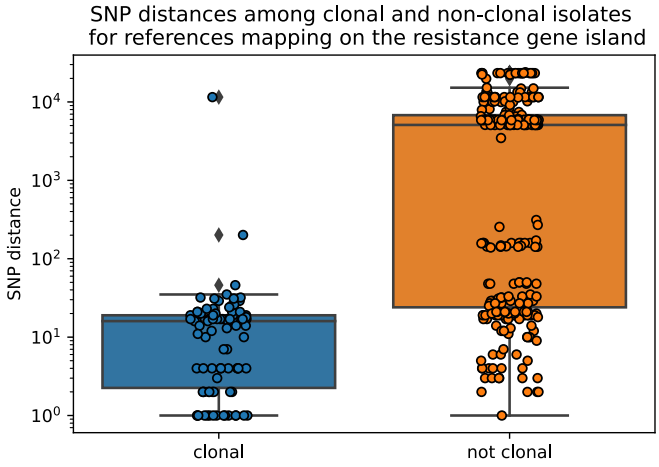
1. Lee Ventola, C. The Antibiotic Resistance Crisis Part 1: Causes and Threats. **40**, (2015).
2. Founou, R. C., Founou, L. L. & Essack, S. Y. Clinical and economic impact of antibiotic resistance in developing countries: A systematic review and meta-analysis. *PLoS ONE* (2017) doi:10.1371/journal.pone.0189621.
3. Thaden, J. T. *et al.* Increased costs associated with bloodstream infections caused by multidrug-resistant gram-negative bacteria are due primarily to patients with hospital-acquired infections. *Antimicrob. Agents Chemother.* (2017) doi:10.1128/AAC.01709-16.
4. Barrasa-Villar, J. I., Aibar-Remón, C., Prieto-Andrés, P., Mareca-Doñate, R. & Moliner-Lahoz, J. Impact on Morbidity, Mortality, and Length of Stay of Hospital-Acquired Infections by Resistant Microorganisms. *Clin. Infect. Dis.* **65**, (2017).
5. Laura Hidalgoa, Mark de Beena, Malbert R. C. Rogersa, Anita C. Schürcha, Jelle Scharringaa, Anneke van der Zee, Marc J. M. Bontena, A. C. F. Sequence-Based Epidemiology of an OXA-48 Plasmid during a Hospital Outbreak. 1–9 (2019).
6. Sheppard, A. E. *et al.* Nested Russian doll-like genetic mobility drives rapid dissemination of the carbapenem resistance gene blakpc. *Antimicrob. Agents Chemother.* **60**, 3767–3778 (2016).
7. Varani, A., He, S., Siguier, P., Ross, K. & Chandler, M. The IS6 family, a clinically important group of insertion sequences including IS26. *Mob. DNA* **12**, 1–18 (2021).
8. Burrus, V. & Waldor, M. K. Shaping bacterial genomes with integrative and conjugative elements. *Research in Microbiology* vol. 155 (2004).
9. Delavat, F., Miyazaki, R., Carraro, N., Pradervand, N. & van der Meer, J. R. The hidden life of integrative and conjugative elements. *FEMS microbiology reviews* vol. 41 (2017).
10. Johnson, C. M. & Grossman, A. D. Integrative and Conjugative Elements (ICEs): What They Do and How They Work. *Annu. Rev. Genet.* **49**, (2015).
11. Guédon, G., Libante, V., Coluzzi, C., Payot, S. & Leblond-Bourget, N. The obscure world of integrative and mobilizable elements, highly widespread elements that pirate bacterial conjugative systems. *Genes (Basel).* **8**, (2017).
12. Kluytmans-Van Den Bergh, M. F. Q. *et al.* Whole-genome multilocus sequence typing of extended-spectrum-beta-lactamase-producing enterobacteriaceae. *J. Clin. Microbiol.* (2016) doi:10.1128/JCM.01648-16.
13. Kluytmans-van den Bergh, M. F. Q. *et al.* Contact precautions in single-bed or multiple-bed

rooms for patients with extended-spectrum β-lactamase-producing Enterobacteriaceae in Dutch hospitals: a cluster-randomised, crossover, non-inferiority study. *Lancet Infect. Dis.* **19**, 1069–1079 (2019).
14. Jamin, C. *et al.* Harmonization of whole genome sequencing for outbreak surveillance of Enterobacteriaceae and Enterococci. *Microb. genomics* **7**, 2020.11.20.392399 (2021).
15. Souvorov, A., Agarwala, R. & Lipman, D. J. SKESA: Strategic k-mer extension for scrupulous assemblies. *Genome Biol.* (2018) doi:10.1186/s13059-018-1540-z.
16. Seemann, T. Abricate, https://github.com/tseemann/abricate.
17. Seemann, T. mlst https://github.com/tseemann/mlst.
18. Jolley, K. A., Bray, J. E. & Maiden, M. C. J. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications [version 1; referees: 2 approved]. *Wellcome Open Res.* **3**, (2018).
19. Harris, S. R. SKA: Split Kmer Analysis Toolkit for Bacterial Genomic Epidemiology. *bioRxiv* (2018) doi:10.1101/453142.
20. Coolen, J. P. M. *et al.* Centre-specific bacterial pathogen typing affects infection-control decision making. *Microb. Genomics* **7**, (2021).
21. Schwengers, O., Blom, J. & Goesmann, A. Bakta: Rapid & standardized annotation of bacterial genomes & plasmids. (2021).
22. Darling, A. E., Mau, B. & Perna, N. T. Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* (2010) doi:10.1371/journal.pone.0011147.
23. Göttig, S., Gruber, T. M., Stecher, B., Wichelhaus, T. A. & Kempf, V. A. J. In vivo horizontal gene transfer of the carbapenemase OXA-48 during a nosocomial outbreak. *Clin. Infect. Dis.* **60**, 1808–1815 (2015).
24. Sheppard, A. E. *et al.* Tetyper: A bioinformatic pipeline for classifying variation and genetic contexts of transposable elements from short-read whole-genome sequencing data. *Microb. Genomics* **4**, (2018).
25. Clausen, P. T. L. C., Aarestrup, F. M. & Lund, O. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinformatics* (2018) doi:10.1186/s12859-018-2336-6.
26. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* (2012) doi:10.1093/jac/dks261.
27. Zwet, W. Van der *et al.* Role of the Environment in Transmission of Multiresistant Enterobacter cloacae in a Hematology-Oncology Department. *Infect. Control Hosp. Epidemiol.* **41**, s379–s380 (2020).
28. He, D. D. *et al.* Antimicrobial resistance-encoding plasmid clusters with heterogeneous MDR regions driven by IS26 in a single Escherichia coli isolate. *J. Antimicrob. Chemother.* **74**, 1511–1516 (2019).
29. Harmer, C. J., Pong, C. H. & Hall, R. M. Structures bounded by directly-oriented members of the IS26 family are pseudo-compound transposons. *Plasmid* vol. 111 (2020).
30. Wang, X. C., Lei, C. W., Kang, Z. Z., Zhang, Y. & Wang, H. N. IS26-Mediated Genetic Rearrangements in Salmonella Genomic Island 1 of Proteus mirabilis. *Front. Microbiol.* **10**, (2019).
31. He, S. *et al.* Insertion sequence IS26 reorganizes plasmids in clinically isolated multidrug-resistant bacteria by replicative transposition. *MBio* **6**, (2015).
32. Harmer, C. J., Moran, R. A. & Hall, R. M. Movement of IS26-Associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS26 and preferentially inserts adjacent to another IS26. *MBio* **5**, (2014).

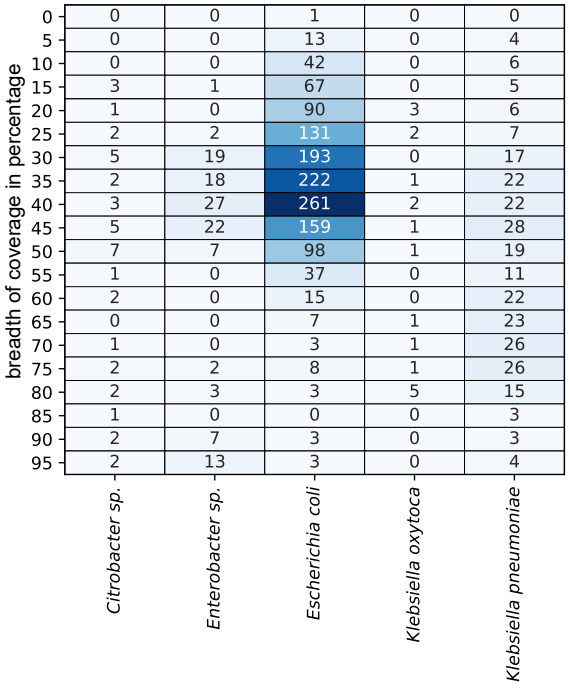
Supplemental Materials



Supplemental figure 1. KP01 resistance plasmid, arrows indicate genes. Visualization was done using Snapgene viewer.



Supplemental figure 2. Boxplot showing the number of SNPs on the resistance gene island between either clonal strains or non-clonal strains. Y-axis is on log scale, zero values were offset to 1 to facilitate log-scaling.



Supplemental figure 3. Heatmap showing the breadth of coverage on the resistance gene island for each isolate split up for each species.

Supplemental spreadsheet 1. Metadata of bacterial isolates, BioSample codes, hospital origin and antibiotic diagram.

species	recoded	ENA	stamnaam_ ENA	hospital recoded	sample type	Amoxicilline	Ampiciline	Amoxicilline Clavulaanzuur
Acinetobacter baumannii	AB-01	ERS7182715	AB-01	Hospital 1	urine			
Acinetobacter baumannii	AB-02	ERS7182716	AB-02	Hospital 1	airways			
E. cloacae	EC-01	ERS7182717	EC-01	Hospital 2	nasopharynx	R	R	R
E. cloacae	EC-02	ERS7182718	EC-02	Hospital 2	nasopharynx	R	R	R
E. cloacae	EC-03	ERS7182719	EC-03	Hospital 2	nasopharynx	R	R	R
E. cloacae	EC-04	ERS7182720	EC-04	Hospital 1	hospital environment		R	
E. cloacae	EC-05	ERS7182721	EC-05	Hospital 2	nasopharynx	R	R	R
E. cloacae	EC-06	ERS7182722	EC-06	Hospital 1	hospital environment		R	
E. cloacae	EC-07	ERS7182723	EC-07	Hospital 1	hospital environment		R	
E. cloacae	EC-08	ERS7182724	EC-08	Hospital 2				
E. cloacae	EC-09	ERS7182725	EC-09	Hospital 2	catheter	R	R	R
E. cloacae	EC-10	ERS7182726	EC-10	Hospital 2	nasopharynx	R	R	R
E. cloacae	EC-11	ERS7182727	EC-11	Hospital 2				
E. cloacae	EC-12	ERS7182728	EC-12	Hospital 2	hospital environment	R	R	R
E. cloacae	EC-13	ERS7182729	EC-13	Hospital 2	nasopharynx	R	R	R
E. cloacae	EC-14	ERS7182730	EC-14	Hospital 2	hospital environment	R	R	R
E. cloacae	EC-15	ERS7182731	EC-15	Hospital 2				
E. cloacae	EC-16	ERS7182732	EC-16	Hospital 2	nasopharynx	R	R	R
E. cloacae	EC-17	ERS7182733	EC-17	Hospital 2	hospital environment	R	R	R
E. cloacae	EC-18	ERS7182734	EC-18	Hospital 2	hospital environment		R	
E. cloacae	EC-19	ERS7182735	EC-19	Hospital 1	hospital environment		R	
E. cloacae	EC-20	ERS7182736	EC-20	Hospital 1	hospital environment		R	
E. cloacae	EC-21	ERS7182737	EC-21	Hospital 1	hospital environment		R	
E. cloacae	EC-22	ERS7182738	EC-22	Hospital 1	hospital environment		R	
E. cloacae	EC-23	ERS7182739	EC-23	Hospital 1	Haematology; urine			R
E. cloacae	EC-24	ERS7182740	EC-24	Hospital 1	Haematology; faeces		R	
E. cloacae	EC-25	ERS7182741	EC-25	Hospital 1	Haematology; urine			R
E. cloacae	EC-26	ERS7182742	EC-26	Hospital 1	haematology; blood culture		R	
E. cloacae	EC-27	ERS7182743	EC-27	Hospital 1	haematology; faeces			R
E. cloacae	EC-28	ERS7182744	EC-28	Hospital 1	haematology; rectal swab		R	
E. cloacae	EC-29	ERS7182745	EC-29	Hospital 1	haematology; rectal swab		R	
E. cloacae	EC-30	ERS7182746	EC-30	Hospital 1	hospital environment		R	
E. cloacae	EC-31	ERS7182747	EC-31	Hospital 1	hospital environment		R	
E. cloacae	EC-32	ERS7182748	EC-32	Hospital 1	hospital environment		R	
K. oxytoca	KO-01	ERS7182749	KO-01	Hospital 1				
K. oxytoca	KO-02	ERS7182750	KO-02	Hospital 3				
K. oxytoca	KO-03	ERS7182751	KO-03	Hospital 3				
K. oxytoca	KO-04	ERS7182752	KO-04	Hospital 3				
K. oxytoca	KO-05	ERS7182753	KO-05	Hospital 3				
K. oxytoca	KO-06	ERS7182754	KO-06	Hospital 3				
K. oxytoca	KO-07	ERS7182755	KO-07	Hospital 3				
K. pneumoniae	KP-01	ERS7182756	KP-01	Hospital 3		R	R	R
K. pneumoniae	KP-02	ERS7182757	KP-02	Hospital 2	urine	R	R	R
K. pneumoniae	KP-03	ERS7182758	KP-03	Hospital 2	urine	R	R	R
K. pneumoniae	KP-04	ERS7182759	KP-04	Hospital 2	catheter	R	R	R
K. pneumoniae	KP-05	ERS7182760	KP-05	Hospital 2	rectal swab	R	R	R
K. pneumoniae	KP-06	ERS7182761	KP-06	Hospital 2				
K. pneumoniae	KP-07	ERS7182762	KP-07	Hospital 2	blood culture	R	R	R
K. pneumoniae	KP-08	ERS7182763	KP-08	Hospital 2	rectal swab	R	R	R
K. pneumoniae	KP-09	ERS7182764	KP-09	Hospital 2		R		R
K. pneumoniae	KP-10	ERS7182765	KP-10	Hospital 1		R		R
P. aeruginosa	PA-01	ERS7182766	PA-01	Hospital 1				
P. aeruginosa	PA-02	ERS7182767	PA-02	Hospital 2				
P. aeruginosa	PA-03	ERS7182768	PA-03	Hospital 2				

piperacilline	piperacilline tazobactam	imipenem	meropenem	cefuroximaxetil	cefuroxim	cefotaxim	ceftriaxon	ceftazidim	amikacine	gentamycine	tobramycine	trimethoprim	trimethoprim/ sulfamethoxazol	nitrofurantoïne	moxifloxacin	ciprofloxacin	colistine	fosfomycine	cefoxitine	tigecycline	gentamicine
		R	R							R	R		R			R	S				
		R	R								R	R	R			R	S				
	R	S	S	R	R	R	R	R	S	R	R		R	S	S		R	S	R		R
	R	R	I	R	R	R	R	R	I	R	R		R	S	S		R	S	S		R
	R	R	I	R	R	R	R	R	S	R	R	R	R	S		R	S	S	R	S	
R	S	S			R	R	R		R	R	R	R			R		S				
	R	S	S	R	R	R	R	R	S	R	R	R	R	S		R	S	R	R		
R	S	S			R	R	R		R	R	R	R			S						
I	S	S			R	R	R		R	R	R	R			R		S				
	R	S	S	R	R	R	R	R	S	R	R	R	R	S		R	S	R	R	S	
	R	S	S	R	R	R	R	R	S	R	R	R	R	S		R	S	R	R		
	R	S	S	R	R	R	R	R		R	R	R	R	S		R	S	R	R		
	R	S	S	R	R	R	R	R	I	R	R		R	S	S		R	S	R		R
	R	S	S	R	R	R	R	R	S	R	R	R	R	S		R	S	R	R		
	R	S	S		R	R	R		R	R	R	R			S		R	S	R		R
R	S	S			R	R	R		R	R	R	R			R		S				
R	S	S			R	R	R		R	R	S	S			R		R				
I	S	S			R	R	R		R	R	R	R			R		S				
R	S	S	S			R	R	R		R	R	R	R	R			R	S	R	S	
	R	S	S	R	R	R	R	R	S	R	R	R	R	R		R	S	S	R	S	
	R	I	S	R	R	R	R	R	S	R	R	R	R	R		R	R	S	R	S	
	R	S	S	R	R	R	R	R		S	R		R	R	R		R	S	S		R
	R	S	S	R	R	R	R	R	S	R	R	R	R	S		R	S	S	R		
	R	S	S	R	R	R	R	R	S	R	R	R	R	S	R		S	S	R		R
	R	S	S	R	R	R	R	R	S	R	R	R	R	S		R	R	S	R	S	
	R	S	S	R	R	R	R	R	S	R	R	R	R	S		R	R	S	R		
	R	S	R	R			R	R		R	R	R	R			R	R	S	R		
	R	S	R	R			R	R		R	R	R	R			R	R	S	R		

Supplemental spreadsheet 2. Identified AMR genes for each bacterial isolate.

isolate	aac(3)-IId	aac(3)-IIe	aac(6')-Ib'	aac(6')-Ib-AKT	aac(6')-Ib-D181Y	aac(6')-Ib-cr5	aac(6')-Ib4	aadA1	aadA2	ant(2'')-Ia	ant(3'')-IIa	aph(3'')-Ib	aph(3')-IIb	aph(3')-Ia	aph(3')-VI	aph(6)-Id	armA	arr-3	blaACT-10	blaACT-12	blaACT-17	blaACT-23	blaACT-24	blaACT-46	blaACT-52
AB-01	.	.	100	.	.	.	.	100	.	.	100	98,31	.	100	.	100	100	.	.	.	.	.	.	.	.
AB-02	.	.	100	.	.	.	.	100	.	.	100	98,31	.	100	.	100	100	.	.	.	.	.	.	.	.
EC-01	.	100	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EC-02	.	100	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EC-03	.	100	.	.	100	.	.	100	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	.	.
EC-04	.	100	.	.	100	.	.	100	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	100	.
EC-05	.	100	.	.	100	.	.	100	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	.	.
EC-06	.	.	.	.	.	.	.	.	98,61	100	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.
EC-07	.	100	.	.	100	.	.	100	.	.	.	100	.	.	.	100	.	.	.	.	100	.	.	.	.
EC-08	.	100	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EC-09	.	100	.	.	100	.	.	100	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	.	.
EC-10	.	100	.	.	100	.	.	100	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	.	.
EC-11	.	.	.	.	100	.	.	.	100	.	.	.	.	.	.	100	.	.	.	100	.	.	.	.	.
EC-12	.	100	.	.	100	.	.	100	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	.	.
EC-13	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100
EC-14	.	100	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	.	.
EC-15	.	100	.	.	100	.	.	100	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	.	.
EC-16	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.
EC-17	.	100	.	.	100	.	.	100	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	.	.
EC-18	.	100	.	.	100	.	.	100	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	100	.
EC-19	.	100	.	.	100	.	.	100	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	100	.
EC-20	.	100	.	.	100	.	.	100	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	100	.
EC-21	.	.	.	100	.	.	.	.	100	100	.	.	.	.	.	.	.	.	84,47	.	.	.	.	.	.
EC-22	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	99,3	.	.	.	.
EC-23	.	.	.	.	.	.	.	.	98,61	100	.	.	.	.	.	.	.	.	.	.	.	100	.	.	.
EC-24	.	100	.	.	100	.	.	100	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	100	.
EC-25	.	100	.	.	100	.	.	100	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	100	.
EC-26	.	100	.	.	100	.	.	100	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	100	.
EC-27	.	100	.	.	100	.	.	.	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	100	.
EC-28	.	100	.	.	100	.	.	100	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	100	.
EC-29	.	100	.	.	100	.	.	100	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	100	.
EC-30	.	.	.	100	.	.	.	.	100	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EC-31	.	100	.	.	100	.	.	100	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	100	.
EC-32	.	100	.	.	100	.	.	100	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	100	.
KO-01	.	100	.	.	100	.	.	.	.	.	.	100	.	100	.	100	.	.	.	.	.	.	.	.	.
KO-02	.	100	.	.	100	.	.	.	.	.	.	100	.	100	.	100	.	.	.	.	.	.	.	.	.
KO-03	.	100	.	.	100	.	.	.	.	.	.	100	.	100	.	100	.	.	.	.	.	.	.	.	.
KO-04	.	100	.	.	100	.	.	.	.	.	.	100	.	100	.	100	.	.	.	.	.	.	.	.	.
KO-05	.	100	.	.	100	.	.	.	.	.	.	100	.	100	.	100	.	.	.	.	.	.	.	.	.
KO-06	.	100	.	.	100	.	.	.	.	.	.	100	.	100	.	100	.	.	.	.	.	.	.	.	.
KO-07	.	100	.	.	100	.	.	.	.	.	.	100	.	100	.	100	.	.	.	.	.	.	.	.	.
KP-01	.	100	.	.	100	.	.	100	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	.	.
KP-02	.	100	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KP-03	.	.	.	.	100	.	.	.	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	.	.
KP-04	.	100	.	.	100	.	.	.	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	.	.
KP-05	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KP-06	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KP-07	.	.	.	.	100	.	.	100	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	.	.
KP-08	.	100	.	.	100	.	.	.	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	.	.
KP-09	100	.	.	.	.	100	.	.	.	.	.	100	.	.	100	100	.	100	.	.	.	.	.	.	.
KP-10	100	.	.	.	.	100	.	.	.	.	.	100	.	.	100	100	.	100	.	.	.	.	.	.	.
PA-01	.	.	.	.	.	.	100	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
PA-02	.	.	.	.	.	.	100	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
PA-03	.	.	.	.	.	.	100	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.

Isolate	blaACT-55	blaACT-57	blaACT-64	blaACT-67	blaACT-8	blaADC-30	blaCTX-M-14	blaCTX-M-15	blaCTX-M-9	blaIMP-13	blaLAP-2	blaNDM-1	blaOXA-1	blaOXA-23	blaOXA-395	blaOXA-48	blaOXA-494	blaOXA-66	blaOXY1-4	blaPDC-374	blaPDC-55	blaSHV-1	blaSHV-106	blaSHV-12	blaSHV-187
AB-01	.	.	.	.	.	100	.	.	.	.	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.
AB-02	.	.	.	.	.	100	.	.	.	.	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.
EC-01	.	.	.	.	99,91	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-02	.	.	100	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-03	.	.	100	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-04	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-05	.	.	.	.	99,91	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-06	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	.
EC-07	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-08	100	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-09	.	.	.	100	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-10	.	.	.	.	99,91	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-11	.	.	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-12	.	.	.	.	99,91	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-13	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EC-14	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-15	.	.	.	.	99,91	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-16	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EC-17	.	.	.	.	99,91	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-18	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-19	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-20	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-21	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EC-22	.	.	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EC-23	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EC-24	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-25	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-26	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-27	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-28	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-29	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-30	.	100	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EC-31	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
EC-32	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
KO-01	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	100	.	.	.	.	.	.
KO-02	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	100	.	.	.	.	.	.
KO-03	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	100	.	.	.	.	.	.
KO-04	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	100	.	.	.	.	.	.
KO-05	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	100	.	.	.	.	.	.
KO-06	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	100	.	.	.	.	.	.
KO-07	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	100	.	.	.	.	.	.
KP-01	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	100	.	.	.
KP-02	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	100	.	.	.
KP-03	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.
KP-04	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	100	.	.
KP-05	.	.	.	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	100
KP-06	.	.	.	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	100
KP-07	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	100	.	.	.
KP-08	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	100	.	.
KP-09	.	.	.	.	.	.	.	.	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	.	.
KP-10	.	.	.	.	.	.	.	.	.	.	.	100	100	.	.	100	.	.	.	.	.	.	.	.	.
PA-01	.	.	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	100	.	.	.	.
PA-02	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	100	.	.	100	.	.	.	.	.
PA-03	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	100	.	.	100	.	.	.	.	.

Supplemental spreadsheet 2. (continued)

isolate	blaSHV-61	blaTEM-1	blaVEB-3	blaVIM-2	ble-MBL	catA1	catB3	catB7	catB8	crpP	dfrA1	dfrA14	dfrA16	floR	fosA	fosA-354827590	fosA2	fosA5	fosA6	fosA_gen	lnu(F)	mcr-4.3	mcr-9.1
AB-01	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.
AB-02	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EC-01	.	.	.	.	.	.	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	100
EC-02	.	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	100	100
EC-03	.	.	.	.	.	98,48	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	100	100
EC-04	.	100	.	.	.	.	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	.
EC-05	.	100	.	.	.	98,64	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	100
EC-06	.	.	.	.	.	100	.	.	.	.	.	.	100	.	100	.	.	.	.	.	.	.	99,88
EC-07	.	100	.	.	.	98,64	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	.
EC-08	.	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.
EC-09	.	.	.	.	.	98,64	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	.
EC-10	.	100	.	.	.	98,64	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	100
EC-11	.	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	100	.	.
EC-12	.	100	.	.	.	98,64	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	100
EC-13	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.
EC-14	.	.	.	.	.	.	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	.
EC-15	.	100	.	.	.	98,64	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	100
EC-16	.	.	100	.	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.
EC-17	.	100	.	.	.	98,64	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	100
EC-18	.	100	.	.	.	98,48	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	.
EC-19	.	100	.	.	.	98,48	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	.
EC-20	.	100	.	.	.	98,48	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	.
EC-21	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.
EC-22	.	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EC-23	.	.	.	.	.	.	.	.	.	.	.	.	100	.	100	.	.	.	.	.	.	.	99,88
EC-24	.	100	.	.	.	98,48	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	.
EC-25	.	100	.	.	.	98,48	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	.
EC-26	.	100	.	.	.	98,48	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	.
EC-27	.	100	.	.	.	98,48	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	.
EC-28	.	100	.	.	.	98,48	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	.
EC-29	.	100	.	.	.	98,48	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	.
EC-30	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	99,88
EC-31	.	100	.	.	.	98,48	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	.
EC-32	.	100	.	.	.	98,48	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	.
KO-01	.	100	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	99,05	.	.	.
KO-02	.	100	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	99,05	.	.	.
KO-03	.	100	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	99,05	.	.	.
KO-04	.	100	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	99,05	.	.	.
KO-05	.	100	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	99,05	.	.	.
KO-06	.	100	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	99,05	.	.	.
KO-07	.	100	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	99,05	.	.	.
KP-01	.	.	.	.	.	98,64	.	.	.	.	.	100	.	.	.	.	.	.	.	100	.	.	.
KP-02	.	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	100	.	.	.
KP-03	100	100	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	100	.	.	.
KP-04	.	100	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	100	.	.	.	.
KP-05	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	100	.	.	.	.
KP-06	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	100	.	.	.	.
KP-07	.	.	.	.	.	98,64	.	.	.	.	.	100	.	.	.	.	.	.	.	100	.	.	.
KP-08	.	100	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	100	.	.	.	.
KP-09	.	100	.	.	100	.	100	.	.	.	.	100	.	100	.	.	.	99,05	.	.	.	.	.
KP-10	.	100	.	.	100	.	100	.	.	.	.	100	.	100	.	.	.	99,05	.	.	.	.	.
PA-01	.	.	.	100	.	.	.	100	.	100	.	.	.	.	.	100	.	.	.	.	.	.	.
PA-02	.	.	.	.	.	.	.	100	.	100	.	.	.	.	.	100	.	.	.	.	.	.	.
PA-03	.	.	.	.	.	.	.	100	.	100	.	.	.	.	.	100	.	.	.	.	.	.	.

isolate	mph(A)	mph(E)	msr(E)	oqxA	oqxA10	oqxA5	oqxA6	oqxA9	oqxB12	oqxB17	oqxB19	oqxB20	oqxB32	oqxB4	oqxB5	oqxB9	qnrA1	qnrB1	qnrB32	qnrS1	qnrS2	sul1	sul2	tet(A)
AB-01	.	100	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	100	.
AB-02	.	100	100	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	100	.
EC-01	.	.	.	.	100	.	.	.	.	.	.	99,11	.	.	.	.	.	100	.	.	.	.	.	100
EC-02	.	.	.	.	100	.	.	.	.	.	.	.	.	100	.	.	.	100	.	.	.	.	.	100
EC-03	.	.	.	.	100	.	.	.	.	.	.	.	.	100	.	.	.	100	.	.	.	.	100	100
EC-04	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.	100	.
EC-05	.	.	.	.	100	.	.	.	.	.	.	99,11	.	.	.	.	.	100	.	.	.	.	100	100
EC-06	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	98,95	100	.	.	.	.	100	.	.
EC-07	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	98,95	.	100	.	.	.	.	100	100
EC-08	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	98,95	.	100	.	.	.	.	.	100
EC-09	.	.	.	.	99,91	.	.	.	.	.	.	.	.	.	100	.	.	100	.	.	.	.	100	100
EC-10	.	.	.	.	100	.	.	.	.	.	.	99,11	.	.	.	.	.	100	.	.	.	.	100	100
EC-11	84,33	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.	.	92,71	.	81,28	.	.	.	100
EC-12	.	.	.	.	100	.	.	.	.	.	.	99,11	.	.	.	.	.	100	.	.	.	.	100	100
EC-13	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	.
EC-14	.	.	.	.	99,91	.	.	.	.	.	.	.	.	.	100	.	.	100	.	.	.	.	.	100
EC-15	.	.	.	.	100	.	.	.	.	.	.	99,11	.	.	.	.	.	100	.	.	.	.	100	100
EC-16	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	98,95	.	.	.	.	100	100	.	.
EC-17	.	.	.	.	100	.	.	.	.	.	.	99,11	.	.	.	.	.	100	.	.	.	.	100	100
EC-18	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	98,95	.	100	.	.	.	.	100	100
EC-19	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	98,95	.	100	.	.	.	.	100	100
EC-20	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	98,95	.	100	.	.	.	.	100	83,33
EC-21	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.
EC-22	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	.	.	.	.
EC-23	.	.	.	.	99,91	.	.	.	.	.	.	.	.	.	100	.	100	.	.	.	.	100	.	100
EC-24	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	98,95	.	100	.	.	.	.	100	100
EC-25	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	98,95	.	100	.	.	.	.	100	100
EC-26	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	98,95	.	100	.	.	.	.	100	100
EC-27	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	98,95	.	100	.	.	.	.	100	100
EC-28	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	98,95	.	100	.	.	.	.	100	100
EC-29	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	98,95	.	100	.	.	.	.	100	83,33
EC-30	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	99,11	100	.	.	.	.	100	.	.
EC-31	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	98,95	.	100	.	.	.	.	100	100
EC-32	.	.	.	.	.	.	.	100	.	.	.	.	.	.	.	98,95	.	100	.	.	.	.	100	83,33
KO-01	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	99,21	.	100	.	.	.	.	100	100
KO-02	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	99,21	.	100	.	.	.	.	100	100
KO-03	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	99,21	.	100	.	.	.	.	100	100
KO-04	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	99,21	.	100	.	.	.	.	100	100
KO-05	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	99,21	.	100	.	.	.	.	100	100
KO-06	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	99,21	.	100	.	.	.	.	100	100
KO-07	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	99,21	.	100	.	.	.	.	100	100
KP-01	.	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	100	.	.	.	.	100	100
KP-02	.	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	100	.	.	.	.	.	100
KP-03	.	.	.	100	.	.	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	100	100
KP-04	.	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	100	.	.	.	.	100	100
KP-05	.	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	.	.	.	100	.	100	.	100
KP-06	.	.	.	.	100	.	.	.	100	.	.	.	.	.	.	.	.	.	.	100	.	100	.	100
KP-07	.	.	.	.	.	.	100	.	.	100	.	.	.	.	.	.	.	100	.	.	.	.	.	96
KP-08	.	.	.	.	.	100	.	.	.	.	100	.	.	.	.	.	.	.	.	.	.	.	100	.
KP-09	.	.	.	.	100	.	.	.	.	.	.	.	100	.	.	.	.	.	100	.	.	100	100	97,08
KP-10	.	.	.	.	100	.	.	.	.	.	.	.	100	.	.	.	.	.	100	.	.	100	100	97,08
PA-01	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	.	.
PA-02	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	.	.
PA-03	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	100	.	.

Supplemental spreadsheet 3. Zone of inhibition of conjugants compared to the acceptor strain J53. FEP = Cefepime, C = Clavulanic acid, CTX = Cefotaxime and CAZ = Ceftazidime

Ø (cm)	Donor	conjugant 3	conjugant 8	J53
FEP+C	22	30	30	30
FEP30	10	12	16	30
CTX+C	23	30	30	30
CTX30	10	10	10	30
CAZ+C	21	30	30	30
CAZ30	10	16	16	30

Supplemental spreadsheet 4. Metadata from SoM study isolates which were identified as carriers of the resistance gene island described in this study.

studie	ENA_run	fasta_file	invriesnummer	collection date	hospital
1	ERR1616165	SCP16-27	SCP16-27	27/11/2012	Hopsital 2
1	ERR1616174	SCP25-79	SCP25-79	13/12/2013	Hopsital 2
1	ERR1616179	SCK52-58	SCK52-58	14/01/2013	Hopsital 2
1	ERR1616180	SCK52-63	SCK52-63	18/02/2013	Hopsital 2
1	ERR1616181	SCK52-65	SCK52-65	25/02/2013	Hopsital 2
1	ERR1616182	SCK52-67	SCK52-67	21/03/2013	Hopsital 2
1	ERR1616183	SCK52-78	SCK52-78	04/12/2013	Hopsital 2
1	ERR1616185	SCK65-13	SCK65-13	07/03/2013	Hopsital 2
1	ERR1616187	SCK65-70	SCK65-70	16/08/2013	Hopsital 2
1	ERR1616205	SCP21-34	SCP21-34	12/03/2013	Hopsital 2
1	ERR1616206	SCP21-36	SCP21-36	12/03/2013	Hopsital 2
1	ERR1616207	SCP21-48	SCP21-48	24/04/2013	Hopsital 2
1	ERR1616208	SCP21-51	SCP21-51	01/05/2013	Hopsital 2
1	ERR1616215	SCP25-71	SCP25-71	13/12/2013	Hopsital 2
1	ERR1616376	SCK52-71	SCK52-71	04/07/2013	Hopsital 2
1	ERR1616441	SCP21-19	SCP21-19	14/02/2013	Hopsital 2
1	ERR1616442	SCP21-22	SCP21-22	14/02/2013	Hopsital 2
1	ERR1616443	SCP21-29	SCP21-29	12/03/2013	Hopsital 2
1	ERR1616643	SCK52-60	SCK52-60	04/02/2013	Hopsital 2
1	ERR1617663	SCK67-02	SCK67-02	07/10/2013	Hopsital 2
1	ERR1617690	SCP21-23	SCP21-23	28/02/2013	Hopsital 2
1	ERR1617691	SCP21-38	SCP21-38	12/03/2013	Hopsital 2
1	ERR1617707	SCP16-22	SCP16-22	21/11/2012	Hopsital 2
1	ERR1617708	SCP16-30	SCP16-30	27/11/2012	Hopsital 2
1	ERR1617711	SCP22-09	SCP22-09	23/10/2013	Hopsital 2
1	ERR1617980	SCK67-32	SCK67-32	26/09/2013	Hopsital 2
1	ERR1618464	SCP25-76	SCP25-76	13/12/2013	Hopsital 2
1	ERR1618466	SCP25-77	SCP25-77	13/12/2013	Hopsital 2
1	ERR1618475	SCP16-32	SCP16-32	20/12/2012	Hopsital 2
1	ERR1618477	SCP16-39	SCP16-39	02/01/2013	Hopsital 2
1	ERR1618478	SCP21-56	SCP21-56	16/08/2013	Hopsital 2
1	ERR1618480	SCP21-59	SCP21-59	16/08/2013	Hopsital 2
1	ERR1618968	SCK22-39	SCK22-39	07/11/2012	Hopsital 2
1	ERR1619328	SCP21-76	SCP21-76	12/09/2013	Hopsital 2
1	ERR1619331	SCP22-04	SCP22-04	15/10/2013	Hopsital 2
1	ERR1619498	SCP26-01	SCP26-01	20/12/2013	Hopsital 2
1	ERR1642495	SCP26-02	SCP26-02	20/12/2013	Hopsital 2
1	ERR1619523	SCP27-46-2	SCP27-46	06/02/2014	external hospital 1
1	ERR1616390	SCP11-14	SCP11-14	19/01/2012	external hospital 1
1	ERR1616647	SCK62-28	SCK62-28	09/01/2012	external hospital 1
1	ERR1616787	SCP24-07	SCP24-07	29/06/2012	external hospital 1
1	ERR1616019	SCK12-05	SCK12-05	05/09/2011	external hospital 1
1	ERR1616033	SCK33-02	SCK33-02	20/06/2012	external hospital 1
1	ERR1616223	SCP28-01	SCP28-01	16/09/2011	external hospital 1
1	ERR1617680	SCP14-43	SCP14-43	29/08/2012	external hospital 1
1	ERR1617688	SCP21-07	SCP21-07	03/09/2012	external hospital 1
1	ERR1616139	SCK56-04	SCK56-04	06/07/2013	external hospital 1
1	ERR1616785	SCP17-15	SCP17-15	15/07/2013	external hospital 1
1	ERR1616031	SCK31-70	SCK31-70	02/11/2012	external hospital 1

9

GENERAL DISCUSSION AND SUMMARY

The tip of iceberg? Implementing WGS for strain typing and outbreak surveillance

Bacteria are becoming more and more resistant to antibiotics due to misuse and overuse of these compounds[1]. As a consequence, common, currently treatable, infections may become untreatable leading to severe complications and death. Therefore, further increase in resistance should be avoided while on the other hand, spread of the resistant bacteria must be mitigated. For infection control for public health and infection prevention in the hospital environment, timely detection of transmission of resistant microorganisms is vital. Bacterial typing can determine if isolates are identical or not, which enables one to identify the spread of these specific isolates and putatively the source of this spread. The emergence of next generation sequencing (NGS) technologies, such as whole genome sequencing (WGS), makes a thorough analysis of DNA content possible. Therefore, the field of molecular epidemiology has rapidly shifted towards the use of WGS to infer phylogeny or relatedness among bacterial isolates based on their genome and thus identify if transmission occurred. However, with the introduction of newer technologies, validation and evaluation of the reproducibility is crucial. This also applies for WGS, where next to the wet-lab (e.g. culturing, DNA isolation and preparation of DNA for sequencing, called library preparation) also the dry-lab (data processing and interpretation) side needs to be thoroughly assessed. Efforts to assess the effect of DNA isolation[2] and sequencing library preparation[3] on WGS based typing have been conducted. The most important findings were the limited isolation of plasmid DNA by some DNA isolation kits and the underrepresentation in WGS data of high GC stretches in the bacterial genome when employing enzymatic tagmentation. As such, caveats for wet-lab procedures for WGS have been identified.

One of the strengths of outbreak surveillance would be the ability to know what strains circulate in the hospital, the region, or even worldwide by means of rapid decentralized data-sharing. An excellent example of successful data-sharing for (public health) decision making is the GISAID (Global Initiative on Sharing Avian Influenza Data, https://www.gisaid.org/) platform[4]. GISAID started for sharing data on the spread of influenza virus worldwide. Yet the platform obtained worldwide recognition and praise for its usefulness for the collection of SARS-CoV-2 genome data since the start of the SARS-CoV-2 pandemic. Although similar platforms for bacterial pathogen typing exist such as PATRIC[5] (https://www.patricbrc.org) and Pathogenwatch[6] (https://www.pathogen.watch), these platforms have yet to achieve the same level of scientific adoption as GISAID.

For successful data-sharing of genomic information, it is of importance to know if raw sequencing data processing may impact subsequent inferred phylogeny of the bacterial isolates. As laboratories rely either on in-house pipelines or commercial

software to analyse WGS data, the different workflows might not be interoperable for the goal of genotyping and outbreak surveillance. The reproducibility and interoperability of the dry-lab side of WGS for outbreak typing has thus far not been thoroughly evaluated. In **chapter 2** we investigated whether decentralized (local) WGS data are comparable between centres and which parameters are of importance, when a standardized procedure for wet-lab and dry-lab was used. The technical variation in genetic distance measured between these replicates was minor for cgMLST, wgMLST and SNP based methods and this variation in genetic distance was within ranges normally used for WGS based genotyping[7,8]. This minor technical variation in genetic distance among replicates is reassuring, as this indicates the feasibility to perform decentralized WGS typing using harmonized protocols. Although only nosocomial *Enterobacteriaceae* and *Enterococcus* species were tested, others have shown similar results for *Staphyloccus aureus*[9], *Listeria monocytogenes, Campylobacter jejuni* and *Salmonella Enterica*[10]. Harmonizing wet-lab protocols and bioinformatic workflows to analyse WGS data would therefore warrant successful decentralized strain genotyping.

Due to the plethora of different genome assemblers, it is unclear if these tools are interchangeable. In **chapter 2** we showed the lack of interoperability of bioinformatic workflows for generating genome assemblies for surveillance of outbreak clones. WGS datasets were subjected to different *de novo* assemblers to assess if the genetic dissimilarity between these replicates was introduced. The number of SNPs between these replicates surpassed thresholds employed for outbreak typing, demonstrating the unsuitability of using different bioinformatic workflows for phylogenetic inference. If a platform like GISAID would be used for tracking the spread of clones nationwide or worldwide, it is essential for WGS data processing, such as *de novo* assembly, to be performed in a uniform way because of the lack of interoperability of *de novo* assemblers. In addition, bioinformatic workflows also impact the outcome of strain genotyping by WGS, as demonstrated in **chapter 3**. In this chapter we demonstrated that different genotyping results are obtained by Dutch laboratories which all used a variety of different bioinformatic workflows even when the same datasets were analysed. Differences in genetic cut-offs used for outbreak typing between centres, were unable to explain differences in typing outcome. By using one single analysis on the genome assemblies obtained from the participating centres, we uncovered that unifying clonal-cut offs is a futile exercise as the typing outcome remains different between centres, which indicates once again the large bias genome assemblers introduce in WGS data. These results demonstrate the risk of using reported cut-offs in literature as described earlier[7,8]. Only if the exact same workflow is used, earlier reported cut-offs can be used for outbreak surveillance. Therefore, if decentralized surveillance of (resistant) bacterial strains by WGS is made readily accessible, it is paramount that bioinformatic data-processing methods are

harmonized and the most accurate bioinformatic workflow is identified. Without a thorough harmonisation it will never be possible to build broad knowledge on routes and mechanisms of spreading of resistant microorganisms and we will be unable to identify and predict outbreaks at all beyond local environments.

Deep down the iceberg: Identifying AMR genes and understanding the role of MGEs

Ever since the surge of WGS for pathogen genotyping took off and became dominant in the field of microbiology, there has been a similar increase in databases containing genetic markers of interest such as resistance genes[11–13], virulence genes[14,15] and mobile genetic elements like insertion sequences, transposons and ICE[16–20]. Often, before new AMR genes are identified and catalogued in databases, they have already disseminated all over the globe, as was the case with *mcr-1*[21–23]. Although (resistant) bacterial strain typing is readily performed worldwide, the role of how these (new) resistance genes spread between bacteria is often overlooked but is equally important. Therefore, it is crucial that i) resistance genes are readily identified, ii) the impact of new resistance genes is thoroughly assessed and iii) we understand the role of mobile genetic elements for antibiotic resistance dissemination.

Identifying (new) AMR genes

Identifying in which bacterial isolate a resistance gene resides within the microflora, is not always straightforward[24]. The commensal gut microbiota forms a natural reservoir of resistance genes, also known as resistome[25], and can hold resistance genes which can be transferred via horizontal gene transfer (HGT) to important pathogens. For instance, the ubiquitous nature of vancomycin resistance gene *vanC* and *vanD* in the human gut microbiome allows for exchange of these genes to *Enterococcus faecium*. Yet, metagenomically screening for these genes by PCR to detect *vanC* and *vanD* positive *Enterococcus faecium* is futile as these genes are mostly carried by the commensal flora. In parallel, in **chapter 5** we examined the importance of the presence of *mcr-9*. Although the prevalence of this *mcr-9* in *Enterobacter* species was alarmingly high, and gut carriage of *mcr-9* was certainly not negligible, we were unable to link this gene to actual colistin resistance. While this specific colistin resistance mechanism seemed of no relevance for hospital surveillance, it remains an important exercise to thoroughly assess the relevance of new resistance genes and how widespread they are. Without doing so, we are at risk of being unable to curb the spread of new resistance genes when they first appear as seen with, for instance, the identification of the already globally disseminated *mcr-1*[21–23].

Reliable identification of AMR genes by WGS is equally important for bacterial genotyping. Although currently, little genotype to phenotype inference is employed for clinical decision making, I foresee more widespread adoption of such practices

when bacterial typing or clinical metagenomics (identification of infectious agents straight from clinical samples using sequencing technologies)[26] becomes more readily accessible. Thoroughly evaluating the bioinformatic workflow for identification of AMR genes is required and suitable benchmark WGS datasets have been compiled, as described in **chapter 4**[27]. In **chapter 3** we identified issues for AMR gene identification from WGS data due to bioinformatic data processing. Some participants of a WGS data ring trial failed to identify ESBL genes for some bacterial isolates. Manual assessment of the genome assemblies from these centres confirmed the absence of these AMR genes. All these centres had one data processing step in common. For the genome assemblies, all contigs smaller than 1000bp were left out of the final assembly, which is a commonly employed step to get rid of small contaminations in the dataset. Currently, using WGS in a clinical setting, bacterial isolates are sequenced after they are cultured in the lab and antimicrobial susceptibility testing already has been performed. Therefore, clinicians already know if certain isolates have, for instance, an ESBL- or carbapenemase-producing phenotype. When NGS technologies are employed for clinical metagenomics, missing these important AMR genes could be of serious consequence, and treatment failure due to treatment with antibiotics to which the pathogen is resistant, might ensue.

The role of MGEs

Finally, the role of MGEs for disseminating AMR genes must not be underestimated. Both plasmids and transposons have been linked to outbreaks, but such outbreaks are much harder to identify[28–31]. Essentially all typing methods with the exception of long-read WGS are unable to elucidate the undercover nature of mobile genetic elements carrying AMR genes. Due to the promiscuous nature of some plasmids, AMR genes encoded thereupon can end up in various different bacterial hosts, spanning different genera. In **chapter 6** this indiscriminate nature of *mcr-1* plasmids was observed. Here, we found a limited number of different plasmid backbones encoding for *mcr-1,* yet the *E. coli* carrying these plasmids were very diverse. Even though these isolates were collected from retail chicken meat, highly similar plasmids were also found in clinical human isolates, indicating the widespread nature of these plasmids. Plasmids encoding the carbapenemase $bla_{\text{OXA-48}}$ is another example of a highly successful plasmid which resides in a plethora of different bacterial hosts[32]. Some lineages of bacteria are well known to be associated with specific environments, a well-known example is *E. coli* ST131 present in humans[33]. These bacterial strains are well adapted through the acquisition of specific genes, making them more suitable for these environments. Yet, plasmids are generally indifferent to the bacterial host they reside in. Essentially only the compatibility with other plasmids or restriction/modification systems and CRISPR/Cas dictate whether a plasmid can thrive in a new bacterial host[34–36]. In **chapter 7** we showed this promiscuous nature of a single *mcr-4*

plasmid in animals and humans throughout the Netherlands, demonstrating that plasmids can also spread between domains. Therefore, surveillance of resistant bacteria should extend to plasmids next to bacterial isolates, as these are also significant contributors in the spread of antibiotic resistance globally.

Finally, the role of transposons for the dissemination of antibiotic resistance should also not be overlooked. The relative contribution for AMR spread of these MGEs is poorly understood and has been studied only in recent years[37]. The nested doll behaviour of these elements whereby they move from and to different plasmids and into the bacterial chromosome highlights the difficult nature to properly identify and track these elements. As described in **Chapter 8** the appearance of what seemed unrelated microbes who shared similar phenotypic resistance profiles was described. Due to the shared phenotype, the possibility of a shared AMR plasmid was considered, and thus long read sequencing was conducted. A large (46kb) gene island, encoding for 14 AMR genes, was identified in different bacterial species, encoded on plasmids but also in the chromosome. Many of these AMR genes were flanked by *IS26* insertion sequences. These *IS26* are often associated with AMR genes[38] and they preferentially insert onto other *IS26* rich regions in the genome[39], generating gene cassettes comprised of AMR genes flanked by *IS26*. The transfer of this gene island can be partly explained by the exchange of genetic material via conjugative plasmids to other members in this taxonomic order[40]. Finally, 1806 WGS datasets from ESBL producing *Enterobacterales* from hospitalized patients in the Netherlands 2011 to 2014 was screened for this resistance gene island, to assess if this gene island was already present before 2019 in The Netherlands. An additional 50 isolates were identified encoding this gene island, of which 37 originated from one of the hospitals of the study of **chapter 8**, implying that this resistance gene island has been present for at least 7 years prior in this hospital. The spread of transposons or these gene islands is often overlooked as they are difficult to identify without generating complete genomes. Only when rare and alarming resistance mechanisms (e.g. carbapenemases) are involved, have these elements been identified as the causative agents for their dissemination[28,29,41]. The long-term presence of this resistance gene island highlights the need for surveillance of these mobile genetic elements to mitigate the spread of AMR to new bacterial strains in a nosocomial setting.

Future perspectives & concluding remarks

Implementation WGS for surveillance and typing

As shown in the work presented in this thesis, multi-centre implementation of WGS for surveillance and typing of antibiotic resistant bacteria is feasible. However, some knowledge gaps are still remaining: I) what is the influence of the chemistry used to produce sequencing libraries? II) Do the different sequencing platforms lead to different or inconsistent results? Other work already ruled out the influence of DNA

isolation methodology on WGS results such as AMR detection and cluster identification[2]. With one important caveat: Some isolation procedures do not isolate plasmid DNA efficiently. As AMR genes often are located on plasmids, poor isolation of plasmids DNA can result in poor identification of AMR genes in the bacterial genome by WGS. The work presented in this thesis, was based on WGS data generated by Illumina MiSeq sequencing and therefore it is not evident that the results demonstrated here are also applicable to other sequencing platforms.

With the shift towards newer sequencing technologies, such as Oxford Nanopore sequencing this exercise to validate the reproducibility of WGS is also required. These new methodologies require different or modified algorithms to process raw signal data to raw sequence data by a process called basecalling, for which several different algorithms exist[42]. Also, established *de novo* assembly algorithms, which utilizes highly accurate sequences are unsuitable to the noisy long reads (>10kb) generated from Nanopore technologies. Based on the outcome of the study in described in **chapter 3,** where the incompatibility for outbreak typing between different short-read WGS outbreak pipelines, was demonstrated, the same incompatibility should be expected when a mix of short or long-read sequenced genomes is used in a single dataset for outbreak typing. As the field of WGS and the associated bioinformatic algorithms keep on improving and changing, harmonization of workflows to achieve interoperability and comparability for bacterial typing is only possible if one accepts the fact that their workflow may have a short shelf life and needs to be revisited and maintained regularly.

Genotyping (new) AMR genes and implications for the clinic

WGS allows for identification of AMR genes. Inference of phenotypic resistance from genotyping of AMR genes using WGS is still in its infancy but has tremendous potential. In theory, one could directly sequence isolates or even metagenomes (all genetic material in a sample, including the microbe of interest) to identify the causative agent and infer its expected phenotypic resistance profile[13,26]. The translation from genotype to phenotype for AMR took the turn into the field of "big-data" when WGS was introduced. Culturing of each individual micro-organism and performance of broth dilution tests to assess resistance of each isolate is no longer necessary, but genomic data can disclose all encoded resistance mechanisms.

However, we should not expect to completely solve genotype-to-phenotype for AMR determination with current standings of NGS alone. More complex matters are at hand than simply identification of an AMR gene which directly translate into a resistance phenotype. The regulation of gene transcription is certainly of impact on specific resistance traits, as seen in expression of multi-drug efflux pumps normally repressed in *Pseudomonas aeruginosa*[43], porin loss[44] or *ampC* de-repression in other *Enterobacteriaceae* species[45]. And finally, modulation of gene expression mediated

by methylation is a known driver for modulating a resistance phenotype in bacteria[46], but with current sequencing technologies these cannot be elucidated. For now, the European committee on antimicrobial susceptibility testing (EUCAST) considers WGS based antimicrobial susceptibility testing (AST)[47] unsatisfactory and further work is required if WGS based AST can be implemented in the clinic. Finally, WGS for strain typing has established itself as a trustworthy tool in infection prevention and surveillance for the spread of nosocomial bacteria. Yet, WGS is used during or after outbreaks occurred and mobile genetic elements are almost never considered.

In conclusion, WGS will be the key to the further reduction of spread of resistant micro-organisms. This technology will continue to improve and its' usage will be more commonly adopted. With the introduction of newer sequencing technologies, the diverse nature of MGEs and their accompanying AMR genes have been exposed and it has become clear surveillance and typing for these elements required.

References

1. Lee Ventola, C. The Antibiotic Resistance Crisis Part 1: Causes and Threats. **40**, (2015).
2. Nouws, S. *et al.* Impact of DNA extraction on whole genome sequencing analysis for characterization and relatedness of Shiga toxin-producing Escherichia coli isolates. *Sci. Rep.* (2020) doi:10.1038/s41598-020-71207-3.
3. Seth-Smith, H. M. B. *et al.* Evaluation of rapid library preparation protocols for whole genome sequencing based outbreak investigation. *Front. Public Heal.* **7**, (2019).
4. Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance* vol. 22 (2017).
5. Gillespie, J. J. *et al.* Patric: The comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. *Infect. Immun.* **79**, (2011).
6. Argimón, S. *et al.* A global resource for genomic predictions of antimicrobial resistance and surveillance of Salmonella Typhi at pathogenwatch. *Nat. Commun.* **12**, (2021).
7. Schürch, A. C., Arredondo-Alonso, S., Willems, R. J. L. & Goering, R. V. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene–based approaches. *Clinical Microbiology and Infection* (2018) doi:10.1016/j.cmi.2017.12.016.
8. Kluytmans-Van Den Bergh, M. F. Q. *et al.* Whole-genome multilocus sequence typing of extended-spectrum-beta-lactamase-producing enterobacteriaceae. *J. Clin. Microbiol.* (2016) doi:10.1128/JCM.01648-16.
9. Mellmann, A. *et al.* High interlaboratory reproducibility and accuracy of next-generation-sequencing-based bacterial genotyping in a ring trial. *J. Clin. Microbiol.* **55**, (2017).
10. Uelze, L. *et al.* German-Wide Interlaboratory Study Compares Consistency, Accuracy and Reproducibility of Whole-Genome Short Read Sequencing. *Front. Microbiol.* **11**, (2020).
11. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* (2012) doi:10.1093/jac/dks261.
12. Alcock, B. P. *et al.* CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **48**, (2020).
13. Bortolaia, V. *et al.* ResFinder 4.0 for predictions of phenotypes from genotypes. *J. Antimicrob. Chemother.* **75**, (2020).
14. 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 Res.* (2016) doi:10.1093/nar/gkv1239.
15. Liu, B., Zheng, D., Jin, Q., Chen, L. & Yang, J. VFDB 2019: A comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res.* **47**, (2019).
16. Siguier, P., Perochon, J., Lestrade, L., Mahillon, J. & Chandler, M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* **34**, (2006).
17. Kichenaradja, P., Siguier, P., Pérochon, J. & Chandler, M. ISbrowser: An extension of ISfinder for visualizing insertion sequences in prokaryotic genomes. *Nucleic Acids Res.* **38**, (2009).
18. Tansirichaiya, S., Rahman, M. A. & Roberts, A. P. The Transposon Registry. *Mobile DNA* vol. 10 (2019).
19. Bi, D. *et al.* ICEberg: A web-based resource for integrative and conjugative elements found in Bacteria. *Nucleic Acids Res.* **40**, (2012).
20. Liu, M. *et al.* ICEberg 2.0: An updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res.* **47**, (2019).
21. Webb, H. E. *et al.* Dissemination of the mcr-1 colistin resistance gene. *The Lancet Infectious Diseases* vol. 16 (2016).
22. Zhi, C., Lv, L., Yu, L. F., Doi, Y. & Liu, J. H. Dissemination of the mcr-1 colistin resistance gene. *The Lancet Infectious Diseases* (2016) doi:10.1016/S1473-3099(16)00063-3.
23. Liu, Y.-Y. *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.* **16**, 161–168 (2017).
24. Flipse, J. *et al.* Appearance of vanD-positive Enterococcus faecium in a tertiary hospital in the Netherlands: prevalence of vanC and vanD in hospitalized patients. *Sci. Rep.* **9**, (2019).
25. Von Wintersdorff, C. J. H. *et al.* Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer. *Front. Microbiol.* **7**, 1–10 (2016).
26. Forbes, J. D., Knox, N. C., Peterson, C. L. & Reimer, A. R. Highlighting Clinical Metagenomics for Enhanced Diagnostic Decision-making: A Step Towards Wider Implementation. *Computational and Structural Biotechnology Journal* vol. 16 (2018).
27. Raphenya, A. R. *et al.* Datasets for benchmarking antimicrobial resistance genes in bacterial metagenomic and whole genome sequencing. 5–10 (2022) doi:10.1038/s41597-022-01463-7.

28. Sheppard, A. E. *et al.* Nested Russian doll-like genetic mobility drives rapid dissemination of the carbapenem resistance gene blakpc. *Antimicrob. Agents Chemother.* **60**, 3767–3778 (2016).
29. Wozniak, A. *et al.* A multispecies outbreak of carbapenem-resistant bacteria harboring the bla KPC gene in a non-classical transposon element. *BMC Microbiol.* **21**, (2021).
30. Adhiratha Boonyasiri, Ashleigh C. Myall, Yu Wan, Frances Bolt, Alice Ledda, Siddharth Mookerjee, Andrea Y. Weiße, Jane F. Turton, Hala Abbas, Ruta Prakapaite, Akshay Sabnis, Alireza Abdolrasouli, Kenny Malpartida-Cardenas, Luca Miglietta, Hugo Donaldson, F. D. Integrated patient network and genomic plasmid analysis reveal a regional, multi-species outbreak of carbapenemase-producing Enterobacterales carrying both blaIMP and mcr-9 genes. *medRvix* (2021).
31. León-Sampedro, R. *et al.* Pervasive transmission of a carbapenem resistance plasmid in the gut microbiota of hospitalized patients. *Nat. Microbiol.* **6**, (2021).
32. Poirel, L., Potron, A. & Nordmann, P. OXA-48-like carbapenemases: The phantom menace. *J. Antimicrob. Chemother.* **67**, 1597–1606 (2012).
33. Woodford, N., Turton, J. F. & Livermore, D. M. *Multiresistant Gram-negative bacteria: The role of high-risk clones in the dissemination of antibiotic resistance. FEMS Microbiology Reviews* vol. 35 736–755 (Blackwell Publishing Ltd, 2011).
34. Price, V. J., Huo, W., Sharifi, A. & Palmer, K. L. CRISPR-Cas and Restriction-Modification Act Additively against Conjugative Antibiotic Resistance Plasmid Transfer in Enterococcus faecalis. *mSphere* **1**, (2016).
35. Bhaya, D., Davison, M. & Barrangou, R. CRISPR-cas systems in bacteria and archaea: Versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.* **45**, (2011).
36. Couturier, M., Bex, F., Bergquist, P. L. & Maas, W. K. Identification and classification of bacterial plasmids. *Microbiological Reviews* vol. 52 (1988).
37. Arredondo-Alonso, S., Top, J., Corander, J., Willems, R. J. L. & Schürch, A. C. Mode and dynamics of vanA-type vancomycin resistance dissemination in Dutch hospitals. *Genome Med.* **13**, (2021).
38. Harmer, C. J. & Hall, R. M. IS 26 -Mediated Formation of Transposons Carrying Antibiotic Resistance Genes. *mSphere* **1**, (2016).
39. Harmer, C. J., Moran, R. A. & Hall, R. M. Movement of IS26-Associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS26 and preferentially inserts adjacent to another IS26. *MBio* **5**, (2014).
40. Redondo-Salvo, S. *et al.* Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids. *Nat. Commun.* **11**, (2020).
41. Sheppard, A. E. *et al.* Tetyper: A bioinformatic pipeline for classifying variation and genetic contexts of transposable elements from short-read whole-genome sequencing data. *Microb. Genomics* **4**, (2018).
42. Wick, R. R., Judd, L. M. & Holt, K. E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* **20**, (2019).
43. Alav, I., Sutton, J. M. & Rahman, K. M. Role of bacterial efflux pumps in biofilm formation. *J. Antimicrob. Chemother.* **73**, (2018).
44. Thomson, J. M. & Bonomo, R. A. The threat of antibiotic resistance in Gram-negative pathogenic bacteria: β-lactams in peril! *Current Opinion in Microbiology* vol. 8 (2005).
45. Tamma, P. D., Doi, Y., Bonomo, R. A., Johnson, J. K. & Simner, P. J. A Primer on AmpC β-Lactamases: Necessary Knowledge for an Increasingly Multidrug-resistant World. *Clinical Infectious Diseases* vol. 69 (2019).
46. Ghosh, D., Veeraraghavan, B., Elangovan, R. & Vivekanandan, P. Antibiotic resistance and epigenetics: More to it than meets the eye. *Antimicrob. Agents Chemother.* **64**, (2020).
47. Ellington, M. J. *et al.* The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. *Clinical Microbiology and Infection* vol. 23 (2017).

10

NEDERLANDSE SAMENVATTING

Het topje van de ijsberg? Implementatie van WGS voor stamtypering en uitbraaksurveillance

Bacteriën worden steeds resistenter tegen antibiotica als gevolg van verkeerd gebruik en overmatig gebruik van deze middelen. Als gevolg daarvan kunnen gewone, momenteel behandelbare infecties onbehandelbaar worden, wat kan leiden tot ernstige complicaties en zelfs overlijden. Daarom is het van belang dat een verdere toename van antibiotica resistentie wordt voorkomen. Een mogelijkheid hierbij is het tegengaan van verspreiding van resistente bacteriën. Voor infectiebestrijding ten behoeve van de volksgezondheid en infectiepreventie in de ziekenhuisomgeving is tijdige detectie van de overdracht van resistente micro-organismen hierbij van vitaal belang. Bacteriële typering kan bepalen of isolaten al dan niet identiek zijn, wat het mogelijk maakt de verspreiding van deze specifieke isolaten tussen patiënten en mogelijk de bron van deze verspreiding te identificeren. De opkomst van "next generation sequencing" (NGS) technologieën, zoals "whole genome sequencing" (WGS), maakt een grondige analyse van de DNA-volgorde van de bacterie mogelijk. Hierdoor is het veld van de moleculaire epidemiologie verschoven naar het gebruik van WGS om de fylogenie of verwantschap tussen bacteriële isolaten op basis van hun genoom te bepalen en zo vast te stellen of er transmissie heeft plaatsgevonden. Met de introductie van deze geavanceerdere technologieën is validatie en evaluatie van de reproduceerbaarheid echter van cruciaal belang. Dit geldt ook voor WGS, waarbij naast het "wet-lab" (bv. kweken, DNA-isolatie en preparatie van DNA voor sequencering) ook het "dry-lab" (dataverwerking en -interpretatie) nauwkeurig moet worden geëvalueerd.

Een van de sterke punten van uitbraaksurveillance is de mogelijkheid om te weten welke isolaten in het ziekenhuis, de regio of zelfs wereldwijd circuleren door middel van snelle gegevensuitwisseling. Een uitstekend voorbeeld van het succesvol delen van data ten behoeve van besluitvorming (op het gebied van de volksgezondheid) is het GISAID-platform (Global Initiative on Sharing Avian Influenza Data, https://www.gisaid.org/). GISAID begon met het delen van data over de verspreiding van het influenzavirus wereldwijd. Toch kreeg het platform wereldwijde erkenning en lof voor zijn nut voor het verzamelen van SARS-CoV-2-genoomsequenties tijdens de SARS-CoV-2-pandemie. Hoewel er soortgelijke platformen voor de typering van bacteriële pathogenen bestaan, zoals PATRIC (https://www.patricbrc.org) en Pathogenwatch (https://www.pathogen.watch), hebben deze platformen nog niet hetzelfde niveau van implementatie bereikt als GISAID.

Voor een succesvolle uitwisseling van genetische informatie is het van belang te weten of de bioinformatische dataverwerking van ruwe sequentiedata van invloed kan zijn op de berekende fylogenie van de bacteriële isolaten. Aangezien laboratoria voor de analyse van WGS-gegevens ofwel eigen analyse pijplijnen gebruiken, dan wel commerciële software implementeren, zijn de verschillende werkwijzen misschien niet uitwisselbaar voor het doel van genotypering en uitbraaksurveillance. De reproduceerbaarheid en uitwisselbaarheid van de "dry-lab" kant van WGS voor uitbraaktypering is tot nu toe niet grondig geëvalueerd. In **hoofdstuk 2** hebben we onderzocht of gedecentraliseerde (lokaal verkregen) WGS data vergelijkbaar zijn tussen centra en welke parameters hierbij van belang zijn, wanneer een gestandaardiseerde procedure voor "wet-lab" en "dry-lab" werd gebruikt. De technische variatie in genetische afstand gemeten tussen deze replica's was klein voor cgMLST, wgMLST en SNP gebaseerde methoden en deze variatie in genetische afstand lag binnen marges die normaal gebruikt worden voor WGS gebaseerde genotypering. Deze geringe technische variatie in genetische afstand tussen de replica's is geruststellend, aangezien dit wijst op de haalbaarheid van gedecentraliseerde WGS-typering met geharmoniseerde protocollen. Hoewel alleen nosocomiale *Enterobacteriaceae* en *Enterococcus*-soorten werden getest, hebben anderen soortgelijke resultaten aangetoond voor *Staphyloccus aureus*, *Listeria monocytogenes*, *Campylobacter jejuni* en *Salmonella enterica*. Harmonisatie van wet-lab protocollen en bioinformatische werkwijzen voor de analyse van WGS-gegevens lijken hiermee toereikend genoeg te zijn voor succesvolle gedecentraliseerde stamgenotypering datavergelijking.

Door de overvloed aan verschillende genoom-assembler algoritmes is het onduidelijk of deze tools onderling uitwisselbaar zijn. In **hoofdstuk 2** hebben we het gebrek aan uitwisselbaarheid aangetoond van bioinformatische werkwijzen voor het genereren van genoom-assemblies voor surveillance van uitbraken. WGS-datasets werden onderworpen aan verschillende *de novo* assemblers om te beoordelen of er genetische dissimilariteit tussen deze replica's was geïntroduceerd. Het aantal SNP's tussen deze replica's overschreed de afkapwaarden die worden gebruikt voor uitbraaktypering, wat aantoont dat het ongeschikt is om verschillende bioinformatische werkwijzen door elkaar te gebruiken voor fylogenetische analyses. Als een platform als GISAID gebruikt zou worden voor het traceren van de verspreiding van klonen over het hele land of de hele wereld, is het essentieel dat WGS-gegevensverwerking, zoals *de novo* assembly, op een uniforme manier wordt uitgevoerd vanwege het gebrek aan uitwisselbaarheid van *de novo* assemblers. Daarnaast hebben bioinformatische werkwijzen ook invloed op het resultaat van stamgenotypering door WGS, zoals aangetoond in **hoofdstuk 3**. In dit hoofdstuk hebben we laten zien dat Nederlandse laboratoria bij analyse van exact dezelfde datasets verschillende genotypering resultaten kregen als gevolg van het gebruik van verschillende bioinformati-

sche werkwijzen. Ondanks verschillen in gebruikte genetische afkapwaarden tussen centra bleken onvoldoende om verschillen tussen typeringsresultaten te verklaren. Door gebruik te maken van één enkele analyse op de genoom-assemblies van de deelnemende centra, hebben we aangetoond dat het gebruiken van uniforme klonale afkapwaarden zinloos is, aangezien de typeringsresultaten tussen centra blijven verschillen, wat eens te meer wijst op de grote bias die genoom assemblers in WGS-gegevens introduceren. Deze resultaten tonen het risico aan van het gebruik van in de literatuur gerapporteerde afkapwaarden. Alleen als exact dezelfde werkwijze wordt gebruikt, kunnen eerder gerapporteerde afkapwaarden voor uitbraaksurveillance worden gebruikt. Als gedecentraliseerde surveillance van (resistente) bacteriestammen door WGS toegankelijk wordt voor een wijder publiek, is het van het grootste belang dat naast de gestandaardiseerde wet-lab procedures ook de bioinformatische werkwijzen worden geharmoniseerd en de meest nauwkeurige bioinformatische werkwijze wordt bepaald. Zonder een grondige harmonisatie zal het nooit mogelijk zijn een brede kennis op te bouwen over de verspreidingsroutes en -mechanismen van resistente micro-organismen en zal men helemaal niet in staat zijn uitbraken buiten de lokale omgeving te identificeren en te voorspellen.

Diep in de ijsberg: Identificatie van antibioticaresistentie (ABR) genen en inzicht in de rol van MGE's

WGS voor de genotypering van pathogenen is in razendsnel tempo dominant geworden in de microbiologie. Daarmee is er een vergelijkbare toename geweest van databases met genetische markers van belang, zoals resistentiegenen, virulentiegenen en mobiele genetische elementen zoals insertiesequenties, transposons en integreerbare & conjugeerbare elementen (ICE). Vaak zijn nieuwe ABR-genen, al over de hele wereld verspreid, voordat ze geïdentificeerd en in databases gecatalogiseerd zijn, zoals het geval was met *mcr-1*. Hoewel de typering van (resistente) bacteriestammen wereldwijd steeds vaker wordt uitgevoerd, wordt vaak over het hoofd gezien hoe deze (nieuwe) resistentiegenen zich tussen bacteriën verspreiden. Daarom is het van cruciaal belang dat i) resistentiegenen betrouwbaar kunnen worden geïdentificeerd, ii) het effect van nieuwe resistentiegenen grondig wordt geëvalueerd en iii) we de rol van mobiele genetische elementen bij de verspreiding van antibioticaresistentie begrijpen.

Identificatie van (nieuwe) ABR-genen

Het is niet eenvoudig om te bepalen welk bacterieel isolaat van de microflora een resistentiegen bevat. De commensale darmmicrobiota vormt een natuurlijk reservoir van resistentiegenen, resistoom genoemd, en kan resistentiegenen bevatten in commensale bacterien die via horizontale genoverdracht op belangrijke pathogenen kunnen worden overgedragen. Het veelvoorkomende vancomycine resistentiegen

vanC en *vanD* in het humane darmmicrobioom maakt bijvoorbeeld de uitwisseling van deze genen naar *Enterococcus faecium* mogelijk. Daarom is screening op deze genen door PCR om *vanC* en *vanD* positieve *Enterococcus faecium* op te sporen zinloos, omdat deze genen meestal door de commensale bacteriën worden gedragen. Parallel aan dit voorbeeld onderzochten we in **hoofdstuk 5** het belang van de aanwezigheid van *mcr-9*, een nieuw mobiel colistine resistentiegen. Hoewel de prevalentie van deze *mcr-9* in *Enterobacter* soorten alarmerend hoog was, en dragerschap van *mcr-9* zeker niet verwaarloosbaar was, waren we niet in staat om dit gen te koppelen aan daadwerkelijke fenotypische colistine resistentie. Hoewel dit specifieke mechanisme van colistine resistentie niet van belang leek voor de patiëntenzorg, blijft het een belangrijke taak voor de medische microbiologie om de relevantie van nieuwe resistentiegenen en de mate waarin deze verspreid zijn grondig te evalueren. Doen wij dit niet, dan lopen we het risico dat we niet in staat zijn de verspreiding van nieuwe resistentiegenen te voorkomen wanneer ze voor het eerst opduiken, zoals we bijvoorbeeld hebben gezien bij de identificatie van het reeds wereldwijd verspreidde *mcr-1*.

Betrouwbare identificatie van ABR-genen door WGS is eveneens belangrijk voor genotyperen van bacteriën. Hoewel momenteel nog weinig gebruik wordt gemaakt van vertaling van genotype naar fenotype voor klinische besluitvorming, is de verwachting dat dergelijke praktijken op grotere schaal zullen worden toegepast wanneer bacterietypering of klinische metagenomics (identificatie van infectieuze agentia rechtstreeks uit klinische monsters met behulp van sequenering technologieën) meer toegankelijk wordt. Een grondige evaluatie van de bioinformatische werkwijze voor de identificatie van ABR-genen is nodig en er zijn geschikte WGS-benchmark datasets samengesteld, zoals beschreven in **hoofdstuk 4**. In **hoofdstuk 3** hebben we problemen geïdentificeerd voor ABR-gen identificatie uit WGS-data als gevolg van bioinformatische werkwijzen. Sommige deelnemers aan een ringonderzoek met WGS-data slaagden er niet in ESBL-genen te identificeren in sommige bacteriële isolaten. Herbeoordeling van de genoom-assemblies van deze centra bevestigde de afwezigheid van deze ABR-genen. Al deze centra hadden één stap in de gegevensverwerking gemeen. Voor de genoom-assemblies werden alle contigs kleiner dan 1000bp uit de assemblies gelaten, wat een veel gebruikte stap is om kleine contaminaties in de dataset te verwijderen. Momenteel worden bij gebruik van WGS in een klinische omgeving de bacteriële isolaten gesequeneerd nadat zij in het lab zijn gekweekt en de antimicrobiële gevoeligheidstests reeds zijn uitgevoerd. Daarom weten de clinici al of bepaalde isolaten bijvoorbeeld een ESBL- of carbapenemase-producerend fenotype hebben. Wanneer NGS-technologieën worden gebruikt voor klinische metagenomics, zou het missen van deze belangrijke ABR-genen ernstige gevolgen kunnen hebben, en kan behandeling, met antibiotica waartegen de ziekteverwekker resistent is, falen.

De rol van MGE's

Ten slotte mag de rol van MGE's voor de verspreiding van ABR-genen niet worden onderschat. Zowel plasmiden als transposons zijn in verband gebracht met uitbraken van ABR-genen, maar dergelijke uitbraken zijn veel moeilijker te identificeren. In de praktijk zijn er geen typeringsmethoden, met uitzondering van "long-read" WGS, in staat om de verborgen aard van mobiele genetische elementen die ABR-genen dragen, goed te identificeren. Sommige ABR-plasmiden kunnen terechtkomen in verschillende bacteriële gastheren, verspreid over verschillende genera, afkomstig uit zowel mens, dier of de omgeving. In **hoofdstuk 6** werd dit gedrag van *mcr-1* plasmiden waargenomen. Hier vonden we een beperkt aantal verschillende plasmide "backbones" die codeerden voor *mcr-1*, maar de *E. coli* die deze plasmiden droegen waren zeer divers. Hoewel deze isolaten afkomstig waren van kippenvlees uit de supermarkt, werden sterk vergelijkbare plasmiden ook aangetroffen in klinische humane isolaten, wat wijst op de wijdverspreide aard van deze plasmiden. Plasmiden die coderen voor het carbapenemase $bla_{\text{OXA-48}}$ zijn een ander voorbeeld van een zeer succesvol plasmide dat in een groot aantal verschillende bacteriële gastheren voorkomt. Van sommige bacteriestammen is bekend dat zij geassocieerd zijn met specifieke omgevingen; een bekend voorbeeld is *E. coli* sequentie type 131 die in de mens voorkomt. Deze bacteriestammen zijn goed aangepast door de verwerving van specifieke genen, waardoor zij geschikter zijn voor deze omgevingen. Plasmiden kunnen echter tussen verschillende bacteriële gastheren verspreiden middels conjugatie. Alleen de compatibiliteit met andere plasmiden of restrictie-/modificatiesystemen en CRISPR/Cas bepalen of een plasmide kan gedijen in een nieuwe bacteriële gastheer. In **hoofdstuk 7** is het *mcr-4* plasmide in dieren en mensen in Nederland uitgezocht. Hier toonde we aan dat plasmiden zich ook tussen domeinen waarschijnlijk kunnen verspreiden. Daarom is het aan te bevelen dat de surveillance van resistente bacteriële isolaten wordt uitgebreid met de resistentie plasmiden aangezien deze ook een potentieel belangrijke bijdrage leveren aan de wereldwijde verspreiding van antibioticaresistentie.

Ten slotte mag ook de rol van transposons voor de verspreiding van antibioticaresistentie niet worden gemist. De relatieve bijdrage van verspreiding van ABR van deze MGE's is nog weinig onderzocht. Het genestelde poppengedrag zoals bij Matroesjka poppen van deze elementen, waarbij ze van en naar verschillende plasmiden en naar het bacteriële chromosoom bewegen, maakt duidelijk hoe moeilijk het is om deze elementen naar behoren te identificeren en te traceren. Zoals beschreven in **hoofdstuk 8** werden niet-verwante microben met vergelijkbare fenotypische resistentieprofielen gedetecteerd. Vanwege het gelijke fenotype werd de mogelijkheid van eenzelfde ABR-plasmide overwogen, en onderzocht met "long read" WGS. Een groot (46 kilobase) genetisch-eiland, coderend voor 14 ABR-genen, werd geïdentificeerd in verschillende bacteriële soorten, gecodeerd op plasmiden

maar ook in het chromosoom. Veel van deze ABR-genen werden geflankeerd door IS26 insertiesequenties. Deze IS26 worden vaak geassocieerd met ABR-genen en zij voegen zich bij voorkeur in andere IS26-rijke regio's in het genoom, waardoor gencassettes worden gegenereerd die bestaan uit ABR-genen die door IS26 worden geflankeerd. De overdracht van dit gen-eiland kan gedeeltelijk worden verklaard door de uitwisseling van genetisch materiaal via conjugatieve plasmiden naar andere leden in deze taxonomische orde. Ten slotte werden 1806 WGS-datasets van ESBL-producerende *Enterobacterales* van gehospitaliseerde patiënten in Nederland 2011 tot 2014 gescreend op dit resistentie-gen-eiland, om na te gaan of dit gen-eiland al vóór 2019 in Nederland aanwezig was. Er werden 50 isolaten geïdentificeerd die coderen voor dit gen-eiland, waarvan er 37 afkomstig waren uit een van de ziekenhuizen uit de studie van **hoofdstuk 8**, wat impliceert dat dit resistentie-gen-eiland al minstens 7 jaar eerder aanwezig was in dit ziekenhuis. De verspreiding van transposons of deze gen-eilanden wordt vaak over het hoofd gezien omdat ze moeilijk te identificeren zijn zonder volledige genomen te analyseren. Alleen wanneer zeldzame en alarmerende resistentiemechanismen (bijvoorbeeld carbapenemases) in het spel zijn, zijn deze elementen geïdentificeerd als de veroorzakers van hun verspreiding. De langdurige aanwezigheid van dit resistentie-gen-eiland wijst op de noodzaak van toezicht op en meer begrip over deze mobiele genetische elementen om de verspreiding van ABR naar nieuwe bacteriestammen in een nosocomiale omgeving te beperken.

Ten slotte heeft WGS voor stamtypering zich bewezen als een betrouwbaar instrument voor infectiepreventie en surveillance van de verspreiding van nosocomiale bacteriën. Toch wordt WGS gebruikt tijdens of na uitbraken en worden mobiele genetische elementen bijna nooit in overweging genomen. WGS zal een belangrijke schakel worden en tot meer inzicht leiden in de rol die MGE's spelen bij verspreiding van antibiotica resistentie. Deze kennis zal vervolgens een bijdrage leveren aan de ontwikkeling van infectiepreventieve interventies die verdere verspreiding van resistente micro-organisme.

11

IMPACT PARAGRAPH

Antimicrobial resistance (AMR) has become a global problem and has been declared an immediate threat to mankind by the World Health Organisation (WHO) requiring urgent and harmonized efforts to curb moving into a post-antibiotic age, where previously treatable common infections become untreatable. This will have serious consequences, such as a higher demands on healthcare, increased morbidity and premature death. One of the major pillars in the fight against AMR is improved surveillance to monitor the presence and dissemination of AMR bacteria. New methods for bacterial typing have been developed, such as second generation (short-read) and third generation (long-read) sequencing, giving the highest resolution in resolving phylogeny among bacterial isolates. Determining phylogeny among bacterial isolates is crucial in order to find key steps in the dissemination of AMR and prevent further spread.

Firstly, the studies in this thesis contribute to the knowledge on how to perform surveillance of bacterial pathogens by whole genome sequencing (WGS). More laboratories are moving to WGS for bacterial typing and use a plethora of methods. When WGS is utilized to characterize isolates for surveillance beyond a local level, protocols need to be harmonized to obtain comparable and interoperable data for surveillance. Without identical procedures, multi-centre outbreaks or outbreaks across a region or country may not be recognized, which may put a burden on healthcare due to decreased patient health, unnecessary isolation of patients and the general waste of time and resources if outbreaks are not contained or prevented.

One of the advantages of WGS is the detection of AMR genes, indicating putative phenotypic antimicrobial resistance of bacterial isolates. Yet, the performance of bioinformatic pipelines to identify AMR genes in the bacterial genomes is still poorly benchmarked. Work outlined in this thesis describe curated WGS datasets for AMR detection of the highest quality to enable such benchmarking, making it a valuable resource for laboratories who perform WGS and want or need to optimize their own bioinformatic data analysis. Furthermore, we identified steps in bioinformatic data processing which can hamper the detection of AMR genes. Improvement of WGS analysis for AMR identification will lead to improved decision making by clinicians and infection prevention specialists, ultimately benefitting the patient. Additionally, all WGS data generated for this thesis and associated metadata has been deposited in public repositories, available to others to retrieve and re-use.

Lastly, one of the main focus-points of this thesis is the plethora of mobile genetic elements (MGE) and their gene content. This field of study remains understudied by genomics due to the difficulties in examining these genetic units. The advances in the field of (bacterial) genomics and sequencing technologies finally made it feasible to do large scale analyses on MGEs. Outcomes of the studies described in this thesis contribute to the knowledge of the role of these MGEs in the spread of AMR and has highlighted the necessity to look further than just the contribution of bacterial iso-

lates in the spread of AMR worldwide. Dissemination of this new knowledge on MGEs and AMR has been distributed on various platforms ranging from scientific journals and conferences, bioinformatic hackathons and social media (Twitter & LinkedIn).

DANKWOORD

Alles komt tot zijn eind, ook mijn promotietraject. Nu is het tijd om in een moment van bezinning mijn dank aan iedereen te verwoorden en neem ik jullie allen graag mee naar memorabele passages wat mij bijgebleven is aan iedereen die van significante invloed is geweest tijdens mijn promotietijd.

Promotieteam

Ik kan het mij nog goed herinneren, ik kwam voor sollicitatiegesprek bij jou, Lieke van Alphen en Christian Hoebe. Christian, Ik gaf een hand aan jullie, en nog voordat ik zat, schoot jij gelijk de eerste vraag af, waarna ik onthutst een monoloog afstak. Lieke moest er nog aan te pas komen om nog vlug koffie aan te bieden voordat het gesprek van start zou gaan. Verder zou dat de laatste keer zijn dat jij, Lieke, mij koffie aangeboden hebt en is het al vlug overgegaan naar speciaal biertjes bij onder andere Thembi. Lieke, onze wekelijke meetings samen verliepen vrijwel vlekkeloos waarbij wij in een vlot tempo over van alles hebben gebrainstormd, wat niet altijd overkwam bij andere die soms aansloten op onze meetings. Ook spoorde jij mij aan om aan te sluiten bij verschillende landelijke symposia en meetings, waaronder Special Interest Group bioinformatica (SIG).

Christian, het is een geweldige eigenschap van jou, hoe goed op de hoogte je altijd bent geweest van mijn progressie tijdens mijn PhD. Verder zijn de mij de activiteiten zoals team building en pubquizzen bij de i-4-1 Health meeting mij het best bijgebleven. Na een succesvolle pubquiz ook samen met Mitch van Hensbergen, hebben wij het plan gemaakt om een Pizza-Bordspellen avond te houden. Op het moment dat wij een datum geprikt hadden, startte de pandemie en is het er helaas nooit meer van gekomen.

Paul, samen met Lieke waren we voor het i-4-1 Health project in Antwerpen. Die avond zaten we in afwachting van een tafeltje in een restaurant aan de praat met de bediening. Van geen kwaad bewust vroeg de desbetreffende meneer of wij op vakantie waren met ons gezin (Paul, Lieke en ik).

Graag wil ik jullie alle drie bedanken voor echt alle vrijheid die ik heb gekregen om tijdens mijn promotietraject om projecten samen met anderen buiten het MUMC+ verder uit te werken en dat heeft dan ook mogen leiden tot eigenlijk ieder hoofdstuk van mijn thesis in samenwerking is geweest met collega's uit andere delen van Nederland en internationaal.

Paranymphen, Heike & Judith. De vrijdag voor carnaval 2020 is er een om nooit te vergeten! Na een avondje stappen met alle MMB-collega's kwamen we erachter wij als laatste drie strijdend overgebleven waren ergens in de binnenstad. Judith, sorry

dat ik altijd gezegd heb dat het feestje bij de scientific springmeeting echt episch is, en dat je zeker mee moet, om vervolgens zelf af te haken en niet kwam opdagen. Graag bedank ik je toch voor alle dropjes waarmee jij mij hebt voorzien toen ik nog samen op het kantoor met je zat. Heike, na vele bakken koffie hadden we echt een puik hobby-research plan. We kwamen met allerlei monster types die we wilde gaan testen, maar uiteindelijk zijn we gestrand dat we tomatensaus wilde gaan testen. Helaas kwam de pandemie en hadden we geen tijd meer om hierop te gaan PCRen. Het enige aandenken aan onze hobby research brainstormen zijn de pakken tomatensaus die nog in een lade stof aan het vangen zijn. Verder, Heike, pas op voor kwaadwillige collega's als je vastzit in de anaerobe kast.

Mitch van Hensbergen, aan het eind van alle verpleeghuisbezoeken samen met Math kwamen we erachter dat wij toch veel zelfde interesses en humor hadden. We spraken toen beide uit dat het jammer was dat we daarna nooit meer samen zouden werken. Maar ineens kwam daar een idee! Omdat we beide bordspellen graag spelen, zou ik aansluiten bij je wekelijkse Gloomhaven bordspelavond met Tom. Dit hebben we twee jaar volgehouden, waarna wij eindelijk na al die avonden bordspellen, Gloomhaven uitgespeeld hebben.

Niels en John, ik kan niet anders dan jullie samen benoemen hier. Er zijn weinig duo's te bedenken die zo goed op elkaar zijn ingespeeld dat je zou zeggen dat ze getrouwd zijn. Maar ik probeer het toch: Bert en Ernie, Buurman en Buurman, Batman & Robin, de Blues Brothers, de bitterbal & Mosterd, enzovoort, enzovoort. Dat jullie enige tijd geleden samen een huis zijn gaan bezichtigen is iets waar niemand vreemd van hoeft op te kijken, toch? Verder kan ik mij een mooie lente dag herinneren, waar de COVID op-hok frustraties door de pandemie de spuigaten uitliepen, en Niels voorstelde om naar John te fietsen in Beek. Na een heuse tocht door berg en dal kwamen we aan en werd onze reis beloond met echt het beste pistache ijs ooit gegeten.

Jordy Coolen, op een winderige en regenachtige oktoberdag een toertocht gaan fietsen in zuid Limburg, waar we al na 10km zeiknat geregend waren, was misschien achter niet zo'n goed idee. Echter was dit wel het voortvloeisel van een vriendschap die is ontstaan nadat wij een paar jaar eerder toegezegd hadden om kartrekker te worden van wat nu bekend staat als onze benchmark studie. Ontzettend gaaf om nu terug te kijken op onze samenwerking en wat er mogelijk is als een klein groepje gelijkgestemde mensen bijeenkomen (de SIG).

Erik Beuken, tot de dag van vandaag ben ik er nog altijd niet in geslaagd om groene duvels te bemachtigen. Samen met jou heeft dit bier een soort van mythische status

bereikt. Ontzettend bedankt voor tomeloze inzet om al het moleculair knip-en-plak werk in het lab te verrichten.

Brian, Gianluca, Matthew, Giang, Jiyang, David, Christel, Mayk, Tessa, Nader, Birke, Kevin, Julius, Daniëlle jullie zijn allen verantwoordelijk voor de fijne sfeer op de (research) afdeling. Misschien nog belangrijker, bedankt voor alle leuke activiteiten zoals de vrimibo, gravelbiken of boulderen buiten het werk.

De moleculaire diagnostiek, Chris, Jozef, Petra Wolffs, Petra Hendrickx, Selma, Frederique, Bart, Ilse, Molly, Anita, Elke, Judith, Raymond, Katrien, Lizzy. Begin pandemie ben ik uitkomen helpen bij de resultaatverwerking van de COVID diagnostiek. Jullie allen hebben ervoor gezorgd dat ik in een warm bad terecht ben gekomen in deze periode.

Pappa, Mamma, Coco en Noortje, nadat ik vertrokken was uit Maastricht om te studeren dachten sommigen van jullie dat ik nooit meer terug zou keren naar het zuiden. Maar niets was minder waar. Graag benoem ik ook nog dat ene autoritje, van Maastricht naar Utrecht door jou, Pappa. De dag ervoor was ik op gesprek geweest voor deze PhD plek, en een dag later zou ik voor een andere baan in Utrecht op gesprek mogen komen. Jij bent mij die ochtend gaan brengen. Echter een uur nadat jij mij afgezet had werd ik door dat telefoonnummer gebeld uit Maastricht. Initieel dacht ik dat Mamma belde dus nam ik ook niets vermoedend de telefoon op. Maar het was Lieke die belde om te vragen of ik in Maastricht wilde komen werken, waarop ik gelijk ja op heb geantwoord. Hele rit voor niets dus!

Last but not least, Mylène. Nadat ik besloot om een half jaar weg te gaan naar Amerika kwam terug en solliciteerde ik op die ene baan in Maastricht. Zelf had jij met je mamma een tripje naar Griekenland gepland, een paar weken na mijn terugkomst uit de VS. Tijdens jouw vakantie kreeg ik dat telefoontje of ik wilde beginnen in Maastricht. Daarmee werden gelijk onze plannen om rond Utrecht te settelen geannuleerd en zijn we aan ons volgend avontuur begonnen. Dankjewel dat je altijd naar mijn geratel blijft luisteren en met mij altijd in de startblokken klaarstaat om een volgend avontuur aan te gaan.

CURRICULUM VITAE

Casper Jamin werd geboren te Maastricht op 13 december 1992. Na het afronden van zijn VWO aan het Sint Maartens College, in Maastricht, in 2011, begon hij aan de bachelor studie Biotechnology aan Wageningen University & Research Centre. Na het behalen van zijn Bachelor diploma vervolgende hij zijn studieloopbaan met de gelijknamige masterstudie Biotechnology. Aan het eind van zijn masterstudie, heeft hij een half jaar aan UC Davis, Californië zijn afstudeerstage volbracht, waarna hij zijn diploma in 2017 behaalde. In datzelfde jaar begon hij als promovendus bij de medische microbiologie van het MUMC+, waar hij onderzoek deed naar de verspreiding van antibiotica resistentie genen. Na het behalen van zijn PhD zal hij werkzaam blijven binnen het MUMC+ als onderzoeker op ditzelfde onderzoek. Gelijktijdig zal hij van start gaan met de opleiding tot medisch moleculair microbioloog.